P-341106

UNLIMITED

304pp.

AGARD-AG-337

AGARD-AG-337

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARDograph 337

Multi-Sensor Multi-Target Data Fusion, Tracking and Identification Techniques for Guidance and Control Applications

(les Techniques de poursuite et d'identification multi-cibles à base de fusion multi-senseur appliquées au guidage et au pilotage)

This AGARDograph has been sponsored by the Mission Systems Panel of AGARD.

NORTH ATLANTIC TREATY ORGANIZATION

Published October 1996

SJS. 9-1-97

Distribution and Availability on Back Cover

UNLIMITED

AGARD-AG-337

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARDograph 337

Multi-Sensor Multi-Target Data Fusion, Tracking and Identification Techniques for Guidance and Control Applications

(les Techniques de poursuite et d'identification multi-cibles à base de fusion multi-senseur appliquées au guidage et au pilotage)

Edited by

Dr. David Liang, Dr. Steve Butler, Dr. Carlos Garriga, Dr. Bruno Mazzetti, Mr. Thierry Uring and Dr. Heinz Winter

This AGARDograph has been sponsored by the Mission Systems Panel of AGARD.

DRIC DISTRIBUTIONS
KENTIGERN HOUSE
65 BROWN STREET
GLASGOW G2 8EX
TELE: 041 224 2434

North Atlantic Treaty Organization
Organisation du Traité de l'Atlantique Nord

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Exchange of scientific and technical information;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published October 1996

Copyright © AGARD 1996
All Rights Reserved

ISBN 92-836-0031-2

Printed by Canada Communication Group
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Multi-Sensor Multi-Target Data Fusion, Tracking and Identification Techniques for Guidance and Control Applications

(AGARD AG-337)

Executive Summary

The aim of this AGARDograph is to provide a quick overview of practical advances in multi-sensor, multi-target tracking (MS/MTT) technology and applications. In order that this AGARDograph could serve as a useful reference for those involved in the design, development, simulation, and applications of the techniques and technology, we have encouraged our authors to take, as much as possible, a tutorial approach. This will provide the general summary of the MS/MTT techniques and technology with emphasis towards practical implementation.

Many examples of sensor fusion involve the methodology of merging various track files taken from different sensors. This allows for more consistent, accurate, and reliable tracks than might be possible with any of the individual systems acting alone. Section I relates to the important use of sensor fusion prior to establishing a firm track file. By combining raw sensor information, greater discrimination of targets from background may be possible from the augmented body of available information.

Tracking and fusion with multiple sensors deals with integration and correlation of data from diverse sources in order to arrive at the best possible situational assessment. In Section II, we present the tutorial on representative data association and filtering techniques, and also address some of the key initiation issues, approaches and track management methodology that simplify and enhance the practical implementation.

Section III presents different types of classification algorithms, Bayesian Belief Networks, and Neural Networks covering the complete Automatic Target Recognition process, including fusion, segmentation and classification, that are very promising for real-time, or quasi-real-time systems applications.

Section IV covers the handling of Automatic Target Recognition (ATR) test data, deals with an effective tool to support the development of precision guided munitions, and presents a study of target acquisition and sensor cueing in air-to-air environment. The last Section presents several practical examples of MS/MTT applications.

Les techniques de poursuite et d'identification multi-cibles à base de fusion multi-senseur appliquées au guidage et au pilotage

(AGARD AG-337)

Synthèse

Cette AGARDographie offre un bref aperçu des avancées concrètes réalisées dans le domaine des technologies et des applications de la poursuite multicible/multisenseur (MS/MTT). Nos auteurs ont été encouragés à adopter, dans la mesure du possible, une approche, afin que cette AGARDographie puisse servir d'ouvrage de référence à tous ceux qui sont associés à la conception, au développement, à la simulation et aux applications des différentes techniques et technologies mises en jeu. Ainsi, cet ouvrage fournit un résumé général des techniques et des technologies MS/MTT en mettant l'accent sur les modalités pratiques de leur mise en œuvre.

Dans beaucoup de cas, le fusionnement des senseurs fait appel à une méthodologie qui permet le regroupement de certains fichiers de poursuite obtenus de différents senseurs. Cette technique permet d'obtenir des pistes plus homogènes, plus fidèles et plus fiables que celles obtenues par un quelconque système particulier, fonctionnant seul. La section I concerne l'application très importante du fusionnement des senseurs en vue de l'établissement d'un fichier de poursuite validé. Le volume accru d'informations résultant de la fusion des données brutes des senseurs permet souvent une meilleure discrimination de la cible par rapport au bruit de fond.

La poursuite et la fusion à l'aide de senseurs multiples consistent dans l'intégration et la corrélation de données de sources diverses afin d'aboutir à la meilleure évaluation possible de la situation. La section II présente un cours sur les techniques d'association et de filtrage des données représentatives et examine en même temps certaines questions clés concernant l'initiation à la gestion de la poursuite, les différentes approches à adopter et la méthodologie permettant de simplifier et faciliter sa mise en œuvre.

La section III présente différents types d'algorithmes de classification, de réseaux de croyances de Bayes, et de réseaux neuronaux couvrant l'ensemble du processus de reconnaissance automatique de la cible, y compris la fusion, la segmentation et la classification. Ces éléments s'annoncent très prometteurs pour des applications systèmes en temps réel ou quasi-réel.

La section IV couvre le traitement des données de test des systèmes de reconnaissance automatique de la cible (ATR). Un outil efficace d'aide au développement des engins guidés de précision est présenté, ainsi qu'une étude sur l'acquisition de la cible et l'alignement des senseurs en environnement air-air. La dernière section présente un certain nombre d'exemples concrets d'applications MS/MTT.

Contents

	Page
Executive Summary	iii
Synthèse	iv
Mission Systems Panel	vii
General Introduction	viii

	Reference
SECTION I: MULTI-SENSOR PHENOMENOLOGY AND SENSOR SIGNAL PROCESSING	
Introduction by Dr. S. Butler	1
Multi-Frequency Phenomenology Fusion via an Ultra-Broadband Hybrid Sensor Technology Enhancing Surveillance and Target Classification/ID Performance by R.F. Ogrodnik	3
Interactive Integration of Passive Infrared and Radar Horizon Surveillance Sensors to Extend Acquisition and Firm Track Ranges by S.R. Horman, R.A. Stapleton, K.C. Hepfer, R.M. Headley and J.K. Stapleton	14
The Fusion of an IR Search and Track with Existing Sensors to Provide a Tracking System for Low Observability Targets by P.V. Coates	32
The Application of MMW/IR Sensor Fusion to Tactical U.S. Air Force Weapon Systems by J.W. Watson, S. Amphay and B. Sundstrom	48
SECTION II: DATA ASSOCIATION AND TRACKING TECHNIQUES	
Introduction by Dr. D.F. Liang	65
A Practical Overview of Multiple Target Tracking Algorithms by S.S. Lim and D.F. Liang	67
Ballistic Missile Track Initiation from Satellite Observations with Extrapolation to Impact by M. Yeddanapudi, Y. Bar-Shalom, K.R. Pattipati and S. Deb	76
Application of Multiple Hypothesis Tracking to Multi-Radar Air Defense Systems by S.S. Blackman, R.J. Dempster and T.S. Nichols	96
Multisensor Tracking and Fusion with MTI Radars by K.-C. Chang and Y. Bar-Shalom	121

System-Level Track Fusion for Command and Control 130
by M.P. Dana and J.L. Dana

SECTION III: PIXEL AND SYMBOL LEVEL IMAGE FUSION, TARGET CLASSIFICATION AND RECOGNITION

Introduction 147
by Dr. C.A. Garriga

Classification par fusion de données incertaines multi-senseurs 149
by A. Appriou

The Use of Bayesian Belief Networks to Fuse Continuous and Discrete Information for Target Recognition, Tracking, and Situation Assessment 160
by L. Stewart and P. McCarty Jr.

Grey Level Segmentation with Selectable Number of Discrimination Levels Using Hopfield-Like Neural Networks with Constraint Satisfaction Criteria 166
by A. Domingo and J. Santamaria

Scene/Object Classification and Segmentation Using MultiSpectral Data Fusion 174
by L.E. Lazofson, T.J. Kuzma, H.C. Choe, E.B. Preston and J.D. Chovan

SECTION IV: SIMULATION AND PERFORMANCE EVALUATIONS

Introduction 185
by Dr. B. Mazzetti

Benchmarking ATR Test Data 187
by A. Hauter, V. Diehl, A. Williams, G. Orsak and M. Sorell

The Modular Algorithm Concept Evaluation Tool (Macet) Workstation 203
by J. Watson, B. Williams, S. Talele and S. Amphay

Studies and Simulations on Sensor Fusion and Cueing for Fighter Application 224
by M. Avalle

SECTION V: DATA FUSION FOR GUIDANCE AND CONTROL APPLICATIONS

Introduction 233
by Dr. D.F. Liang

Air Defence Radar Surveillance Systems Tracking Assessment 235
by S.S. Lim, D.F. Liang and M. Blanchette

Low Cost Multi-Sensor Suites for Surveillance and Weapon Guidance 251
by B.R. Suresh

Sensor Data Fusion for Air to Air Situation Awareness Beyond Visual Range 268
by C.A. Noonan

Multi-Sensor Data Fusion in Command and Control and the Merit of Artificial Intelligence 284
by R.G. Zuidgeest

Mission Systems Panel

Chairman: Mr. J.K Ramage
Chief, Flight Control Systems Branch
WL/FIGS, Bldg 146
2210 Eighth St, Ste 11
Wright-Patterson AFB, OH 45433-7521
United States

Deputy Chairman: Prof Dr Heinz Winter
Director,
Institut für Flugführung für Luft und Raumfahrt e.V. Flughafen
Postfach 32 671
D-38022 Braunschweig
Germany

TECHNICAL PROGRAMME

Programme Director and Editor: Dr David F Liang

Head, Space Systems and Technology
Defence Research Establishment Ottawa
Department of National Defence
Shirley Bay, Ottawa K1A 0Z4
Canada

PANEL EXECUTIVE

Lieutenant-Colonel P. Fortabat, FAF

Mail from Europe:
AGARD-OTAN
Attn: MSP Executive
7, rue Ancelle
F-92200 Neuilly-sur-Seine
France

Mail from USA and Canada:
AGARD-NATO
Attn: MSP Executive
PSC 116
APO AE 09777

Tel: 33 (1) 47 38 57 80
Telex: 610176 (France)
Telefax: 33 (1) 47 38 67 20/57 99

ACKNOWLEDGEMENT/REMERCIEMENTS

The Programme Director and the Mission Systems Panel wish to express their appreciation to all authors who contributed to this AGARDograph and made its publication possible.

Le Directeur du Programme et la Commission missions et systèmes tiennent à remercier tous les auteurs qui contribuèrent à la réalisation et la publication de cette AGARDographie.

GENERAL INTRODUCTION

Startling advances in stealthy, high speed, accurate weapon and target systems have imposed stringent requirements on the performance of advanced surveillance, detection, tracking, identification, and classification systems to support military guidance and control applications. The effectiveness of advanced weapon systems as well as the defence against them will be highly dependent upon the timely availability of accurate battle-sphere sensor and intelligence information. The utilization of any single sensor will no longer be sufficient to cope with increasingly demanding operational requirements and challenging mission environment. The effective use of multiple and multi-spectral sensors has become increasingly important for situational assessment, planning support and command and control operational decisions. The importance of timeliness and accuracy requirements was clearly demonstrated in Desert Storm operations.

Therefore the development and fielding of effective multi-sensor, multi-target tracking (MS/MTT) systems together with the associated classification and identification systems becomes essential for effective detection, tracking and identification of increasingly sophisticated targets in challenging threat environments.

The aim of this AGARDOGRAPH is to provide a quick overview of practical advances in this broad but important area of MS/MTT technology and applications. In order that this AGARDOGRAPH could serve as a useful reference for those involved in the design, development, simulation, and applications of the techniques and technology, we have encouraged our authors to take, as much as possible, a tutorial approach. This will provide the general summary of the MS/MTT techniques and technology with emphasis towards practical implementations.

Even though in recent years there have been numerous conferences and books published on this subject, this AGARDOGRAPH has the advantage of having been able to identify contributions from government, defence, academic, and industrial establishments within all NATO nations. Therefore this AGARDOGRAPH has attempted to present specific advances and applications made in recent years on both sides of the Atlantic. There are many other excellent contributions which we would love to have included, but we were constrained by a limited budget and a fixed schedule for publication.

This AGARDOGRAPH is organized into the following five Sections:

— Section I: Multi-Sensor Phenomenology and Sensor Signal Processing;

— Section II: Data Association and Tracking Techniques;

— Section III: Pixel and Symbol Level Image Fusion, Target Classification and Recognition;

— Section IV: Simulation and Performance Evaluations; and

— Section V: Data Fusion for Guidance and Control Applications.

The Editorial Committee consists of Dr. Steve Butler of the U.S., Dr. Carlos Garriga of Spain, Dr. David Liang of Canada, Dr. Bruno Mazzetti of Italy, Mr. Thierry Uring of France, and Dr. Heinz Winter of Germany.

Dr. David F. Liang
Head, Space Systems and Technology
Defence Research Establishment Ottawa
Department of National Defence
Shirley Bay, Ottawa
Canada K1A OZ4.

SECTION I

MULTI-SENSOR PHENOMENOLOGY AND SENSOR SIGNAL PROCESSING

INTRODUCTION

by
Dr. Steve Butler
Technical Director, HQ AFM/EN
4375 Chidlaw Rd, Ste 6
Wright-Patterson AFB, OH 45433-5006
USA

Many examples of sensor fusion involve the methodology of merging various track files taken from different sensors. This allows for more consistent, accurate, and reliable tracks than might be possible with any of the individual systems acting alone. This chapter, however, relates to the important use of sensor fusion prior to establishing a firm track file. That is, to merge sensor inputs prior to establishing the separation between target and background at each sensor. By combining raw sensor information, greater discrimination of targets from background may be possible from the augmented body of available information.

The classic example is the multispectral imager. Co-boresighted images are formed with each image representing a different optical passband. Each image provides spatial information and contrast in each passband. New images formed by the ratio of two scenes may highlight features unrecognized in either original. This is widely used in agricultural analysis of overhead imagery. Crop conditions appear in ratio images with much higher contrast than in individual scenes. In this case, the "target" is lost in the background in each independent image but is statistically separated from background when the images are fused. The air-to-air seeker works in this fashion to separate targets from decoys. A reticle used to chop the signal for tracking purposes can be produced to pass different infrared passbands in alternating spokes. The signal formed on the detector is, in effect, a result of the fusion of two signals in two passbands. This difference signal provides greater target discrimination possible with a single reticle. Aerosol chemical agents are detected using a ratio or subtraction technique similar to decoy detection.

Many fail to recognize these simplistic techniques as sensor fusion. If the concept is taken a step farther, it may be more clearly seen as a form of pre-detection fusion. In the case of a dual mode sensor comprised of boresighted Infrared and Millimeter wave detectors, the Constant False Alarm Rate (CFAR) thresholds of each detector might depend on (or be a function of) the other sensors output. That is, the radar might be alerted to possible target existence by the infrared sensor. The convergence of signals in both bands might create a "product" signal

with greater contrast or discriminatory ability than either sensor acting alone. Many combinations of sensors might act in this fashion where one sensor provides information which helps set gains, offsets, thresholds, filter parameters, or in general, focuses the hypotheses which are tested in the search for a target. The signal processing parameters are modified according to the information gathered and optimize the performance of each sensor by taking advantage of information gathered by other sensors.

The basic phenomenology of some targets lend themselves well to this approach. Examples include the radiance and reflectivity of a hard body. The infrared signature might be reduced by lowering the emissivity of the target coating. However, if the sensor is equipped with an active emitter in the same band (e.g. laser radar), the lowered emissivity naturally causes a higher reflectivity. A potential target cell which elicits a high velocity feature from a radar sensor would be expected to have a measurable infrared feature. The lack of correlating features might indicate a countermeasure or false alarm. Scalar signals from acoustic and seismic sensors might indicate the presence of a tracked vehicle. The two signals can be fused to create a new information (range) which was not present in either signal independently.

While this discussion might be applied to the broader use of sensor fusion, it is intended here for the case when no detection or track has been established. The sensors work together to test all target hypothesis (all pixels could be the target). Only after the fused information yields the existence of a target, is a detection noted and a track file established. Such pre-detection fusion is an important area of research but is less common in the literature. The work at Rome Laboratory is widely published and comprises a significant portion of the pre-detection research in the United States. Mr. Watson's paper represents the great interest in the Army and Air Force to use fused signals from low cost millimeter wave radars and infrared sensors to detect targets with the same fidelity associated with considerably more expensive single mode devices. The paper from the Naval Surface Warfare Center is an example of the significant work ongoing in the Navy to improve long range detection ability through the use of fusion technologies. Mr. Coat's paper is similarly demonstrative of a greater body of work at Thorn EMI in the United Kingdom where fusion has demonstrated its ability to increase performance synergistically. The papers in this chapter highlight applications of such fusion but are not intended to distract from diversity of possibility.

MULTI-FREQUENCY PHENOMENOLOGY FUSION VIA AN ULTRA-BROADBAND HYBRID SENSOR TECHNOLOGY ENHANCING SURVEILLANCE AND TARGET CLASSIFICATION/ID PERFORMANCE

Robert F. Ogrodnik

Rome Laboratory (RL/OCTM)
26 Electronic Pkwy
GAFB NY USA 13441-4514
Tel: 315-330-4431

SUMMARY

Target phenomenology has observables which occur across the electromagnetic spectrum, and, are linked in their multispectral characteristics through time in the multi-frequency observation space, making target phenomenology an ideal candidate for detection domain level signal fusion. That is, given the ability to multispectrally observe target activity, as well as target phenomenon (i.e. missile multispectral launch transits, or, target vehicle body parts which may be dimensionally comparable with the surveillance wave length employed, etc.), an ideal surveillance approach, which supports detection domain fusion, would be to implement a broadband multispectrum surveillance sensor technology which could optimize its surveillance sensitivity and operating spectrum simultaneously on differing multispectral phenomena throughout the scenario life profile of the target. This paper addresses such a surveillance technology which exists today at Rome Laboratory. It is based on the integration of passive signal awareness monitoring combining real time signal parameter processing, and, passive coherent radar. This technology readily provides multispectral signal fusion by exploiting in real time background multiple ambient signal sources in order to synthesize a passive radar surveillance capability simultaneously at multiple carrier frequencies (as many diverse spectral sources as there are which constitute the electromagnetic ambient background in the target domain). This allows the target and its associated phenomena to be observed simultaneously in a multispectral fashion, optimizing detection performance whenever an optimal, or near optimal signal source is available for the purposes of enhanced detection and target classification.

1. INTRODUCTION

The focus of this work is on the benefits and the results of simultaneous multispectral signal observables fusion, and the merits which are derived from a broadband sensing technology which preserves critical exploitable signal components as well as the characteristics of these signals in support of detection enhancements and target characterization/ID operations. Target phenomena can be key in high confidence target classification/ID as well as the means for selective (template) signal component fusion processing to enhance interference/noise rejection, and, consequently improve small signal target surveillance sensitivity. The fusion of multispectral, multifrequency signal phenomenology, and the resulting sensing benefits provided by means of a very broadband sensor technology, are addressed here, both from a test data results aspect as well as from a sensor technology and signal domain fusion aspect.

The sensing technology is the integration of all signal awareness monitoring with passive coherent radar. This hybrided passive sensor technology merges multispectral signal phenomenology by multispectral sensing and multi-frequency signal fusion processing techniques. The principles of the broadband passive sensor technology addressed here are shown in Fig. 1 as a multi-mode operating sensor system. The first primary principle is the ambient signal or illuminator awareness is afforded by a passive signal intercept and waveform parameter analysis mode. This mode provides ambient illuminator monitoring, cataloging and selection for the passive radar exploitation (based on radar functions governed by waveform ambiguity properties versus desired passive radar operations, and, general surveillance operations). The secondary principle of this multi-mode passive sensor is its passive radar (broadband, multi-frequency noncooperative) which receives the illuminator selection controls from the signals awareness mode and coherently exploits its direct path for target echo processing. This processing produces real-time measurements of target range, range rate (and higher order target kinematics) and angular position. These measurements directly support real-time multiple target tracking and positional display. In order to explain the broadband nature of this sensor technology, examine Fig 2. where the signal awareness signal domain is displayed on a three-axis information graph. Fig 2. shows multiple ambient background signals cataloged along the signal power (Z-) axis, the angular or bearing (X-) axis, and the carrier frequency (Y-) axis. The signal domain shown here is composed of a high power level signal (at 9.4 GHz) and several associable target echoes (target reflections of this strong signal distributed along the Azimuth angle X-axis but lying close to the direct path frequency axis, 9.4 GHz) as well as several spurious non related background signals which could represent co-channel interference. All signals in this informational signal domain are digitized, stored in a

digital Radio Frequency memory (DRFM) and cross correlated with themselves, which leads to the direct measurement of target range, velocity and angle from all coherent direct path and target echo combinations which are randomly present. The output is a Planned Position Indication (PPI) track display giving the air situation report, or data such as that displayed in Fig 3.This figure shows the real time track display performance for the hybrid sensor equipment mounted in an airborne testbed flying over the Delaware MD area, where the exploited background signal was on the ground and the target echoes processed were from aircraft flying in the vicinity of the testbed.

2. TEST RESULTS FOR SIGNAL PHENOMENOLOGY FUSION

Several tests were conducted with this hybrid sensor technology to determine its multispectral phenomenological performance and potential support to detection signal domain fusion. Fig. 4 shows the hybrid sensor system simultaneous processing multi-frequency target returns for the purpose of contrasting radar cross-section results at differing near resonant and non resonant frequencies. Here in this figure the target radar cross-section (RCS) with target aspect angle is measured simultaneously at different carrier frequencies at UHF (495 MHz) and at VHF (175 and 205 MHz). Notice these RCS polar plots provide target shape information and directly support target typing operations. They reveal the scattering nature of the target surface at differing frequencies simultaneously.

The passive sensor technology was tested against a singular noncooperative target which possessed a rich target phenomenology observable set, namely a solid propellant missile during its launch (boost phase) stage. This was done at a test range where several tracking radars were netted to provide highly calibrated ground truth data for sensor data calibration and analysis purposes. The hybrid sensor exploited both background low frequency signals as well as the high frequency radar tracker signals. Fig 5. shows the multitude of sensor observables afforded by this missile target. These observables have their own and separately optimal signal domain set with very little commonality. Therefore, jointly they collectively demand a very broad sensor operating frequency domain if all these observables were to be simultaneously exploited for target typing and optimized detection/tracking operations. This broadband operation is easily met by the broadband sensor technology we are addressing here, as long as there is a broadband ambient illuminator set available to be exploited in the region and accessible to the missile launch site. Two widely separated illuminator frequencies were available for this purpose at the range site, namely VHF and C-Band ambient background illuminations. This diversity and frequency region was ideal to exploit launch phenomenon (see Fig 6. and Fig 8. for vertical axis acceleration observables, namely missile staging events versus powered flight and ballistic periods). The VHF sensor missile velocity tracking data is shown in Fig 7. (individual data dots) versus missile test range derived ground truth data (solid curve). From Fig 6., the velocity data (Fig 7.) extends over the first stage powered flight period, into its booster engine cut-off (BECO) event (occurring at 24 seconds) and through its ballistics region (26 seconds through 40 seconds), and finally second stage firing (40.6 seconds) seen in Fig 6. as the acceleration period beginning at 40.6 seconds). BECO is accurately measured by the missile kinematics measurements displayed in Fig 8. Note there the fairly steady acceleration from 20 to 23 seconds, with acceleration profile showing a strong downward trend at 25 seconds up to 40 seconds (the ballistics period). If we note the sensor-target geometry as depicted in Fig 9 in reference to the target aspect angle, as well as the radar range equation containing the on-site realized surveillance parameters, we can use these relationships to determine the target RCS with target aspect angle. Target measured RCS is shown in Fig 10. versus time from launch. Overlayed on this RCS data (data dots) is the missile acceleration profile with time from launch (solid curve derived from missile test range ground truth data). Note target RCS (missile with plume) remains relatively constant for a time period of 10 through 24 seconds, following the steady relatively constant acceleration profile of the missile first stage up to 24 seconds. After this time event the acceleration falls to ballistic values (o m/s/s at 25 seconds, -13 m/s/s at 29 seconds), with the target RCS falling also and following the overall trend of the first stage power profile (see Fig 10.). If we return to Fig 9 to note the target aspect geometric relationships in this passive sensor versus target setup, we can replot the missile RCS and acceleration profile versus target aspect angle. We do this in Fig 11. and add target predicted RCS (lower solid curve, methods of moments procedure used here for target 40 degree role and for this VHF frequency case [1]) in order to compare the missile hardbody RCS without plume effects (predicted) versus the plume plus hardbody RCS (actual measured data). Notice (with reference to the acceleration profile versus aspect) the hardbody plus plume measured data remains invariant over a 20 degree aspect angle change, suggesting that during the powered flight the RCS is dominated by plume effects. Compared to the predicted missile only RCS, the measured RCS data shows and cross-section enhancement of nominally 10 to 15 dB over the hardbody only. Fig 11. strongly suggests that observing this solid propellant missile target at VHF a 10 dB detection improvement is realizable via ambient illuminator exploitation. Note in Fig 11. the measured RCS at the ballistic phase falls back down to the hardbody only prediction results (112 through 122 degrees aspect angle interval. The range tracking performance at VHF for this passive sensor technology is shown in Fig 12. (measured data as dots, ground truth data as solid curve). The performance differed from ground truth by only 10 to 20 meters (for times earlier than 70 seconds). Similar range performance data exploiting a much high frequency, namely the on range C-Band trackers is shown in Fig 13. This higher frequency results supported better target kinematics detection and resolution, as it would be

expected as one goes up in frequency. Finally, the measured misslie plus plume data (measured data dots) at C-Band is shown in Fig 14., giving a much lowere RCS result, namely o dbsm, which compares favorably with predicted data (see solid curve in Fig 14.). The overall target RCS templates which are useable for rocket motor typing is shown in Fig 15, along with other typing observables.

3. CONCLUSION

The passive sensor technology which can observe targets at multiple frequencies simultaneously, can optimize the detection and analysis of target phenomenology, which is demanding in its frequency coverage. A broadband sensor suite, which is very broadband in nature, exploited VHF and C-band background ambient illumination in order to optimize detectivity and to analyze observables like kinematics based BECO and RCS measurements during launch transient periods (rocket motor characterization). This research will continue, focusing on automating many analysis operations as well as fusing the results at the detection domain rather than at the after track combining domain (were much of the data phenomenology is lost).

4. REFERENCES

[1] J. W. Crispin, et al, "A Theorectical Method for the Calculation of Radar Cross Section of Aircraft and Missiles", *Univ of Michigan Radiation Lab*, 2591-1-H, July 1959.

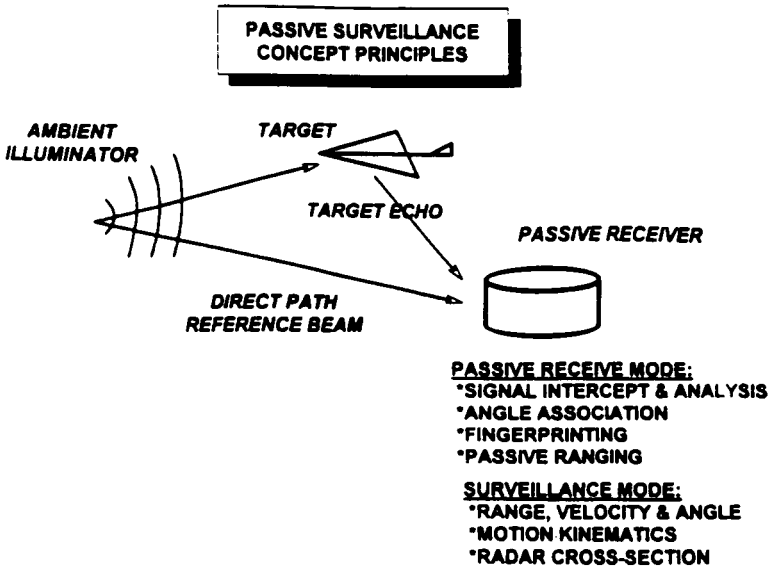


Fig. 1 Passive Sensor Operating Principles

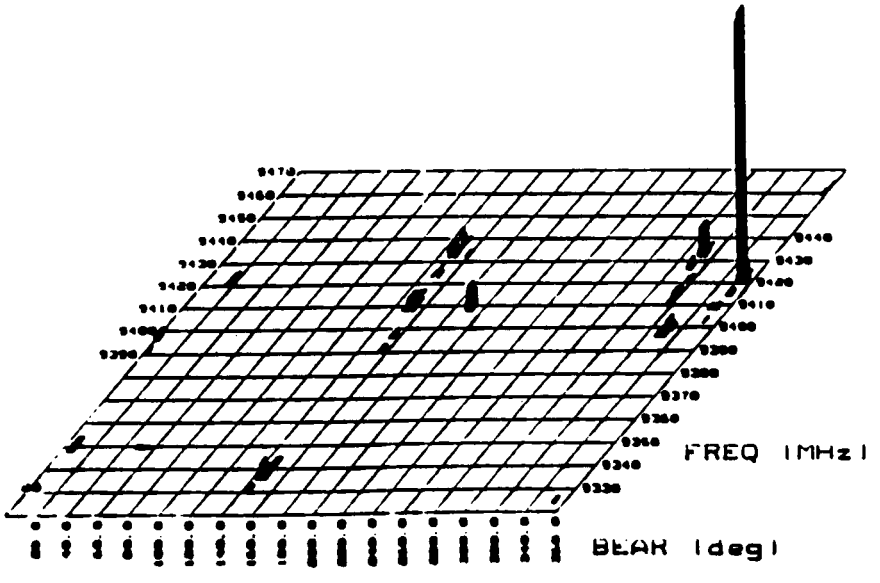


Fig. 2 Passive Sensor Signal Awareness Domain

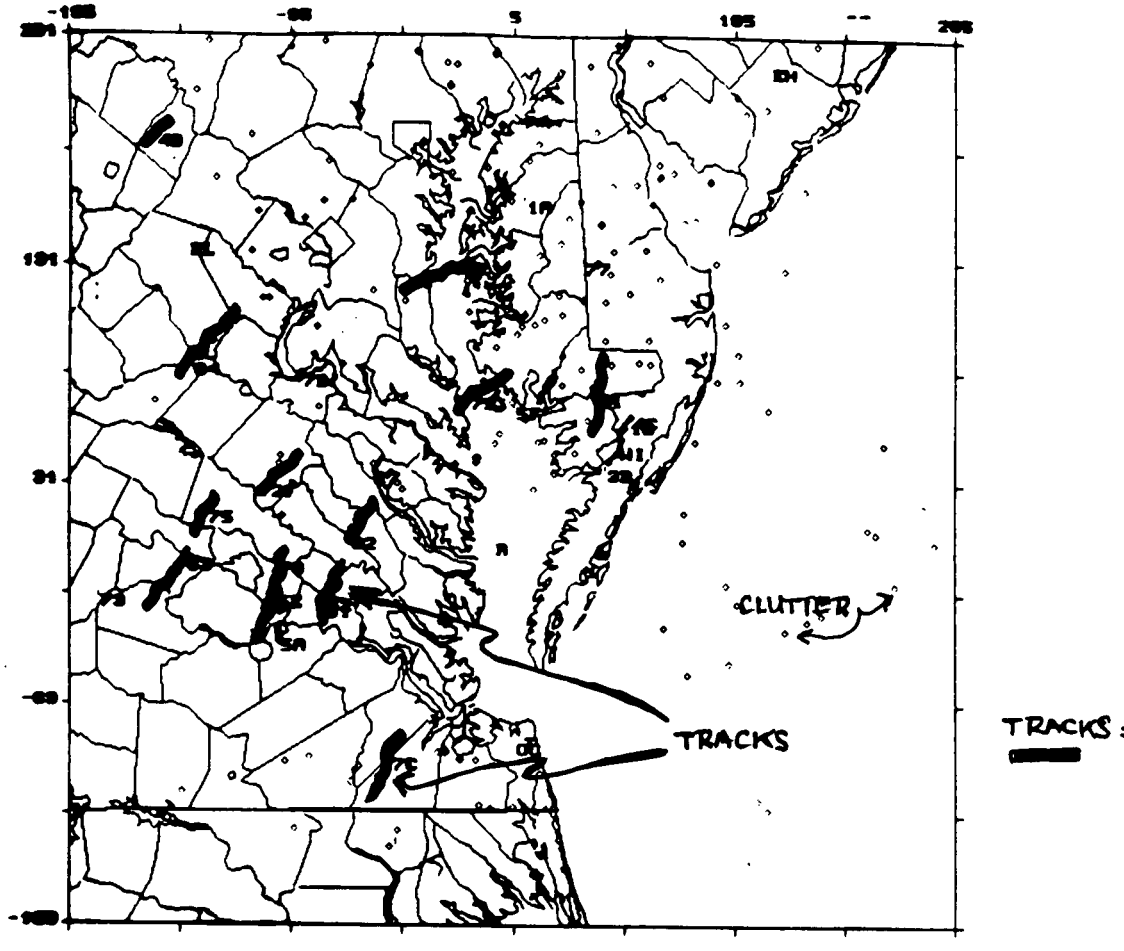


Fig. 3 Pulse Position Indication (PPI) Display

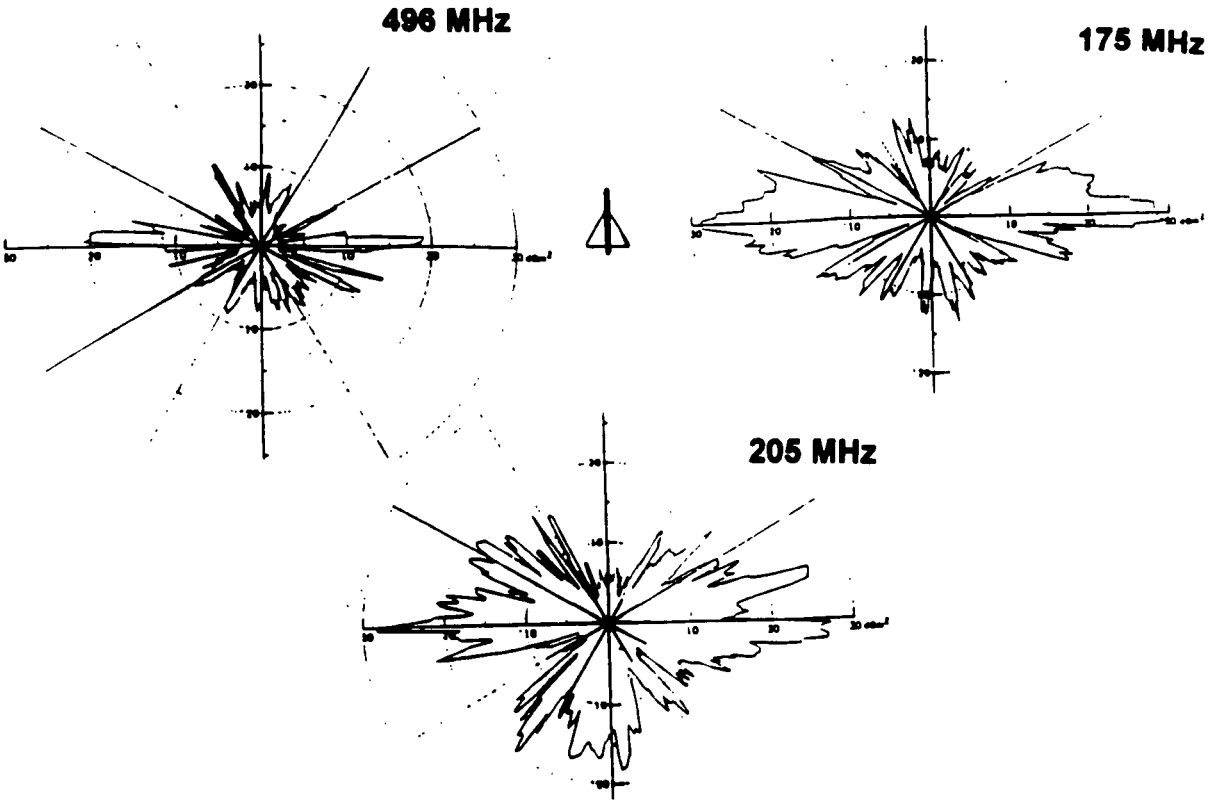


Fig. 4 Target Typing via Multi-Frequency Surveillance Modes

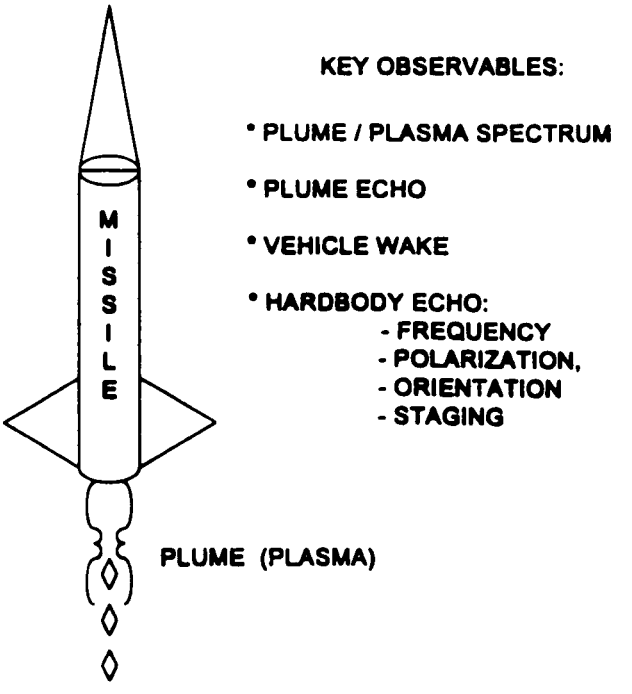


Fig. 5 Target Observable from a Multi-Frequency Aspect

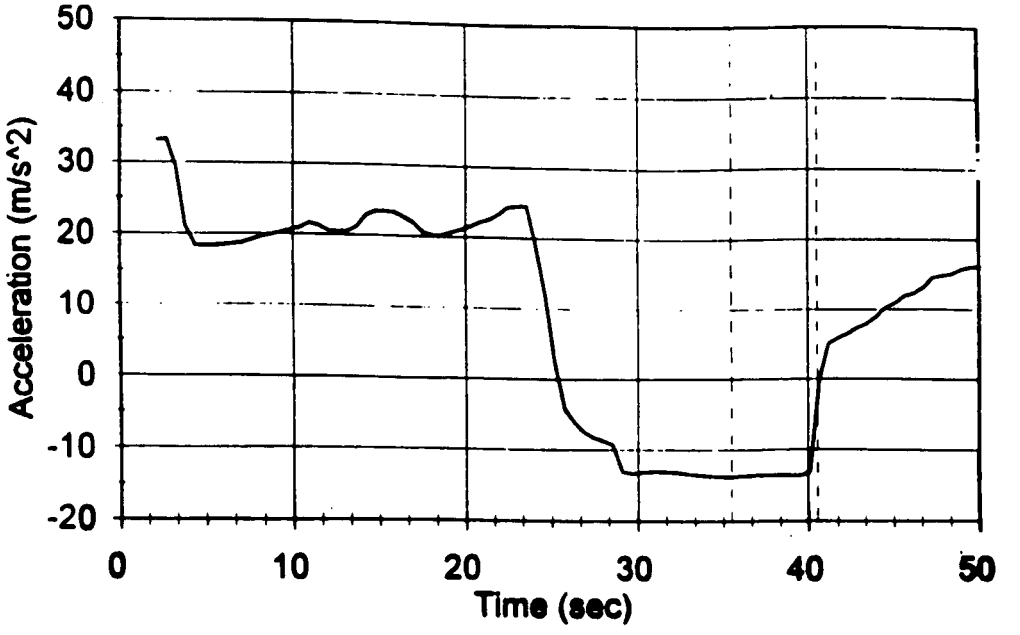


Fig. 6 Missile Target Verical Acceleration Profile

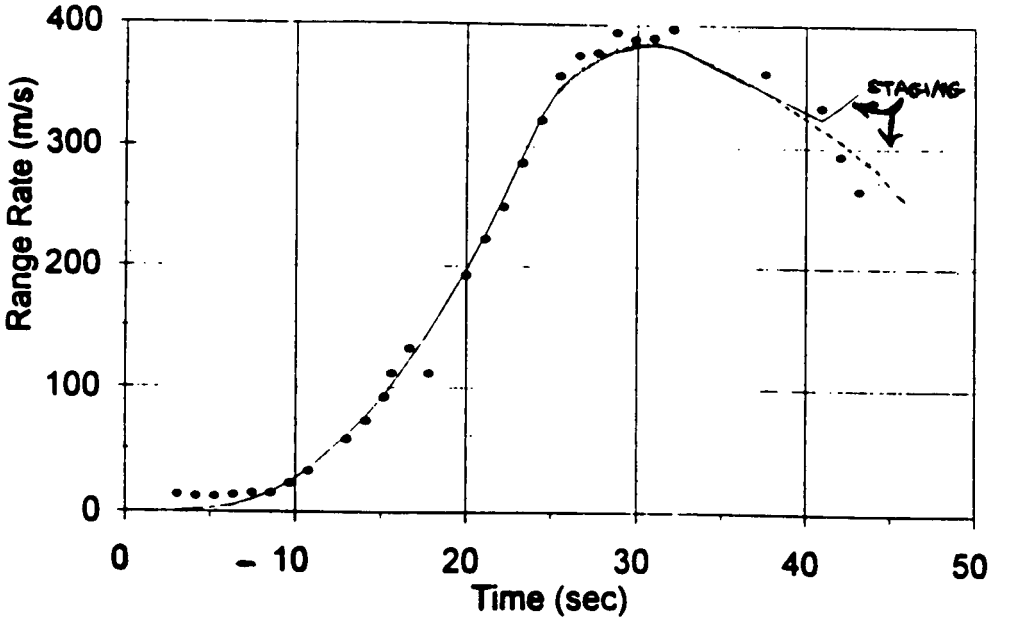


Fig. 7 Missile Target Range Rate vs Ground Truth

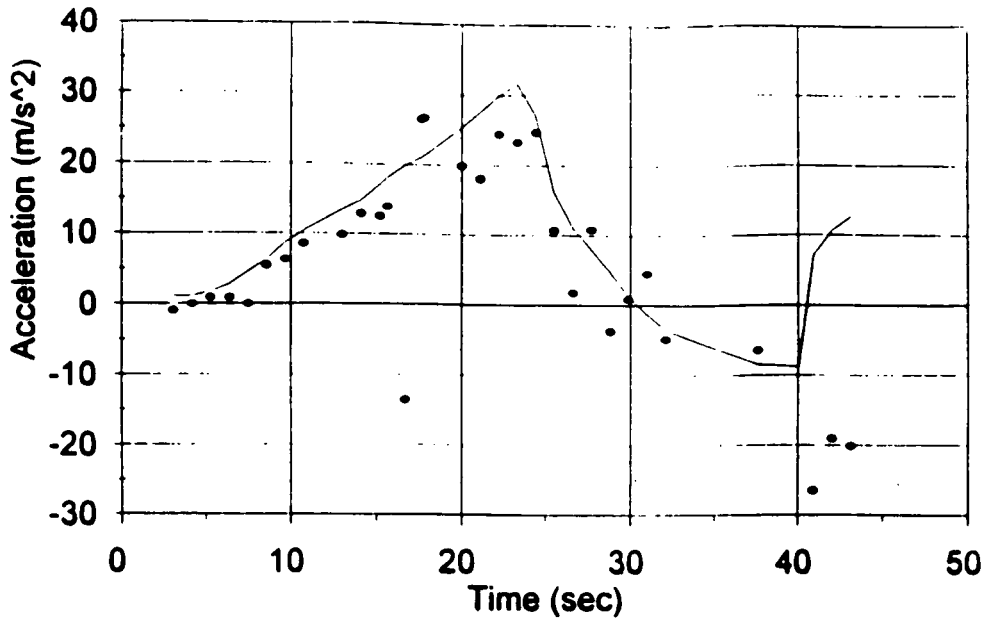
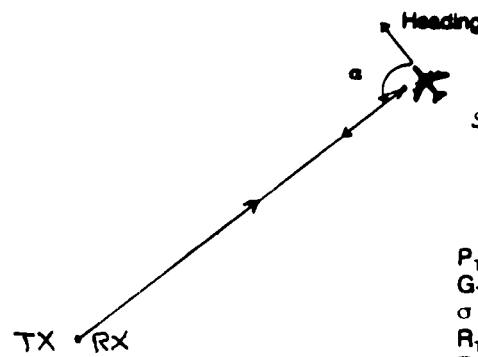


Fig. 8 Missile Target Acceleration Profile Data vs Ground Truth



$$SNR = \frac{P_T\, G_T}{4\pi R_T^2} \; \frac{\sigma}{4\pi R_R^2} \; \frac{G_R \lambda^2}{4\pi k T B N_f L}$$

P_T = Average Transmit Power
G_T = Transmit Antenna Gain
} = 74 dBm

σ = Radar Cross Section
R_T = Range from Transmitter to Target
R_R = Range from Receiver to Target
G_R = Receiver Antenna Gain = 10 dBi
λ = Wavelength = 1.75 dBm
kT = -174 dBm / Hz
B = Bandwidth = 3 dBHz
N_f = Noise Figure = 5 dB
L = Loss = 10 dB

TX = Transmitter
RX = Receiver
α = Aspect Angle

Fig. 9 Passive Sensor Geometry vs Aspect Angle

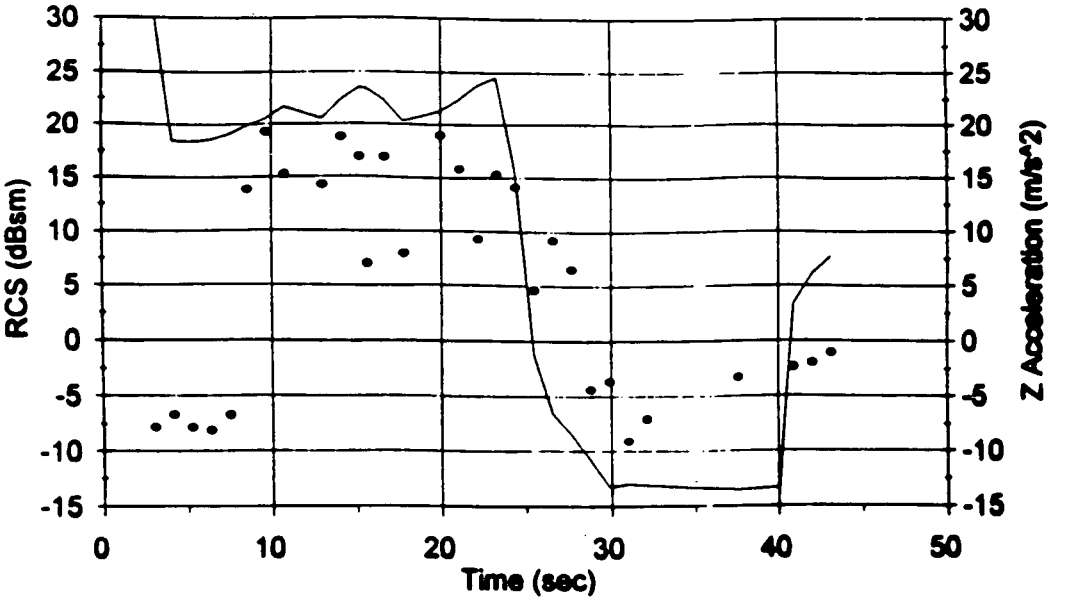


Fig. 10 Target Radar Cross-section(RCS) and Ground Truth Acceleration vs Time

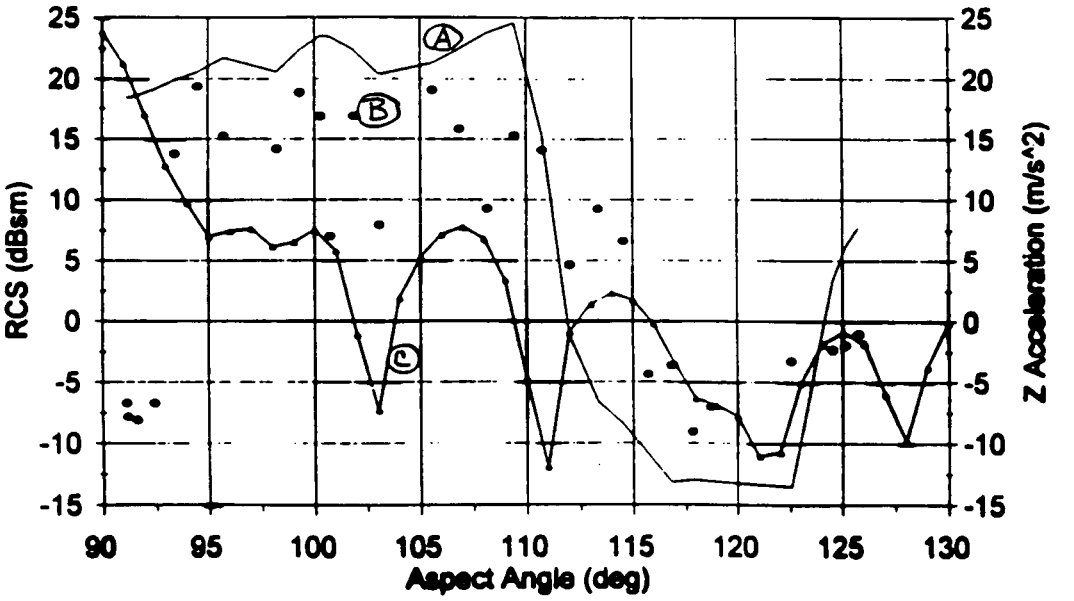


Fig. 11 Target RCS and Acceleration vs Aspect

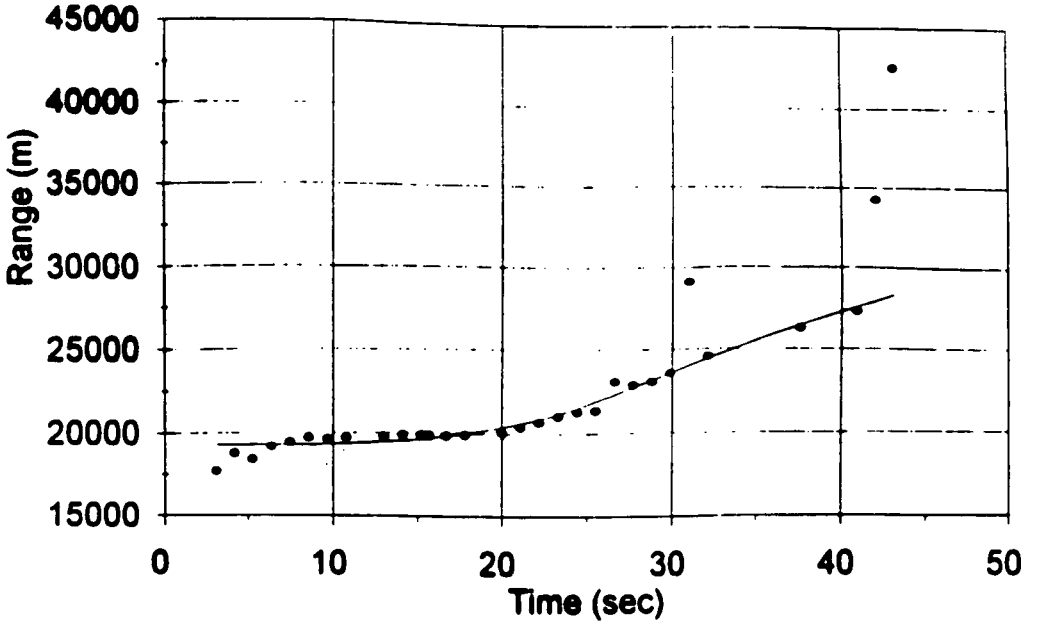


Fig. 12 Missile Target Range Profile via VHF Illuminator (U)

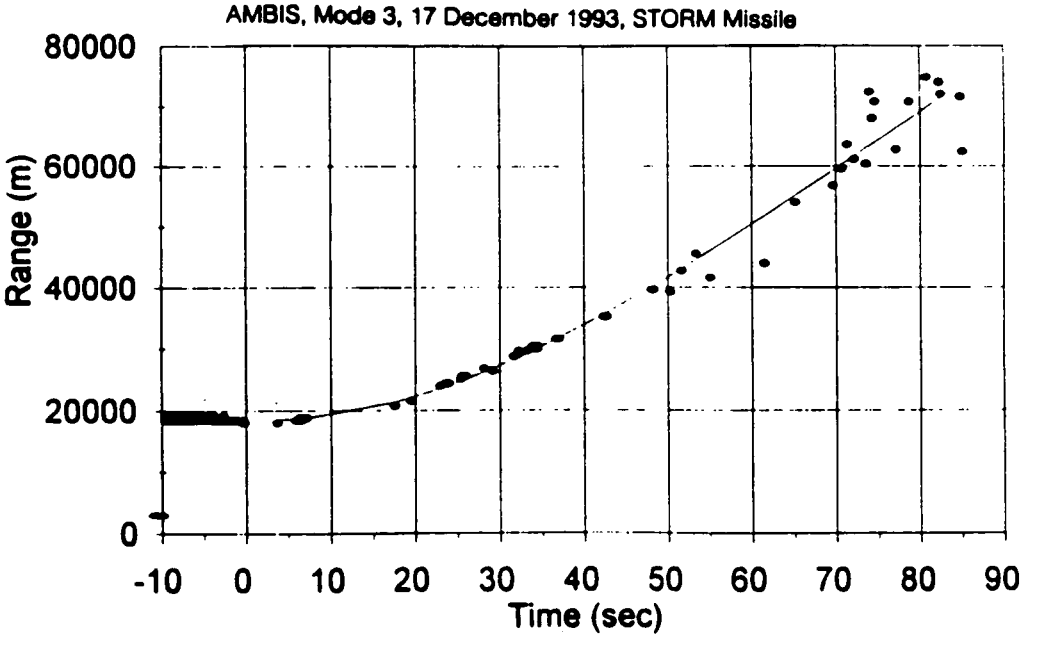


Fig. 13 Missile Target Range Profile via C-Band Illuminator

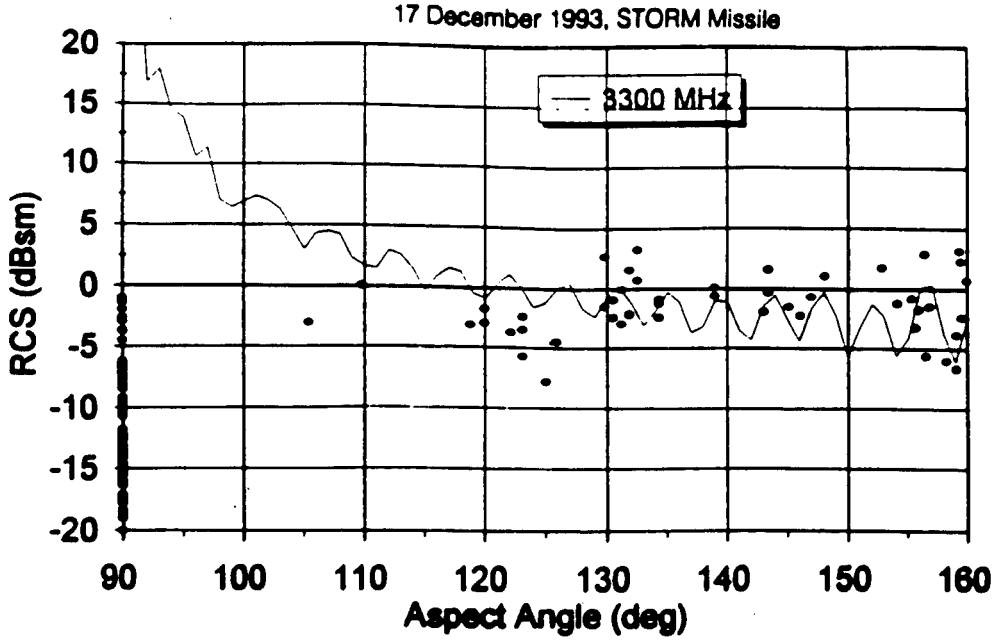


Fig. 14 Target Measured RCS vs Predicted RCS at C-Band vs Aspect

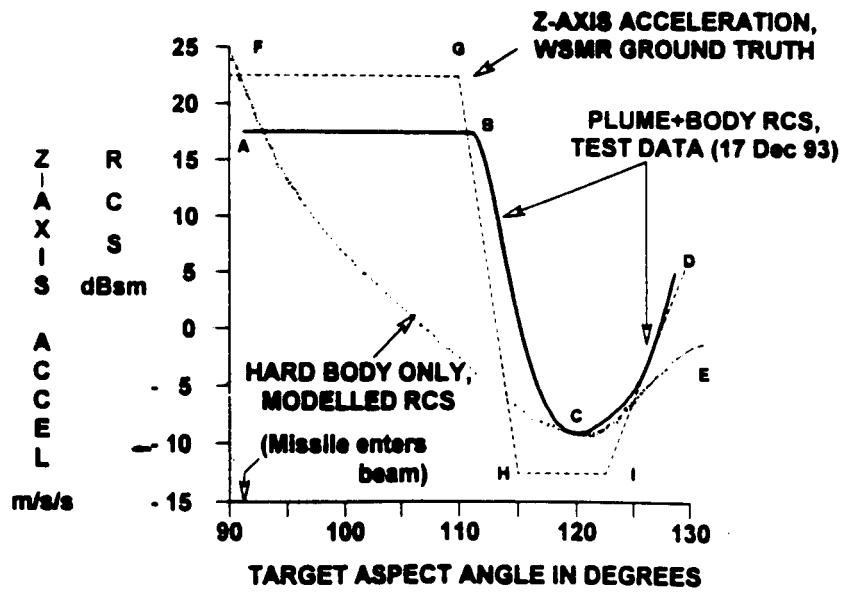


Fig. 15 RCS and Acceleration Kinematics Templates for Solid Propellant Missile Targets

INTERACTIVE INTEGRATION OF PASSIVE INFRARED AND RADAR HORIZON SURVEILLANCE SENSORS TO EXTEND ACQUISITION AND FIRM TRACK RANGES

S.R. Horman, R.A. Stapleton, K.C. Hepfer, R.M. Headley and J.K. Stapleton
Naval Surface Warfare Center, Dahlgren Division
Dahlgren, VA 22448-5000

1.0 SUMMARY

The *interactive* integration of sensors has shown great promise as a means of significantly improving the range at which ships can detect and acquire high-speed, low altitude Anti-Ship Cruise Missiles (ASCM). Modest sensors, designed for the purpose, have been developed and successfully tested in a field environment. Near-horizon clutter was measured, producing some surprises. A real time integration system has been built and successfully field tested with sensors which had most of the desired characteristics against representative targets. In addition, an unprecidented capability has been developed to measure low altitude propagation to high fidelity as a function of frequency, time, target altitude, radar antenna height and target range, and to measure dynamic infrared refraction effects. This provided a dramatic means to better understand the performance of multi-sensor systems when operated in the field under conditions that produced anomalous propagation. Improvements in firm track ranges were commensurate with our predictions. A FY95 real time demonstration is planned using a radar with an agile beam antenna.

2.0 PREFACE

The low altitude, high speed, reduced radar cross section ASCM has evolved as a means to exploit intrinsic vulnerabilities of shipboard radar surveillance systems. When a threat is at low altitude, and near the horizon, propagation losses are severe, reducing the amplitude and thus the signal to noise of the reflected return. In addition, the target is embedded in sea clutter, which displays characteristics similar to targets. The demands on radar systems are further exacerbated by target maneuverability, Electronic Countermeasures (ECM) and, in littoral environments, returns from land clutter that can appear to be at much closer ranges. Traditional solutions to these problems have been to increase radar power-aperture products (and cost) to increase the amplitude of the return and to use Moving Target Indication (MTI) or doppler waveforms and signal processing to suppress clutter. The extremely low altitudes at which current and projected ASCM can fly, coupled with other countermeasures to radar, produce a threat class that will stress affordable radar-only surveillance systems beyond both current and projected capabilities. Present ship combat systems are dependent on radar-only solutions for Anti-Aircraft Warfare (AAW) surveillance leading to engagement of ASCM. As a consequence, the late detection and track of such fast, low threats would result in greatly compressed engagement timelines, often resulting in single round intercepts at dangerously short ranges or complete loss of engagement capability using surface-to-air missile systems. For example, some ASCM may impact their target less than 30 seconds after crossing the radar horizon. The resulting requirements placed on detection time, threat evaluation and weapons assignment, missile preparation time, and missile flight time are great. Therefore, relaxation of these timelines by even a few seconds produce both a significant increase in battle space and substantially reduced demands on the entire combat system. The use

of Multi-Sensor Integration and Control (MSI&C), to make sensors interactive and mutually supportive, has demonstrated great promise in countering the low altitude threat. This performance increase results from the use of a horizon surveillance and track radar and an Infrared (IR) horizon surveillance sensor, both of modest performance, cost and weight, compared to a radar-only solution. The operative principle is to utilize the IR surveillance sensor to detect high speed, low altitude threats at or near the horizon. The radar, which also provides independent horizon surveillance, is then directed to interrogate the detected threat with a pulse-doppler waveform of approximately 12-20 dB greater energy than that of the normal surveillance waveform. This greater dwell energy provides the required signal to noise, while improved radar stability and doppler processing is required to suppress clutter. The result has often provided a dramatic increase in initial firm track range. Under fluctuating propagation conditions, a significant increase in firm track range and track continuity has also been obtained through the use of multi-sensor processing to establish and maintain tracks by combining intermittent IR and radar detections.

3.0 DESCRIPTION OF INVESTIGATIONS

The Office of Naval Research has supported an exploratory development MSI&C task by the Naval Surface Warfare Center, Dahlgren Division with the support of the Naval Research Laboratory since 1988.[1,2] This task is focused on demonstrating increasing radar firm track range through real time MSI&C of Infrared (IR) and radar horizon surveillance sensors designed for interactive operation. To differentiate this objective from the numerous other potential benefits of Multi-Sensor Integration (MSI), the task is titled Multi-Sensor Detection (MSD). The most stressing projected ASCM threats will fly at low altitude, utilize high speed and maneuverability, have reduced radar cross sections and will operate in the countermeasures environment created by standoff and self-screening jammers. Advances in radar design have produced surveillance radars which have adequate to excellent performance against mid to high altitude threats, even against the most stressing of such threats in an ECM environment. Conversely, IR sensors often have severely degraded performance at elevation angles above about one degree. This is because there is a 50% worldwide probability that a cloud may mask a threat flying above 1000 feet, and the effects of cloud clutter become pronounced above 1 degree elevation. Therefore, the MSD task was further focused to address the most stressing combination of threat characteristics and ECM environments over a narrow range of elevation angles about the horizon. This provides the potential for maximizing IR sensor performance in the region where surveillance radars need the most help, and where IR systems perform best. This focus on the horizon and the use of sensor interaction, while leaving volume surveillance to other sensors, provides the basis for projecting significant performance improvements to the combat system using affordable sensors.

Before an MSI&C system could be built, a number of fundamental questions needed to be answered through measurements and experiments. These questions concerned phenomenological as well as technological issues. The most basic were:

a. Can a modest horizon surveillance radar using special confirmation waveforms:

- Provide sufficient clutter rejection, and
- Detect a low altitude, low observable target at the required range while maintaining a realistic False Alarm Rate (FAR)?

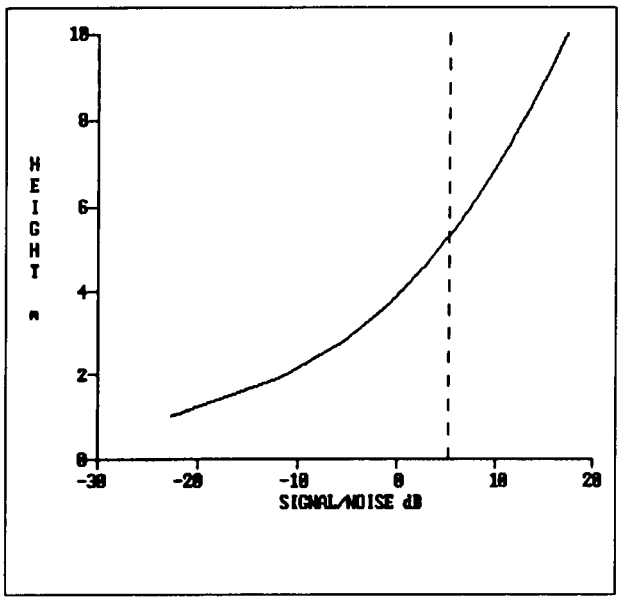


Figure 1. S/N vs. altitude at 10 nmi

b. Can an affordable IR horizon surveillance sensor:

- Detect the target at the horizon a useful percentage of the time, and is this range greater than the radar's detection range when using its standard search waveforms,
- Detect the target at or near the optical horizon while rejecting severe ambient sea clutter, and
- Detect the target against land clutter with an acceptable FAR?

c. Will the combination of IR and radar sensors, properly integrated and designed for the purpose, provide performance superior to that provided by them operating independently? These questions as well as less fundamental technology and engineering issues have shaped the planning and execution of the work described below. This paper is intended to provide a quantitative rationale for the use of MSI&C and to provide a snapshot of our present status and plans. Numerous analyses, experiments and measurements by a number of scientists and engineers are discussed below. Therefore, for purposes of clarity, the paper is organized by design drivers and associated data and analyses rather than by chronological order or individual investigations.

4.0 PHENOMENOLOGICAL DRIVERS

Volume surveillance radars provide surface combatants a critical sensing capability . However, such sensors are designed to provide long range detection and track of threats at altitude. Detection of low altitude, high speed threats at or beyond the optical horizon is needed to provide adequate combat system response time. When near-horizon threats fly at 2-5m above the sea surface, they are well embedded in the radar diffraction region. In this case, propagation is highly dependent on both target and radar altitudes as well as radar frequency and the specific environment along the path. Threat studies have shown that low, fast (Mach 3) ASCM with cross sections of -XXX to -YYY dBsm can be expected in the 2010 timeframe [3]. Figure 1 shows the signal to noise vs. target altitude that would be obtained by a ***notional*** S band volume surveillance radar for a -ZZZ dBsm missile at 10 nmi (18.52 km), under nominal propagation conditions. The shape of this function is the figure's salient feature. Signal to noise can be changed through changes in radiated power, antenna gain or other radar design parameters. On the other hand, the propagation induced loss of S/N as a function of target altitude constitutes a fundamental phenomenological limitation for any fixed radar frequency, if all other factors are held constant. The vertical dotted line denotes the signal to noise (S/N) required for this radar to detect the target under benign clutter and 5 m evaporative surface ducting conditions. In this particular case, at 10 nmi, the target could be detected at an altitude of approximately 5 m or greater. However, if the radar observed the same target flying at an altitude of 2 m, a S/N of 16.3 dB below the detection threshold would result. One

can see, therefore, that if a given missile were to fly at 2 m rather than 5m at 10 nmi, this 3 m drop in altitude would result in a reduction of this radar's S/N equal to that resulting from a 16.3 dB reduction in missile radar cross section (were the missile flying at 5 m), a very dramatic effect. One can as readily calculate the increase in radar energy on target required to extend detection range, with all other variables fixed. This provides another way to look at the problem. Figure 2 shows the propagation loss as a function of range for the same notional radar and the same target flying at 2 m. The dashed curve represents free space propagation loss, the solid line, that for a target at 2 m. The dotted line represents the radar's detection threshold. One can see that this target can first be detected by this notional radar at 8.3 nmi. To extend the detection range by 3 nmi, an additional 14 dB energy on target would be required. A 4 nmi increase would require nearly 20 dB. To increase the average power of a capable surveillance radar by 14 to 20 dB would be extremely costly. To increase the power-aperture product of the radar to provide the required S/N would be costly and would likely result in significant increases in sensor weight and size. Significantly increasing the energy on target through use of longer surveillance dwells would result in unacceptably long search frame times, given constant radar average power. However, if one were to design the radar such that it performed surveillance with a conventional waveform, and could occasionally change its waveform on command to provide a single confirmation dwell (with 12 to 20 dB greater energy than the surveillance waveform) in the direction specified by a cue, a large performance increase in that direction could be obtained with a much smaller impact on radar search frame time and power-aperture product. Naturally, this capability would only be useful if another sensor were able to reliably detect the target at or beyond the desired firm track range, and be able to provide a timely cue to an

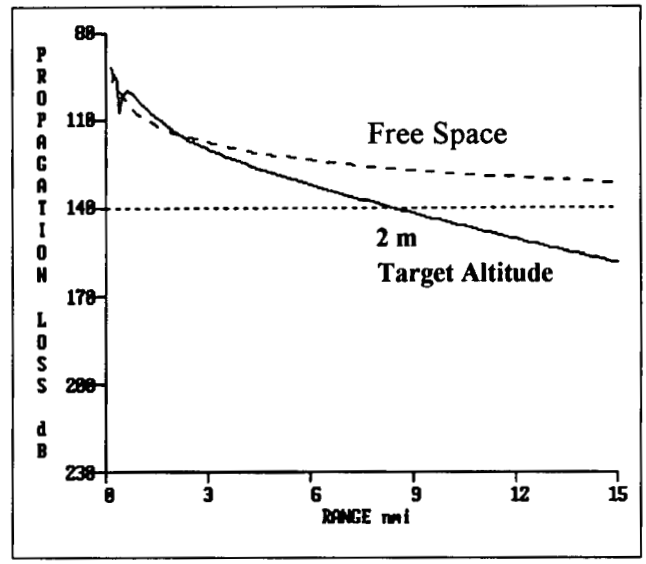


Figure 2. Propagation Loss vs. Range

accuracy of a fraction of the radar's beamwidth. In addition, the cueing sensor would have to provide false cues at a sufficiently low rate to limit the extra loading to the radar to an acceptable limit, such as 25-30% of the total energy budget per search frame time. The optimal percentage of the total energy budget will be a function of environment and the knowledge of it resident within the sensor resource manager. This question will be future addressed in future testing, and the answer and its fidelity will probably continue to evolve for years. The ability to remotely sense the environment and adapt a multi-sensor system to it for optimal operation is the focus of a new ONR task, ***Interactive Adaptation of Fire Control Sensors to the Environment***. Based on findings to date, the total allocation of energy to this purpose, and the energy used as a function of bearing will ultimately be governed by existing propagation and clutter conditions, and the confidence of the combat system in its knowledge of them. Some candidate sensors that could provide useful cues are passive IR surveillance sensors, precision electronic support measures (PESM) Radio Frequency (RF) sensors and sensors located on other platforms.

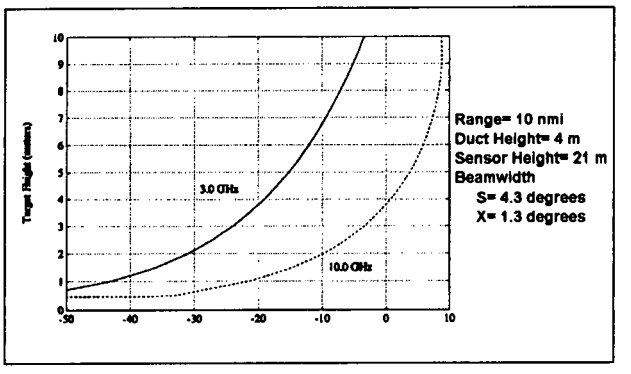


Figure 3. X vs. S Band Two Way Pattern Propagation Factor (dB), Constant Aperture

The compressed timelines associated with the high speed, low altitude ASCM require that a confirmation dwell be requested shortly after the target is detected. Therefore, an IR surveillance sensor would be of most use if it were able to provide designations to targets upon initial detection (at the contact level), rather than waiting to develop a track file. In the case of slower ASCM, with lower signatures, it may be necessary to develop a firm IR track to have sufficient confidence to request a confirmation dwell. However, since the threat is slower, under many cases, track level integration of sensors will still provide enhanced radar firm track ranges. In the case of very fast or highly maneuverable targets, the radar would also need to be able to transmit the confirmation waveform before the target elevation and bearing could move a ponderable fraction of a beamwidth. Therefore, for tactical surveillance applications, where no firm track of the target exists, the radar generating the confirmation dwell must use an agile beam antenna, such as a phased array. Such a properly designed suite of interactively integrated sensors could provide the necessary S/N to significantly increase radar firm track ranges. However, adequate S/N is a necessary but insufficient condition for establishing a firm radar track.

S/N is frequently not the dominant factor limiting radar performance against low altitude targets. The radar must have sufficient stability and signal processing capability to suppress residual clutter to the noise level to achieve the gains described above. A horizon surveillance radar must be pointed at or near the horizon, maximizing the potential for sea clutter and/or land clutter to degrade the radar's performance. One means to improve S/N against low altitude targets is to utilize a higher transmitted frequency. Figure 3 shows the difference in pattern propagation factor (PPF) between two notional radars which are identical except that one operates in S band, the other in X. The improvement in PPF, and therefore S/N, for a 2 m target at 10 nmi would, under the assumed conditions, be 22 dB. However, this improvement would come at the expense of greatly increased clutter, which would have to be suppressed to realize the benefits of the improved S/N. In addition, propagation as a function of target height and range frequently becomes highly structured in time and space as the operating band of a radar increases in frequency. Shipboard experience, as well as carefully controlled field measurements, have shown that the requirement for clutter cancellation is often the strongest driver when searching for small low targets. This is further complicated by the non-stationary nature and azimuthal anisotropy of propagation and background clutter.[4]

Supersonic ASCM can traverse 3-4 nmi in a matter of 6-8 seconds. It is reasonable to question the true value of this additional time and battlespace. This can be addressed quantitatively only through detailed analysis of every element of the combat system and how it would react and perform against specific threats and scenarios. However, the relative benefits of increased firm track range can be quickly illustrated through use of a greatly simplified model of the combat system, including the weapon, and a well behaved, but stressing target.

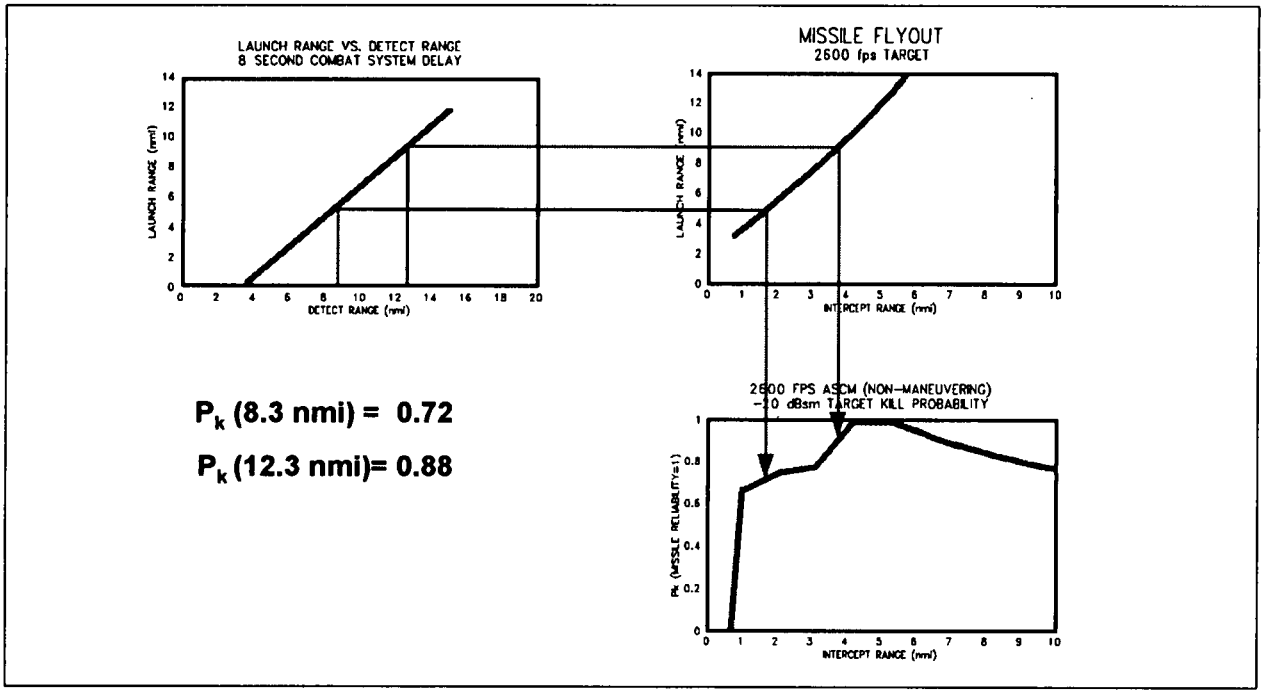


Figure 4. Notional Combat System Performance against a 2600 fps ASCM

Consider the case shown in Figure 4. The probability of kill as a function of intercept range of this notional weapon against the notional, non-maneuvering ASCM was determined though detailed Monte Carlo analysis and is outside the scope of this paper. Of primary interest is the shape of this function.

P_k (CUM) AS A FUNCTION OF FIRM TRACK RANGE (AAW MISSILE RELIABILITY = 1.0)			
FIRM TRACK RANGE= 8.3 nmi		FIRM TRACK RANGE= 12.3 nmi	
8.3 nmi	DETECT ASCM	12.3 nmi	DETECT ASCM
4.8 nmi	1st MISSILE AWAY	8.8 nmi	1st MISSILE AWAY
4.4 nmi	2nd MISSILE AWAY	8.4 nmi	2nd MISSILE AWAY
1.7 nmi	1st INTERCEPT P_k = 0.72	3.7 nmi	1st INTERCEPT P_k =0. 88
1.4 nmi	2nd INTERCEPT P_k =0.68	3.5 nmi	2nd INTERCEPT P_k = 0.85
P_k (CUM)	0.910	P_k (CUM)	0.989
P_k (4 ASCM)	0.68	P_k (4 ASCM)	0.956

Figure 5. Notional Combat System Performance against a 2600 fps ASCM

The probability of kill of a typical AAW missile falls off dramatically near minimum intercept range. Therefore it is important to take this into consideration when engagement timelines are highly compressed. In this example, a fixed combat system delay time of 8 seconds between establishment of firm track and first missile launch was assumed, along with interceptor flyout and P_k characteristics as shown. Also, optimistically assumed for simplicity, was that a shoot-shoot-look-shoot doctrine was used, all interceptors were perfectly reliable, interceptors could be launched at 1 Hz, 1 second was required to assess kill and launch the third missile, there was never a shortage of illuminator assets and attacking missiles were sufficiently spaced in time that multiple engagements did not interfere with each other through competition for combat system assets. Figure 5 compares the result of engagements where the initial firm track range was 8.3 and 12.3 nmi. In each case, the probability of raid annihilation against a single missile and 4 missile raid is shown. The difference is quite dramatic. Increasing the firm

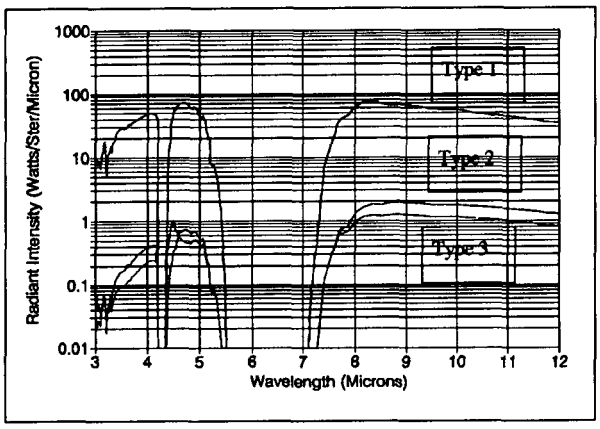


Figure 6. Spectral IR Signatures of Representative ASCM

track range from 8.3 to 12.3 nmi decreased the probability of a "leaker" from 32% to 4.4%. Removing the simplifying assumptions greatly complicate such an analysis, and generally result in lower estimates of combat system performance. However, significant benefits can be shown in most cases. The ability of a combat system to successfully counter an attack is strongly a function of the performance of every one of its elements. However, given that the combat system can engage a threat successfully, if provided sufficient warning, initial firm track range is clearly a critical driving factor. Interactive integration of IR and RF sensors is one way of achieving this extra battlespace.

Passive IR surveillance sensors have the ability to provide extremely accurate designations to supersonic targets. In addition, they perform very well near the horizon, the environment where radar performance degrades. They cannot detect threats through clouds or heavy fog and rain, however. IR sensor performance degrades under adverse weather conditions due to molecular absorption by the atmosphere and scattering by aerosols (dust, haze and fog). In addition, like RF, IR radiation is subject to refractive ray bending. This can reduce or increase the range at which low altitude targets become obscured by the horizon. Low elevation refractive effects in the IR are due primarily to temperature gradients close to the water surface. A positive air-sea temperature difference tends to extend the horizon, a negative difference moves the horizon closer to the sensor. There are benefits and limitations associated with each refractive condition. These will be discussed below. Where atmospheric extinction vs. refraction limits the range at which a given IR sensor can detect a particular ASCM, the threat's IR signature is a critical factor. Figure 6 shows the IR spectral signatures of three ASCM that roughly bracket existing threats. One can see that there are approximately three orders of magnitude variation in IR signatures between these threats. In addition, atmospheric extinction can easily vary by another three orders or more of magnitude. The dominant sources of IR emission are blackbody radiation from the ASCM skin and emissions from the plume. Careful design can limit a missile's plume signature, but operation at supersonic speeds creates substantial, unavoidable skin heating. A modest IR horizon surveillance sensor can be built that has sufficient sensitivity to detect a supersonic ASCM at the horizon, under all but the worst weather conditions. To provide this performance against very slow, low signature ASCMs would require a much more capable and costly IR sensor. However, the greatest stress to the combat system is created by high speed threats. Therefore, for purposes of the MSD exploratory development task, only supersonic or high IR signature threats have been addressed. Given the ability to detect a threat, clutter can be an even more serious limitation to IR surveillance sensor performance than it is to radar sensors, since no range resolution is available to assist in clutter rejection. The angular separation of a low altitude missile from the horizon is small. Figure 7 shows that no low altitude ASCM ever exceeds 1 mrad elevation angle above the horizon, when viewed from a typical shipboard IR surveillance sensor height, and that a large fraction of them never

exceed a few hundred μrad. For example, Figure 8 is a high resolution (88 μrad detector angular subtense) mid-IR image showing a missile-like target at an altitude of 12 feet at a range of 10 nmi. This image was gathered during testing at Wallops Island in 1992. The largest angle above the horizon observed for this target at any range was approximately 200 μrad. Inspection of Figures 8 and 9 also shows that sea clutter can be extensive and of amplitude much greater than near-horizon threats. Previous analyses and measurements showed that IR sea clutter fell off quickly near the horizon, suggesting that it might be benign at the horizon. This belief was founded on extensive previous clutter measurements with sensors of modest resolution. [5,6] It was also predicted and observed that the location of the optical horizon was often ambiguous and that targets frequently appeared above the apparent optical horizon.[7] The 1992 measurements performed at Wallops Island with the above high resolution IR sensor showed that clutter was often significant up to the horizon.[8,9] This is easily seen through inspection of Figure 9. This

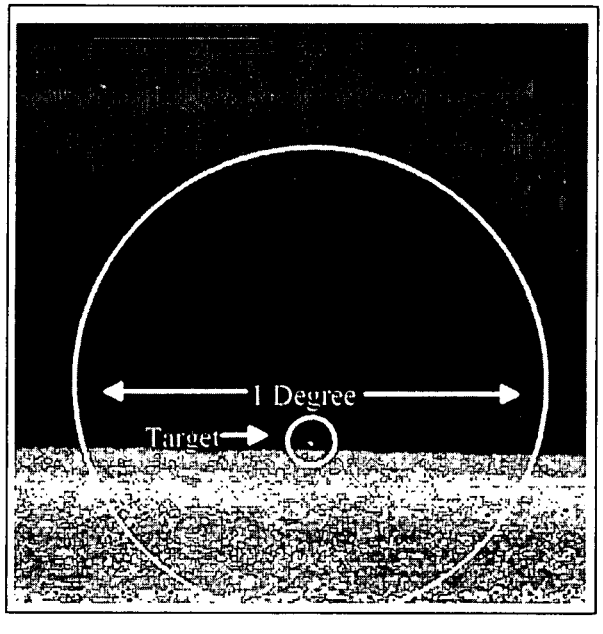


Figure 8. The Threat in its Environment

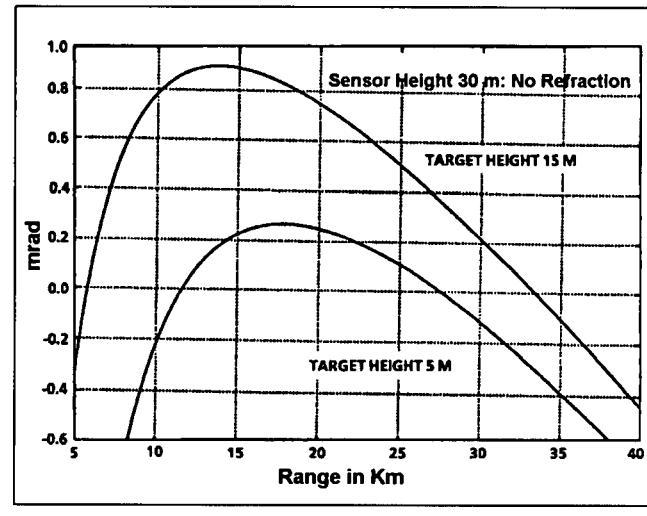


Figure 7. Threat elevation angle relative to the horizon as a function of altitude

figure shows an example of the measured standard deviation of clutter, measured in sensor A/D counts, as a function of elevation angle for a representative sea background of moderate clutter. Detector count starts at the bottom of the field of view. The relatively flat, low amplitude "clutter" observed above the horizon is actually the noise floor of the sensor. Inspection of Figures 7, 8 and 9 clearly shows that an IR horizon surveillance sensor does not require a large vertical angular coverage, but

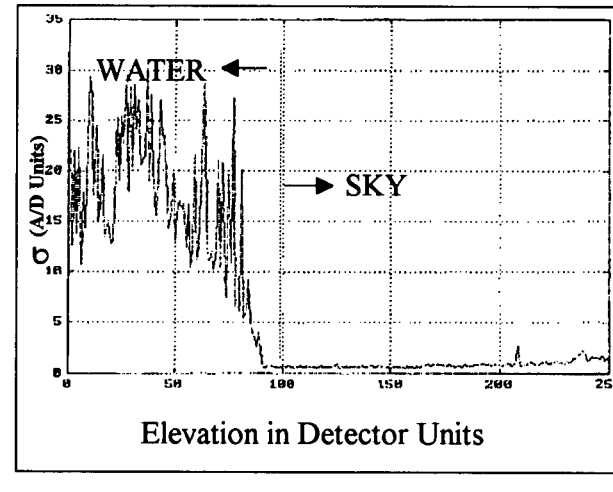


Figure 9. Background Clutter vs. Elevation Angle

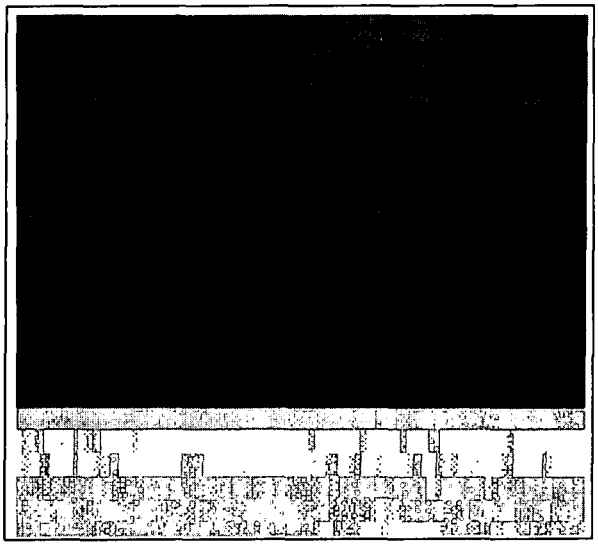


Figure 10. Low Altitude Threat as Seen by a Volume Surveillance IRST

must have high resolution to separate near-horizon threats from sea clutter. The requirement for high vertical resolution would increase a volume IR surveillance system's signal processing load as well as providing the opportunity for producing more clutter detections. Limiting vertical coverage to a small angle about the horizon is a highly effective means to reduce both demands on the sensor. To illustrate the importance of high resolution, it is useful to compare the performance of the above horizon surveillance IR sensor to a traditional IR Search and Track (IRST) volume surveillance sensor. IR volume surveillance sensors typically have detectors of large vertical angle subtense, about 1 mrad. This is driven by the need to provide 10-30 degrees of elevation coverage with a reasonable number of detectors. These detectors would necessarily couple large amounts of below horizon clutter, when it is present, with the target contribution, resulting in unacceptable performance. In addition, the unresolved target would lose contrast even in the absence of intense clutter. This effect can be demonstrated by processing the image in Figure 6 to replicate the output of a typical volume surveillance IRST. Figure 10 shows what an IRST with a detector angular subtense of 0.3 X 1.2 mrad would "see." The target is completely obscured by near-horizon clutter, and becomes undetectable, even to sophisticated signal processors, on a single look basis. The conclusion that an IR horizon surveillance sensor must have high vertical resolution is, therefore, driven by experimental data as well as theory. Such an IR sensor design is a major departure from traditional design approaches.

Using validated computer models, the effects of refraction and atmospheric transmittance on an IR sensor's performance can be calculated using the calculated or measured IR signatures of the targets of interest, a complete description of the atmospheric conditions and sea temperature and a description of the background. Although general agreement on the accuracy and precision of atmospheric transmission models has existed for some time, the effects of refraction are calculated[10,11,12,13,14] with models that sometimes disagree. Additionally, they all have serious limitations in their ability to predict refractive effects under conditions with positive air-sea temperature differences. In the past, interpretation of the statistical impact of refraction effects to IR sensor performance varied significantly within the surveillance sensor community. A substantial minority maintained that the effects of refraction would unacceptably reduce the range of IR horizon surveillance sensors a majority of the time. In addition, most of the data from which these models were derived were obtained through visible band radiation measurements. A widely held perception was that a negative air-sea temperature difference would reduce the range to the visible horizon and would therefore be "bad" for an IR sensor, and the converse would apply to positive air-sea temperature difference conditions. Although the refractive effects in the mid-IR and the visible should be nearly identical, based on theory and measured indices of

refraction, one documented experiment in the Long Wave IR (LWIR) measured the effects of refraction to have virtually no dependence of the maximum intervisibility range on air-sea temperature difference.[15] Therefore, carefully instrumented and controlled measurements of the maximum range at which a low altitude target could be seen by a shipboard IR sensor, mounted at a representative height, were necessary to develop a statistically significant body of data to validate or correct refraction models and to demonstrate the utility of IR sensors for multi-sensor horizon surveillance. This body of data necessarily required a wide range of air-sea temperature differences to provide a basis for conclusions.

The Naval Surface Warfare Center, Dahlgren Division, assisted by the Naval Research Laboratory and the Applied Physics Laboratory, performed the above referenced simultaneous IR/RF refraction and clutter measurements at Wallops Island in March-April, 1992. A target that had the mid-IR signature characteristic of a very small supersonic missile, 20 watt-sr^{-1} ,was mounted 12 feet above the water on a boat. Since the projected intensity of a Mach 3 seaskimmer is nominally 300 watt-sr^{-1}, this comparatively small target provided a substantial conservative margin for extrapolation of test

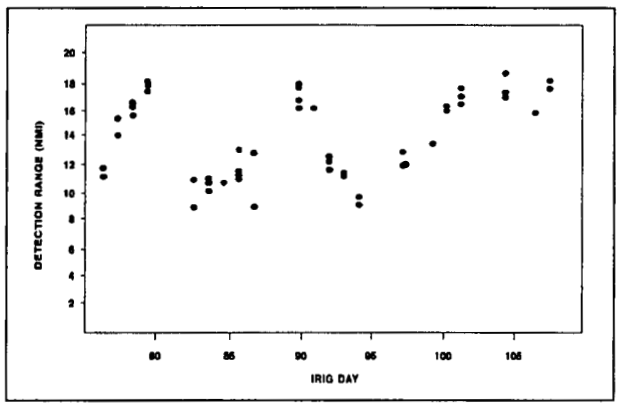


Figure 11. IR Sensor Refraction Limited Range vs. Time.

results. During IR and RF measurements of sea and near-horizon background clutter, the boat was sent outbound until the IR sensor could no longer detect it. The range at which this occurred was the Maximum InterVisibility Range (MIVR). This measurement was sometimes performed several times per day. Figure 11 shows the observed MIVR for a period of 5 weeks. During this period there were three days where the IR sensor had inadequate performance due to rain and fog. The quasi-periodic variation of the data is the result of the passage of air masses and frontal activity. Cold days resulted in negative air-sea temperature differences, warm ones trended toward positive differences. In addition, as the day progressed, and the air warmed, the range at which the boat could be detected increased since the beneficial effects of refraction moved the horizon out. Although a very wide variation of air-sea temperature difference was observed, and this range is representative of most environments, it did not cover the full range possible. Extreme refractive effects were observed by Takken, et. al. when viewing the horizon in cold weather while looking over the Gulf Stream.[16] Here the intense negative air-sea temperature difference foreshortened the horizon an unknown, but large amount. In addition, very large positive air-sea temperatures can extend the horizon so far that the target may first be observed against the sea surface.[17] This was observed a number of times during additional subsequent testing of interactively integrated sensors at Wallops Island during the period of December 1993 to April 1994. It was determined that while a negative air-sea temperature difference reduced the range to the visible horizon, the horizon was almost invariably crisp, and the target appeared above the horizon, and was therefore quite easy to detect with relatively simple signal processing. In addition, for a brief period of time after crossing the horizon, the apparent intensity of the target was enhanced by refraction. In the

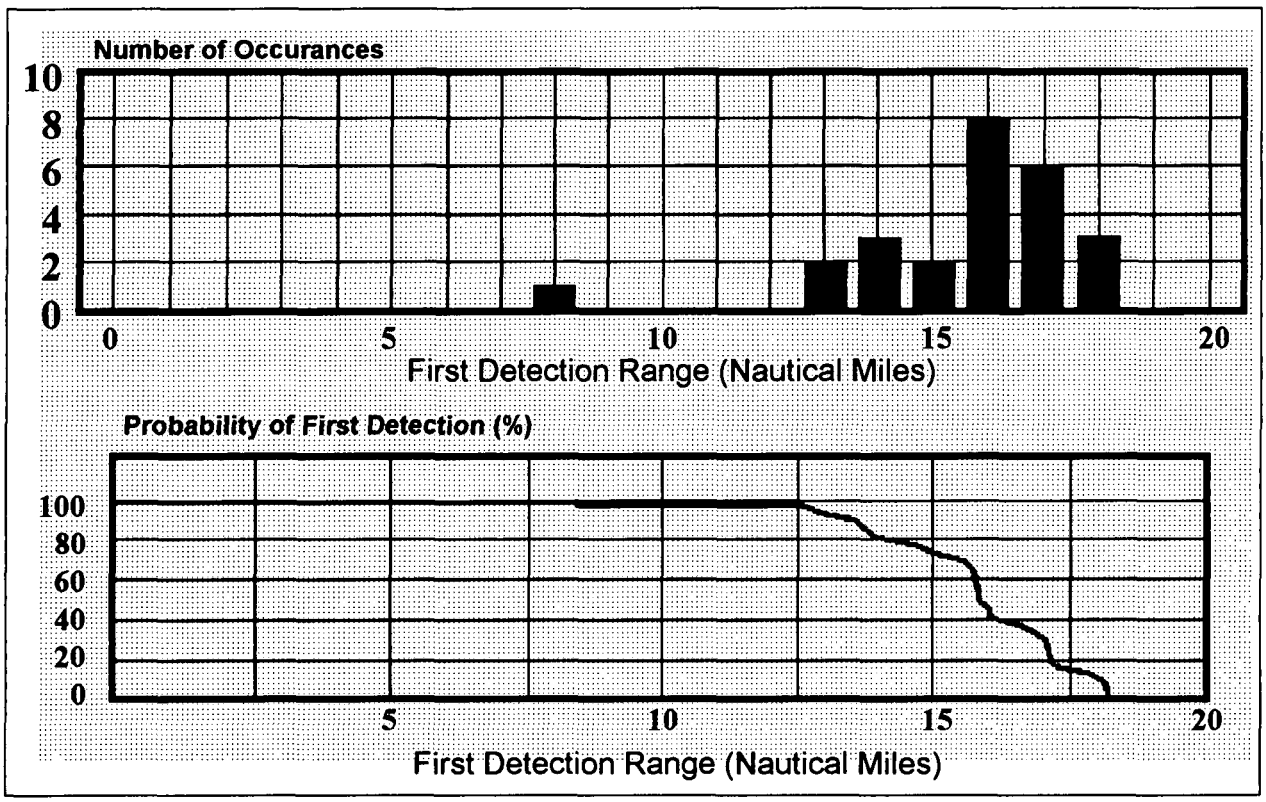


Figure 12 . Probability of Detection of a TLX Towed Target by HISS from Wallops Island Test (Comprehensive Target Metric =0.11)

case of positive air-sea temperatures, the target generally first became detectable against the sea surface, and was therefore embedded in sea clutter. The signal processing utilized included multiplicative terms associated with target elevation angle relative to the horizon, density of adjacent clutter and target spatial characteristics. The output contacts were assigned an associated Comprehensive Target Metric (CTM) based on those characteristics. With a CTM of 0.11, the IR sensor produced a false designation rate that averaged about 1 Hz, the maximum projected to be tolerable. With an adaptive CTM, the false designation rate was typically 0.01 Hz, and first detection ranges were reduced by about one nmi. The details of the relevant IR signal processing and sensor performance are the focus of several reports and papers. [18,19,20,21,22,23,24] With negative air-sea temperature differences, there was no obvious dependence of IR sensor performance on target altitude, while under positive temperature difference conditions, the detection range of the lowest targets was degraded somewhat, since they first became detectable against sea clutter. At extended ranges, the apparent intensity of the targets were reduced below what might otherwize be expected by refractive effects, and some significant mirage effects were observed, although acquisition ranges remained acceptable. A useful measure of the performance of the integrated system of sensors was obtained through testing with an aircraft towed target, the TLX. This target had a measured in-band intensity of 10 watt-sr^{-1}, and was controlled by an altimeter to fly at low altitudes. Performance of the IR sensor against the TLX Towed Target is shown on Figure 12. The IR-enhanced version of the TLX towed target is shown on Figure 13. The average range improvement obtained through interactive

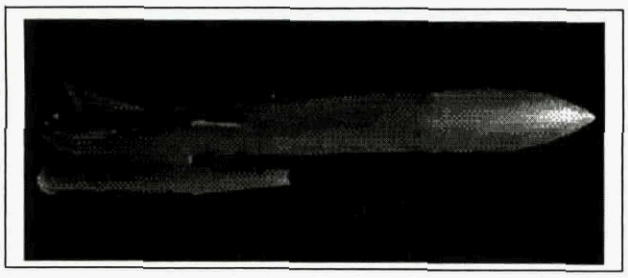


Figure 13. TLX Towed Target

integration of disparate sensors during the test was 20-25%. Figure 14 shows an example from a typical test run. For purposes of this test, a near real time comparison of the performance of the radar with and without the cued waveform was obtained through allowing the radar to accept a cue only every eight seconds. If an acquisition was made, this was indicated on Figure 14 by a plus sign (+). Once firm track was established, the system automatically terminated track and resumed operation in its normal surveillance mode. When acquisition was obtained through use of the surveillance waveform, this was indicated by an "o". One can see that in this case, the use of interactive integration produced an increase in initial firm track range of 2 miles. In other test events, the improvement ranged from 1.4 to 5 nmi, depending on propagation conditions. This is in good agreement with the theory developed previously at NSWCDD.

In summary, the effects of compressed engagement timelines, severely degraded radar performance and the demonstrated high availability of IR horizon surveillance sensors designed to compliment horizon surveillance radars provide a strong motivation for using interactive integration of these dissimilar sensors to improve combat system performance.[25,26] Both analytical prediction and measurement of RF and IR propagation and clutter support the use of sensors designed for the purpose. It has been shown above that the IR surveillance sensor must have high resolution, and need not have a large vertical field of regard. The radar must use an agile beam antenna and be designed to be interactive to support the generation of confirmation dwells as well. These requirements and others create the engineering challenges of MSI&C.

5.0 ENGINEERING DRIVERS

Phenomenological drivers and target characteristics constrain the design of the

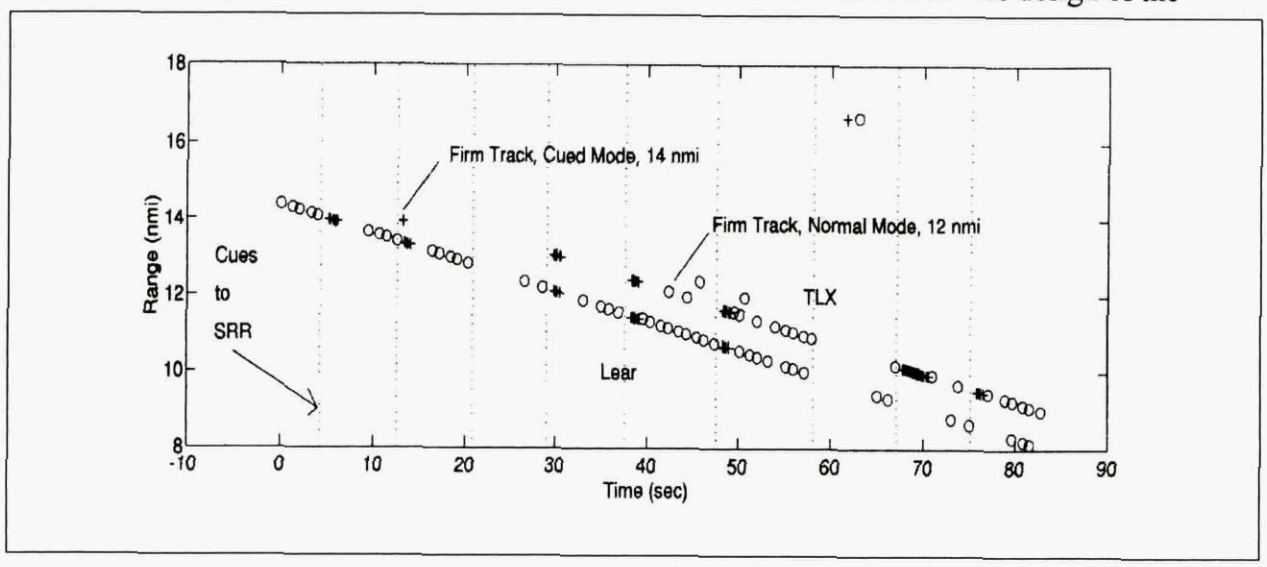


Figure 14. Firm Track Range Improvement Through Interactive Integration of Sensors

sensors. Other system level drivers further constrain the sensors and the MSI&C system which controls them and processes their data. Although the potential for improvement in combat system performance through interactive sensor integration is clear, abundant evidence exists that an incorrect selection of sensors or implementation of the integration can result in no improvement or even significant degradation in overall system performance.[27]

The integration of multiple sensors is widespread on U.S. Navy ships. When integration of sensor data is performed, it is usually performed at the track level. In other cases, integration may occur as a result of a voice or teletype communication link. Virtually all contact level integration is performed using data from similar sensors. The integration of dissimilar sensors is complicated by a significant number of factors. The most important are:

a. Dissimilar Information. A radar may produce 2-D or 3-D detections and tracks, i.e. bearing and range or range, bearing and elevation. IR surveillance sensors do not measure range, only bearing and elevation. Therefore, correlation and association of contacts is complicated by a limited number of common observables. An unavoidable consequence of the integration of dissimilar data is the potential for association (and the formation of a single declared track) of different objects, objects with clutter, or of disparate clutter. Any MSI&C system must cope with these fundamental issues. Failure to do so can result in greatly increased system loading and/or inappropriate engagement decisions.

b. Dissimilar Information Quality. The precision with which disparate sensors perform measurements significantly impacts the integration process. The resolution of a horizon surveillance sensor may be on the order of 100 μrad, that of a horizon surveillance radar is at best several milliradians. This mismatch results in an association window that is dominated by the radar's limitations.

c. Static and Dynamic Misalignment. Inaccurate alignment of sensors can result in improper associations, track degradation and improper direction of confirmation dwells. Static misalignment is a function of the installation and periodic alignment process. Dynamic misalignment results from improper stabilization, ship flexure, gyro inaccuracy and sensor artifacts. When any of these unavoidable factors become large enough to create significant beam shape losses during the confirmation dwell, system performance will degrade, unless active compensation techniques are employed. Unless sensors are mounted in close proximity, and on a stiffened portion of the superstructure, ship flexure alone can exceed acceptable limits.[28,29,30,31]

d. Dissimilar Detection Ranges. The primary benefit of interaction is derived from exploiting the dissimilar ranges of disparate sensors in specific environments. However, this mismatch can create problems as well. Against low altitude, high speed targets, the detection range of a horizon surveillance IR sensor may greatly exceed that of the radar, even when a confirmation waveform is used. This can be exacerbated by anomalous propagation and ECM. One challenge to the MSI&C system is deciding when (and how long) to ignore an apparently high priority threat that is persistent, but cannot be detected by the radar. When multiple, similar candidate threats exist, a process to determine how long to wait before attempting another confirmation dwell against the same contact is needed. The details of this process have not been determined to date analytically or empirically. An additional tradeoff is required for this situation, and that is the allocation of energy to the confirmation

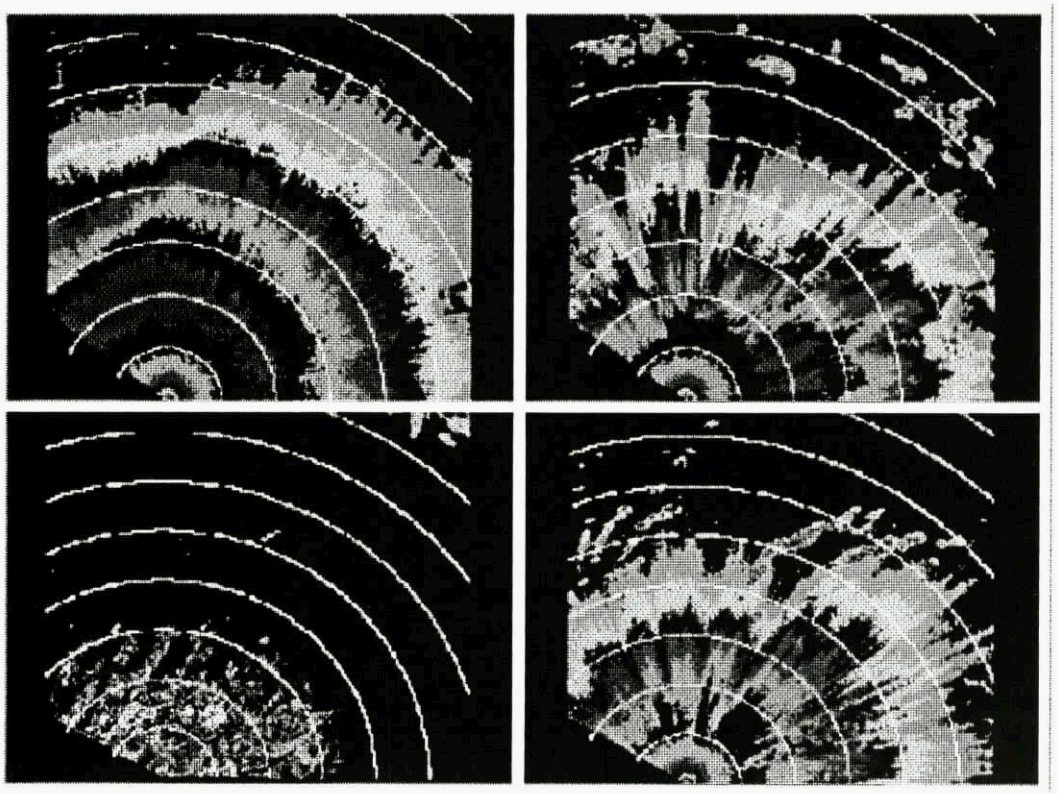


Figure 15. Typical variations in clutter and propagation.

dwell. It may, under some circumstances, prove better to reduce the energy in each confirmation dwell and increase the rate of interrogations.

e. Sensor Resource Management. The potential benefits of MSI&C come at a price. In addition to the additional system complexity required to support interaction, interactive operation will result in degradation of at least the radar's performance as a stand-alone sensor. For example, every confirmation dwell places demands on the total time/energy budget of the radar. Therefore, some time and energy that could be dedicated to stand-alone operation, must be reserved for interactive support to the MSI&C system. The optimum balance of design for stand-alone operation vs. operation to optimize overall combat system performance has not yet been established. It will almost certainly vary with environmental conditions, the state of combat alertness and the existence and location of threat sectors. Under many conditions, propagation and clutter will vary widely with bearing.[32] Examples of the extreme variability of propagation and clutter at sea can be seen in Figure 15. This figure shows four azimuth scans with a powerful radar at Kwajalein. Each range ring corresponds to 32 KM, and each gray scale corresponds to 5 dB change in volume backscatter corrected for range. Conditions ranged from well behaved ducting at the upper left, to virtually no ducting on the lower left, to the nearly chaotic variability seen on the right. Therefore, the use of different surveillance and confirmation waveforms as a function of bearing will be necessary to achieve optimum performance. To achieve this, the combat system must either sense, or be provided, a detailed clutter and propagation map for each sensor. This information would then be used to determine the optimum surveillance, track and

confirmation waveforms as a function of bearing. Adaptive resource management will be a major focus in the ultimate design of any MSI&C system.

f. Data Latency. Data latency, and the similar effects of delayed sensor response, can degrade the system by creating confirmation dwell beam shape losses. During the delay between the time that the IR sensor detects the target and a confirmation dwell is transmitted, target angular motion and uncompensated ship flexure will result in a misalignment of the radar beam and the target. In addition, since the primary means of obtaining increased energy in the confirmation dwell is to increase its duration over a surveillance dwell, the target may become further misaligned with the beam during the period of transmission. In practical terms, the total latency budget to be allocated to IR sensor response time, data transmission, MSI&C processing and radar response runs from tens to a few hundreds of milliseconds, depending on the threat.

6.0 SYSTEM DESIGN

The requirements for low latency, coupled with the large amount of data that must be processed has driven the architecture of the system presently being developed at NSWCDD. A high level diagram of this architecture is shown on Figure 16. Data is transmitted to and from the sensors by a commercially available fiber optic local area network (LAN) that utilizes replicated memory. This LAN, SCRAMNET, provides low data latency and this latency is deterministic, simplifying algorithm design. Data processing is performed on both 68040 processor boards and i860 programmable vector processors, all mounted on a VME backplane. The i860 processors perform the required vector calculations at about 200 times the speed of the 68000 processors. Rapid communication between the i860 processors is obtained through use of a separate VSB bus, thereby reducing the loading of the VME backplane. Standard workstations connected to the 68000 boards via ETHERNET are used for control and displays. At present there is significant room for growth on the backplane. The open architecture and design of the software will allow later use of additional backplanes linked via the SCRAMNET LAN if required.

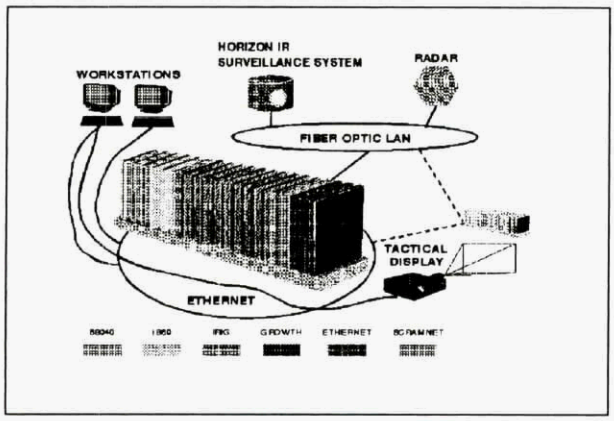


Figure 16. MSD System Architecture

Since the focus of the MSD task is to increase available battlespace through extending the radar's firm track range, the multi-sensor track filters utilized are only sufficient to minimize inappropriate track and contact associations. However, since there is a significant mismatch in radar and IR sensor resolution, there is only one strong common observable, bearing. Therefore, a multi-hypothesis test is iterated on all tracks to prune out inappropriate associations. The computer program architecture is shown at a high level in Figure 17. Initiation of cues utilizing individual high priority contacts produces the maximum obtainable performance. When this is possible, the data flow runs from the upper left to the lower right, resulting in confirmation dwells as well as surveillance and track dwells. However, this is not always possible, due to environment or the existence of numerous simultaneous high probability contacts. Contacts not associated with active tracks are passed into the multi-sensor, multi-hypothesis track initiation and cue request

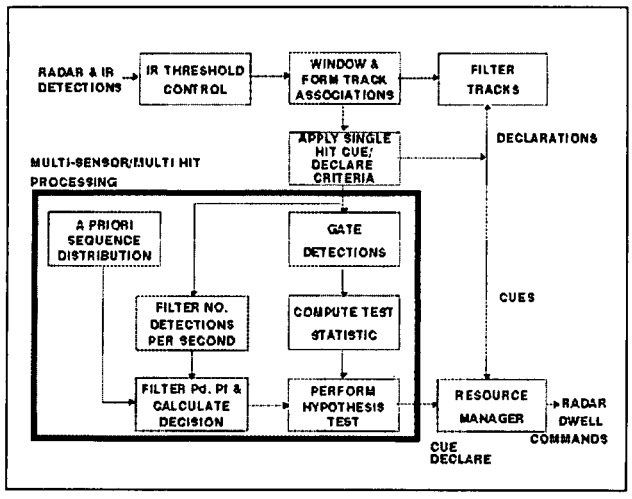


Figure 17. MSD Computer Program Architecture

processor, here shown enclosed by a bold rectangle. This processor can initiate tracks directly from multi-sensor contact data or recommend confirmation dwells to the sensor resource manager.

7.0 STATUS AND PLANS

All the programs have been tested against both real and simulated multi-sensor data, in the lab and through real time interactively integrated sensor field testing. They have performed well. The open architecture shown in Figure 17 readily allows the insertion of recorded data and/or sensor simulators into the system, greatly simplifying computer program development and testing. All interfaces are designed to allow direct substitution of sensor emulators for sensors, and to substitute recorded data where needed. In the latter case, sensor interaction is lost. All initial real time coding was completed and tested against recorded data in July, 1993 . A complete real time system interfaced to a horizon surveillance radar and to an IR horizon surveillance sensor was tested against representative targets at Wallops Island between December, 1993 and April, 1994. The range was heavily instrumented with ground truth equipment to provide better understanding of the test results and to provide a basis for extrapolation to other systems. Along with a robust suite of meteorological instrumentation, provisions for detailed measurements of radar and IR propagation was in place. In addition, a differential Global Positioning System (GPS) was utilized to localize targets to high accuracy and precision.

8.0 SUMMARY

Significant advances have been made in the development and application of interactive Multi-Sensor Integration (MSI) technologies to counter the high-speed, low altitude Anti-Ship Cruise Missile (ASCM). Special pulse-doppler waveforms that can detect and track such threats at tactically useful ranges have been developed and demonstrated in a marine environment using a radar of modest power-aperture product. The capability of a high resolution IR surveillance sensor to reliably detect such threats at required ranges has also been demonstrated in a marine environment. Simultaneous measurements of sea clutter using the above sensors have been made, and initial representative data reduced to determine spatial dependencies and RF/IR clutter correlation. The algorithms that implement single contact MSI&C have been tested with both simulated and real multi-sensor data in real time on a real time system. The more generalized algorithms required to perform MSI&C have been developed and demonstrated in a simulated environment. They are presently being converted to real time code and baselined on the multi-sensor detection processor developed in-house for the purpose. System integration testing began in July, 1993 and a real time demonstration with representative sensors and targets is scheduled for the latter part of FY95.

9.0 REFERENCES

1. Task 5WFCS08, of the Shipboard Fire Control/Mission Planning Project, sponsored by ONR351.
2. Trunk, G.V., *et. al.*, "Radar-IR Cueing" ,

Proceedings of the 5th National Symposium on Sensor Fusion, *Vol II*, April, 1992.
3. Henderson, T.C., *et. al.*, *Future Airborne Antiship Threat*, NSWCDD/TR-92/55, September, 1992.
4. Queen, J.L., *et. al.*, *High-Powered Radar Collection of Ocean Backscatter Data: Kwajalein Atoll, Marshall Islands*, NSWC TR-88-241, December, 1988.
5. Kessler, B.V., *et. al.* "New IRAMMP Dual-Band Radiometric Sensor", *IRIA-IRIS Proceedings of the 1989 Targets, Backgrounds and Discrimination Meeting*, 1989.
6. Takken, E.H., *et. al*, "IR Horizon Phenomenology: Report on IRAMMP Field Test", *IRIA-IRIS Proceedings of 1991 Targets Backgrounds and Discrimination Meeting*, 1991.
7. Priest, et. al, *High Resolution Imagery of the Sea Horizon: Data and Modeling*, NRL Technical Report (In Preparation), 1992.
8. Hepfer, K.C., *et. al.*, *Horizon Infrared Surveillance Sensor (HISS) Data Analysis Results*, NSWCDD Technical Report (In Preparation), 1993.
9. Takken, E.H., *et. al.*, *Analysis of Ocean Horizon Measurements Acquired with a Staring InSb IR Focal Plane Array*, NRL Technical Report (In Preparation), 1993.
10. Yates, H.W., *Atmospheric Refraction over the Water*, NRL Report 4786, 1956.
11. Paulus, R.A., *Validation of the Bulk Method for Overwater Optical Refractivity*, NOSC white paper.
12. Beaulieu, A.J., *Atmospheric Refraction Model and the Effects of Surface Waves*, DREV Report 4661/92, 1992.
13. Freiesleben, H.C., "Investigations into the Dip of the Horizon", *Journal of the Institute of Navigation*, Volume 3, Number 3, July, 1950.
14. Larsen, R. "Propagation of Infrared Radiation in the Evaporative Duct", *Proceedings Of The Sixth International Conference On Antennas And Propagation (ICAP89), Part 2: Propagation*, April, 1989.
15. Feinberg, R. and Hughes, H.J., "Marine Boundary Layer Refractive Effects in the Infrared", *Applied Optics*, Vol 18, number 15, August 1, 1979.
16. Takken, E. H., *et. al.*, *op.cit.*.
17. Hepfer, K.C., *et. al.*, *op. cit.*
18. Trahan, J.W., "Infrared Refraction and Mirages; Wallops Island, Feb-Apr 1994", *Proceedings of the IRIS Symposium on Targets, Backgrounds and Discrimination*, 31 January, 1995.
19. Trahan, J.W., *Infrared Refraction and Mirages; Wallops Island, Feb-Apr 1994*, NSWCDD MP 94/365, January, 1995.
20. Hepfer, K.C., "False Alarm Analysis (Preliminary), Horizon Infrared Surveillance Sensor; Analysis of Test Data Collected at NSWCDD, Wallops Island Detachment, January-April 1994", *Proceedings of the IRIS Symposium on Targets, Backgrounds and Discrimination*, 31 January, 1995.
21. Hepfer, K.C., *False Alarm Analysis (Preliminary), Horizon Infrared Surveillance Sensor; Analysis of Test Data Collected at NSWCDD, Wallops Island Detachment, January-April 1994*, NSWCDD MP-94/271, September, 1994.
22. Dezeeuw, P.A., "Detection Range Performance, Horizon Infrared Surveillance Sensor; Analysis of Test Data Collected at NSWCDD/Wallops Island Detachment, November 1993 through April 1994," *Proceedings of the IRIS Symposium on Passive Sensors*, 13 March, 1995.
23. Hepfer, K.C., *False Alarm Analysis (Preliminary), Horizon Infrared Surveillance Sensor; Analysis of Test Data Collected at NSWCDD, Wallops Island Detachment, January-April 1994*, NSWCDD MP-94/363, January, 1995.
24. Hepfer, K.C., *et. al.*, "Horizon Infrared Surveillance Sensor," Proceedings of the IRIS Specialty Group on Infrared Countermeasures, 27 April, 1994.
25. Hepfer, K.C., *et. al.*, "Horizon Infrared

Surveillance Sensor, Current Status and Future Plans," *Proceedings of the 1994 National Infrared Information Symposium*, 25 May, 1994.
26. Vastag, B.V., *et. al.*, *NATO Anti-Air Warfare Critical Experiment 101A-- Phase II Test Report*, NAVSWC TR 89-451, June 1990.
27. *Multi-Sensor Integration Critical Experiment 3-2-101A Final Report*, APL Report NAAW-89-225, FS-89-049, April 1989.
28. DeWoody, R.T., *High Speed Turn Hull Flexure Test On Hull No. 4226 (DD-988), Builder Trials*, Report No. R 5384-160, February 1980, Ingalls Shipbuilding Division, Litton Systems, Pascagoula, Miss.
29. Sikora, J.P. and Dinsenbacher, A.L., *Progress Report On Structural Flexure Measurements For Proposed Destroyer Trials*, SD 79-1730-40, Jan 1979, DTNSRDC, Bethesda, Md.
30. Greco, D.A., *Tartar Tracking Loop Study*, Contract N00024-82-C-5212, February 1985, General Electric, Pittsfield, Ma.
31. Schulze, A., *Small Ship AAW IR&D Task Ship Flexure*, *#209/41-0576*, January 1991, GE Aerospace, Moorestown, NJ.
32. Queen,J.L.,*et. al.*, *op. cit.*

The Fusion of an IR Search and Track With Existing Sensors To Provide a Tracking System for Low Observability Targets

P.V. COATES
Pilkington Thorn Optronics
(Formerly THORN EMI Electronics)
120 Blyth Road
Hayes
UB3 1DL
England

SUMMARY

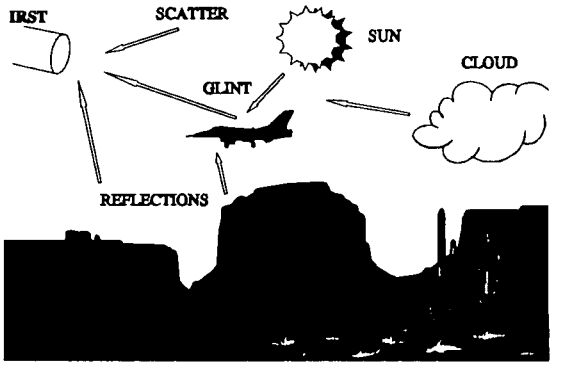
This chapter covers the topic of the application of multi-sensor fusion, in a modern fighter aircraft, with the objective to improve the performance of pointing and tracking systems. It deals specifically with various techniques, including the use of other sensor ,data to improve the performance of a passive IR Search and Track (IRST) system to provide an enhanced tracking solution for targets at all ranges, from low observability long range stealth targets to short range pop-up targets.

The timing is particularly appropriate in that Pilkington Thorn Optronics is the Technical lead partner in the European International Consortium responsible for the development of the IRST/FLIR system [Ref 1] for Eurofighter 2000 (European Fighter Aircraft). It is one of the first airborne passive Electro Optic (EO) detection systems, in Development, capable of simultaneously detecting and accurately tracking multiple targets. It will also have a passive ranging capability. It will be the first time that such a high performance passive target acquisition and tracking system will be available in a modern fighter aircraft. The availability of such systems presents a unique opportunity to combine this information with other on-board sensing systems to obtain a performance "force multiplier".

1. INTRODUCTION

The IRST is a relatively new sensor system which can provide an airborne platform with an increased capability in roles such as surveillance, target acquisition and gun fire control. Its intrinsic high spatial resolution is a vital requirement to interface with modern armaments, such as; to cue hit to kill weapons, to point directed countermeasure systems to extremely high line of sight accuracies and to allow reliable passive target engagements at extremely long ranges.

An IRST sensor gathers information by repeatedly scanning a wide field of regard, typically 150° x 60°, by means of a multi-element infrared detector. The detector measures the radiated energy that is naturally emitted by all objects. The radiated energy covers a frequency spectrum, the peak of which is dependant upon the temperature of the object, in accord with Plank's law [ref 2]. Additional energy will come from sources such as conduction and reflection. Atmospheric attenuation limits the use of the infrared waveband to two "windows" at 3-5μm and 8-11μm.



As indicated in the picture above, the total energy incident on the IRST detector from any external object is a complex combination of components from many sources. These components are not just related to the object temperature, but varying with its surface texture, its surface attitude and its surroundings. If the object is a distant aircraft then it will probably be viewed against a sky background, which might contain clouds and the likelihood that it will appear to radiate the same energy as its background is very remote. The apparent temperature difference between the target object and its surroundings is the signal that the IRST system can exploit.

It is very difficult, if not impossible, to employ stealth techniques to prevent an aircraft from being "seen" by an IRST system. Unlike radar, it is not relying on detecting its own returning emitted radiations, which could feasibly be attenuated or dissipated by stealth techniques. Also, unlike radar, if it were possible to

totally eliminate all radiations from the target, it would actually present a **more detectable** signature, since it would appear as a "black hole" against the natural background emissions. The only way for a target to prevent itself from being seen would be to maintain its level of emissions at **precisely that of its background as viewed by the IRST sensor**, clearly an impossible task. Even if it were possible to invent a method of rapidly varying an aircraft IR signature, it would not be known at what level to set it to exactly match its own apparent background. The IRST viewed background might be sky, cloud or even ground features and may well be many kilometres beyond the target.

In addition, the IRST system's ability to detect the target is not degraded by any of the ECM devices traditionally used to defeat radar systems.

Thus as a passive target detection system, the IRST has many operational advantages. However, it also has a number of performance limitations which usually limit its long range detection capability, particularly with adverse weather conditions.

A modern aircraft platform will contain many sensors, operating in different wavebands. It is logical to consider the use of information from these to improve the performance of the IRST system. From a theoretical viewpoint it would be ideal to consider the fusion of all sensor data in some "central processor" which could optimally combine all information to ensure the best possible utilisation of sensor assets. However, such a scheme requires both considerable processing power and the ability to distribute vast volumes of high speed data.

A radical change in both component technology and avionic system architectures is required before this ideal data fusion system can be fully realised.

This paper concentrates on the use of other sensor data within the realistic constraints of available avionic architectures and data distribution systems to give a true picture of currently achievable performance.

The paper is broadly divided into two parts; the first dealing with the need for sensor fusion as a performance multiplier and the second describing its potential application, in a modern fighter aircraft, to improve the performance of an IRST system.

2. THE NEED FOR SENSOR FUSION

2.1. Introduction

The rapid advances in both weapon and platform technology have lead to the requirement to be able to detect the enemy platform at increasing ranges.

The likely combat environment of tactical aircraft in the near future is likely to be increasingly more hostile, particularly with regard to the increasing numbers of hostiles and their improving capabilities. It is generally assumed that the future pilot will often be faced with situations where he is numerically outnumbered, by at least two to one. Given this situation, it is of increasing importance that emphasis be placed on the ability to detect and engage the enemy at Beyond Visual Range (BVR) where the benefits of superior weapon delivery systems can be exploited.

In addition, since close engagements Within Visual Range (WVR) are traditionally decided in favour of the force with the numerical advantage, it is of increasing importance that systems are developed that can rapidly acquire and accurately designate medium/close range targets to provide the tactical "edge" to combat the increased numbers of hostiles.

As well as the tactical issues , there are the technical issues resulting from the increased capability of hostile forces. Some of the more significant of these are covered in the following sections;

2.2 Stealth Technology

Over recent years there has been a significant advance in the techniques of signature reduction. By the judicious use of materials, coatings and structural design, this new "stealth" technology has resulted in a considerable reduction in the observability of all types of targets in all wavebands. Recently, particular emphasis has been placed on reducing the equivalent radar reflecting areas of targets.

However, this technology is not simple and is not cheap to implement. Its effectiveness has been demonstrated at radar frequencies and in selected EO wavebands, but the achievement of a target which is completely stealthy across the whole electromagnetic spectrum is not a practicable proposition within the foreseeable future.

Hence, the more that effective use can be made of target emissions across wider regions of the electromagnetic spectrum, through the use of multi-sensor fusion, the more detectable the target will be.

2.3 The Modern Battlefield

The effect of a loss in detection range is made worse by both the longer engagement ranges achieved by new missile systems and the increase in the speed with which modern battles are fought. New generation missiles will contain highly sensitive multiwaveband sensors and sophisticated target detection and tracking systems enabling them to autonomously acquire targets, after launch, detect them at long ranges and track them through evasive manoeuvre and countermeasures.

These factors lead to the requirement to maintain (or even improve on) the long range target detection and tracking capability, when operating against targets of low observability.

2.4 Sensor Technology Limitations

There are fundamental and physical limitations to the design of individual sensors which prevents the achievement of the large performance improvements necessary, for an individual sensor to reliably detect and track a low observability target at a strategically significant range. For example, whilst the achievement of a very low aircraft radar cross-section has been shown to be technically viable, increasing the performance of the radar system, to regain the required detection range, presents significant practical, and technical, problems.

Similarly, limitations in detector technology and available optical apertures will prevent significant performance increases in individual EO sensors.

2.5 The Benefits of Sensor Fusion

The required step increase in detection performance will not come from either significant technical developments in existing sensors nor from the discovery of revolutionary new sensors. It will lie in the more efficient use of the sensors that already exist. The combination, or fusion of data from several sensors has the potential not only to greatly increase the overall target detection performance but has the potential to provide many other additional benefits, such as:

- EXTENDED COVERAGE FROM MORE SENSORS
- IMPROVED CONFIDENCE IN DETECTION AND DECISION MAKING
- IMPROVED DETECTION AND TRACKING
- IMPROVED RELIABILTY THROUGH MULTIPLE REDUNDANCY
- INCREASED COUNTERMEASURE RESISTANCE
- IMPROVED ALL-WEATHER PERFORMANCE
- BETTER PERFORMANCE AGAINST STEALTH TARGETS

3. CURRENT LIMITATIONS TO THE APPLICATION OF SENSOR FUSION

As stated in Section 2, there are clear benefits to be derived from the application of sensor fusion techniques.

However, although the potential for such performance improvements is clear, the practical achievement of them, in the near future, is limited by a number of practical problems. The more significant of these are discussed in the following sections.

3.1 Sensor Design

It is usually the case that individual sensors are designed separately, generally without fusion in mind. Each sensor is designed to exploit its own waveband and usually has its own dedicated signal processing chain, its own dedicated aperture and is often separately mounted, independently of other sensors.

Ideally, to allow the benefits of sensor fusion to be fully realised, consideration must be given, at the very outset to issues such as; co-location and/or shared apertures of sensor systems and the exploitation of information synergy to allow the fusion of data within the processing chains of individual sensor systems.

3.2 Data Rates

Considering some typical sensors, such as IRST, Radar and ESM, the front-end video signal is usually at a very high data rate. However, within this raw video will lie all of the relevant information that the sensor system can produce. Hence, the availability of this data, at the point of any combined sensor fusion operation, would ensure the least loss of information. However, in practical terms this option is not very attractive. The volume of data from one sensor, at this level, is very high, that from two or three sensors would be excessively high, typically in the Gbits/s region. Moreover, at the raw video level, little commonality will exist between the structure of the data from the various sources and the fusion process would be extremely difficult to implement.

It is generally considered to be both more practical and more effective to combine data at an information level further down the processing chain of each sensor, where the type of information being combined is likely to have a greater degree of commonality. However, it is essential to trade off the reduced complexity of the fusion process, when applied to higher level sensor data, against the inevitable loss of information within the individual sensor processing systems.

Thus, it is clear that the achievement of effective sensor fusion will require the distribution and combination of a wide range of significant volumes of sensor data.

3.3 Avionic System Architectures

Consideration of the requirements of sensor data distribution, as outlined in Section 3.2, lead to the inevitable conclusion that a radical improvement in avionic system architectures is necessary before the full benefits of sensor fusion can be realised.

Current data highways and distribution systems severely

limit the quantity, variety and rate of data distribution. Additionally, they also restrict the common access to high speed data and signal processing hardware.

3.4 Summary

It is clear that avionic platform system architectures and the associated hardware and software implementations need to be designed with sensor fusion in mind from the outset.

The radical changes required include; improved parallel data distribution systems and distributed, reconfigurable processing hardware and software. A recommended approach, for future systems, is outlined in [ref 3]. This will only be achieved in the long term.

However, in the shorter term, some significant benefits can still be obtained by the use of sensor fusion technology, even though it must be implemented within the constraints of the platform architecture.

Section 4 considers the design of a modern IRST system and outlines the benefits and performance improvements that can be obtained by the fusion of data from other sensors, implemented within the defined practical architecture constraints.

4. SENSOR FUSION IN A MODERN IRST

4.1 Introduction

This section considers the detection and tracking processes employed in a modern IRST and shows how the application of fusion with other information can significantly improve the acquisition of targets.

There are significant operational advantages to be obtained from a sensor that can passively locate targets at very long ranges.

There is increasing emphasis on being able to engage targets Beyond Visual Range (BVR) and many of the future weapon systems will need accurate and reliable BVR target detection and tracking information.

However, there are also equally important requirements to detect targets appearing at medium/close ranges and to accurately track nearby targets for close-in engagements.

It is possible to use an IRST system very effectively in both situations. But it is important to note that the processing requirements for distant target detection are quite different from that of close-in encounters.

The following sections consider both the long range and medium to short range target acquisition processing in an IRST system.

4.2 Long Range Detection and Tracking

At long ranges, the target size and its differential signature will be very small and a detection system with high thermal and spatial resolution is required to reliably locate it.

Typically an IRST system will have an instantaneous field of view (IFOV), or resolution cell, of better than 1mrad. In search mode it operates in a non-imaging wide angle scan mode, much like a Plot-While-Scan radar and targets will usually subtend less than the IFOV. As with all detection systems the IRST requires a gathering lens with sufficient aperture to resolve the target.

Table 4.1 shows the range at which a 2m x 2m target (head-on aircraft) matches the IFOV, for typical IRST resolutions. It also shows the required aperture for a Modulation Transfer Function (MTF) of 0.75, in both IR windows [Ref 4].

RESOLUTION (MRAD)	RANGE FOR SUB-PIXEL RESOLUTION	8-11 SYSTEM APERTURE	3-5 SYSTEM APERTURE
2.00	1.00 km	12.00mm	5.00mm
1.00	2.00km	24.00mm	10.00mm
0.75	2.70km	32.00mm	13.00mm
0.50	4.00km	48.00mm	19.00mm
0.25	8.00km	96.00mm	39.00mm
0.10	20.00km	240.0mm	96.00mm

TABLE 4.1 Limiting Pixel Resolution Range

Clearly, the higher the resolution, the better the range performance, but the greater the aperture needed to achieve it.

There is an additional problem in that it is desired to scan as wide an overall Field of Regard (FOR) as possible in as short a time as possible, to achieve good performance with a moving host platform. Ideally, a 150° x 60° FOR should be covered in at least 1 second.

Table 4.2 shows the output video data rate for various IFOV values, assuming a 150° x 60° FOR scanned in 1 second, at maximum efficiency. It can be seen that with current data transmission limitations, the achievement of a very low IFOV is not possible.

Thus, practical design limitations on aperture size and scanning the required wide FOR mean that a head-on aircraft target, at long range, must be detected when at sub-resolution.

SENSOR IFOV	VIDEO DATA RATE
2.00 mrad	0.69 MHz
1.00 mrad	2.74 MHz
0.75 mrad	4.87 MHz
0.50 mrad	10.97 MHz
0.25 mrad	43.87 MHz
0.10 mrad	274 MHz

TABLE 4.2 IRST Video Data Rates

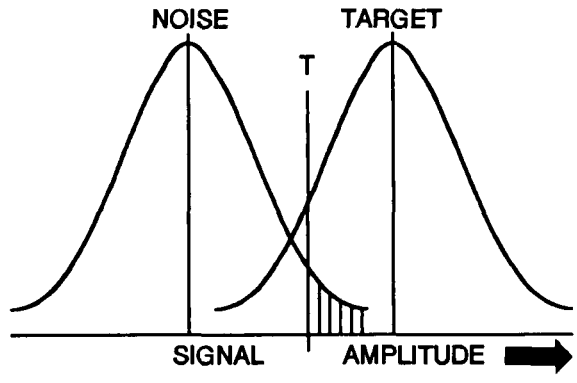


FIGURE 4.1 Target Detection in Noise

The problem resolves into the detection of a consistent signal relating to a specific object in the scanned field. This signal will usually be mixed with both random noise and with structured background "clutter".

Much work has been done on this aspect, generally by taking the work done by the Radar processing community as a starting point and adapting those principles to IRST systems [Ref 5].

This work has concerned the detection and subsequent tracking of small signals. The basic objective is to partition the sensor data into sets of observations, obtained over a period of time, derived from the same source.

The following are the basic elements of an IRST tracking system:

1. Detection
2. Correlation
3. Filtering/prediction

The following sections consider these in turn:

4.2.1 *Detection Processing*

For the purposes of calculation, it is generally assumed that for a small point size, long range target, viewed against a high altitude sky background, then both the noise and the target signals can be approximated to a Guassian distribution.

Example noise and target distributions are illustrated in Figure 4.1.

The separation of these distributions is the target signal to noise ratio.

The detection operation usually consists of setting a threshold level T such that when the signal amplitude is greater than T a target is declared present.

The area under the target distribution, to the right of T, represents the probability of target detection, the area of the target distribution to the left represents the low observability targets lost. The area under the noise distribution to the right of T, shown hatched, represents the probability of false alarm.

It can be seen that, for a given signal to noise ratio, there is a trade-off, dependant upon the setting of T, between the probability of detection and the false alarm rate.

The threshold level is usually fixed by the desire to limit the number of false alarms (FA). This is a crucial parameter and a high number of FA can render an IRST system useless.

The minimum level of threshold is determined by the Threshold to Noise Ratio (TNR);

$$TNR = T/\sigma_n$$

where σ_n = standard deviation of the noise

The TNR value is set for a given value of False Alarm Rate (FAR).

The operating Signal to Noise Ratio (SNR) is the separation of the two distributions and is fixed by the required Probability of Detection (Pdet);

such that:

$$SNR = TNR + \Omega s$$

where the value of Ωs is set by the desired Pdet.

For instance:

for Pdet = 90%, Ωs = 1.28

Consider the example of the host aircraft, with IRST system, and target aircraft, on a co-altitude approach path. The IRST system and target details are assumed to be as follows;

IRST System Details

Host platform altitude 6km
Field of Regard 150° x 60°
Scan time 1.0 second
Instantaneous FOV 0.25 mrad
Aperture 100 mm
Wavebands 8-11μm and 3-5μm
NETD 0.013° (8-11μm) and 0.05° (3-5μm)
D*pk 6.0 e10 @ 7.5-10μm
D*pk 1.5 e11 @ 3.4-5μm

Target Details

Target Altitude 6km
Equivalent radiating area 2m x 2m (head on Aircraft) @ M0.9 and M2.0
Emissivity 100%
Background radiation is mid latitude summer @ 261K

NB It is assumed that no exhaust plume or hot engine parts are visible and the only radiation is due to the target skin temperature rise. A mid latitude summer temperature and absorber vertical profile is assumed with ground level visibility of 23km. Target skin temperature rise is assumed proportional to (Mach No)^2.

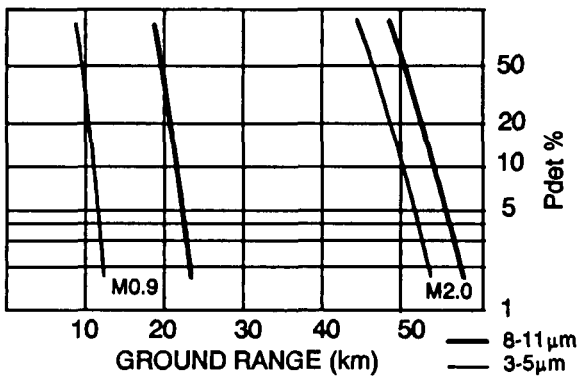


FIGURE 4.2 Detection Range vs. Pdet

Figure 4.2 shows the detection range performance with varying Probability of detection, assuming a FAR of 1/hour. The performance for both 3-5μm and 8-11μm wavebands is shown from which it can be seen that, for low observability targets, the 8-11μm band has a superior range performance under the defined conditions. Two speeds of target aircraft have been assumed. At M2.0, the increased aerodynamic heating gives a greater detection range and there is less difference between the performance in the two wavebands.

The simple thresholding operation used discriminates on amplitude only and assumes that the target is among the "hottest" signals. In the presence of large areas of high emission background, the threshold will have to be raised to preserve a low false alarm rate and small dim targets will not be detected.

The detectability of small targets is improved by the use of a 2-Dimensional Matched Filter which will reject large, intense areas and will optimise the small point target to background noise ratio, as follows;

$$H(f) = \frac{S^*(f)}{|N(f)|^2} \qquad .....(4.1)$$

where: $H(f)$ = Filter Transfer Function
$S^*(f)$ = Target Frequency
$|N(f)|^2$ = Background Noise Power Spectrum

In the spatial domain;

$$ACF(x)*h(x) = S(x) \qquad .....(4.2)$$

where the autocorrelation function ACF(x) is the Fourier Transform of the background noise power spectrum which is convolved with S(x) to give the impulse response of the matched filter

Equation (4.2) is a continuous function whereas, in practice, the signal will be digitised and hence, in sampled data form.

$$[H] = [S].[ACF]^{-1} \qquad .....(4.3)$$

[H] is the set of spatial filter tap weights and for a sub-resolution target, the target signal [S] will be the total effective point spread function of the IRST system.

It can be seen that the background noise is an important consideration in obtaining the optimum matched filter.

In imaging sensors the background will often be highly structured and will vary considerably over the sensor field of regard. Hence, optimum performance cannot be achieved over the whole field with a single filter. Thus, in practice, an adaptive filter is used which analyses the local background statistics and selects the co-efficients [H] (equation 4.3) for the optimum local filter. Thus, as the FOR is scanned, in real-time the filter statistics are altered to best extract the target from its immediate background signal.

This process has been described previously, by THORN EMI (now Pilkington Thorn Optronics), for use in IRST systems [Ref 5].

However, there is still a limit on detection range performance as a result of the need to set a high threshold value to limit the number of false alarms. Random noise will not be spatially correlated whereas targets at long range will maintain a spatial consistence from scan to scan.

This can be exploited to further reduce the noise detections by the process of correlation.

4.2.2 *Correlation Processing*

The IRST sensor gathers information by repeatedly scanning its FOV and the detection operation is performed anew for each scan, producing a stream of detections.

The purpose of correlation is to decide whether an observation (detection) belongs to an established track or whether it is a new target. In the case that the observation belongs to an existing track then an assignment is made and that track file is updated by the new observation characteristics.

The assignment process traditionally uses gating procedures to eliminate the more unlikely observation-to-track pairings. The gate size is chosen based on the anticipated relative target/sensor sightline velocity and the sensor scan update rate. Circular gates would be optimum for equal target manoeuvrability in azimuth and elevation, ellipsoidal gates could be used for non equal target velocities but, in general, it is computationally more efficient to use rectangular gates.

The gating system is fine for situations when object spacing and frequency is such that a single observation falls within a single track gate. However, in a dense object environment conflict situations can occur. For instance, an observation might fall within the gates of multiple target tracks or multiple observations might fall within the gate of a single track. This leads to the requirement for additional logic to address the conflicts. The simplest solution is to use Nearest Neighbour (NN) logic. In conflict situations, the closest observation to the predicted position is used to update the track. This is clearly subject to error.

It is possible that the observation was due to system noise. In this case a gate is established, but the track would be unlikely to be propagated by further observations. However, it is possible for genuine tracks to have missed detections, especially at long detection ranges.

A more modern approach uses *a posteriori* probabilities, calculated using Bayes' Rule, for Hypotheses concerning the observed data. These are used as the *a priori* probabilities for use on receipt of the next data set. Thus as each subsequent data set is received, the probabilities for each hypothesis, for a given data association, is updated. Genuine tracks will establish high probability values whilst false assignments will generate low values as the logical progression of the track is not continued. In effect, this operation postpones difficult decisions, allowing the use of later measurements to resolve association conflicts.

Reid [Ref 6] has presented a Multiple Hypothesis Tracking (MHT) structure, known as Reid's Algorithm to perform this process.

This is shown diagrammatically, in Figure 4.3, as a "Multiple Hypothesis Tree".

The notation is as follows;

FA = Observation is a false alarm

NT1 = Observation is the first detection of a new track number 1.

T1 = Observation is associated with existing track number 1.

The first scan (data set 1) is assumed to have two detections $y_1(1)$ and $y_2(1)$. The hypotheses are derived as follows;

On receipt of $y_1(1)$, in the absence of any other data, it can only be either a false alarm (FA) or the first observation of a new track (NT1). Thus the first two hypotheses are;

$$H_1\text{: } y_1(1) = \text{FA} \qquad H_2\text{: } y_1(1) = \text{NT1}$$

On receipt of the second observation, further hypotheses are generated. It is possible that the two previous hypotheses are false alarms, thus H_1 and H_2 become;

$$H_1\text{: } y_1(1) = \text{FA};\ y_2(1) = \text{FA}$$
$$H_2\text{: } y_1(1) = \text{NT1};\ y_2(1) = \text{FA}$$

It is assumed that a single target produces no more than one observation per scan hence, there is no possibility, at this stage, for an association with an existing track. Thus the only two remaining hypotheses are that new tracks are started, as follows;

$$H_3\text{: } y_1(1) = \text{FA};\ y_2(1) = \text{NT2}$$
$$H_4\text{: } y_1(1) = \text{NT1};\ y_2(1) = \text{NT2}$$

Thus after the first scan, with only two observations,

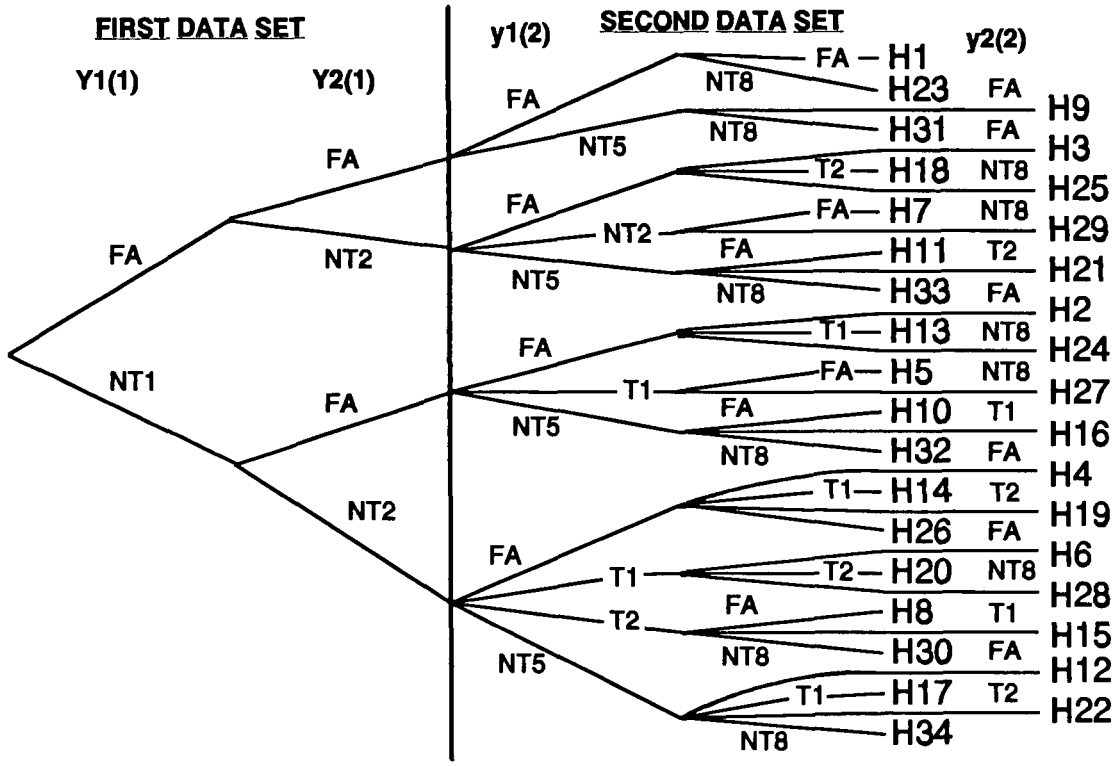


FIGURE 4.3 Multiple Hypothesis Tree

there are four hypotheses.

After the second scan, with two further observations, additional hypotheses are added, giving a total of 34 hypotheses.

It can be seen that the same track will appear in more than one hypothesis and it is necessary to cross reference which hypotheses contain what tracks and which tracks are contained in what hypotheses.

The addition of only one more data set with two further observations, increases the number of hypotheses to over 500.

Clearly, in any realistic target scenario, after a few scans, the escalation of hypotheses would lead to an unacceptably high computational load.

Also, if it is required to detect targets at very long ranges, then, as has already been noted, it is necessary to operate at very low detection thresholds, which will result in larger numbers of FA's thus exacerbating the problem.

However, there are methods for pruning and combining hypotheses to allow for a reasonable implementation.

For instance, if it is assumed that the primary aim is to detect a low contrast, distant target, then specific rules can be generated to eliminate hypotheses, as follows;

Assuming an (earth axis) stabilised sightline, then distant targets will exhibit very low (or zero) sightline motion and subsequent genuine target observations will be highly spatially correlated.

Thus a gating technique can be used to restrict the window size, specifically for the case of distant target detection. For this case all hypotheses falling outside the gate distance are rejected.

In addition, for each scan, the probabilities associated with each hypothesis are recalculated. The probabilities for each hypothesis, which fall inside the gate, can be weighted based on a distance function.

In operation, a 7x7 pixel gate is defined, centred on each observation, with Bell Function probability weights

as shown in Figure 4.4.

0.1	0.2	0.2	0.2	0.2	0.2	0.1
0.2	0.4	0.4	0.4	0.4	0.4	0.2
0.2	0.4	0.8	0.8	0.8	0.4	0.2
0.2	0.4	0.8	1.0	0.8	0.4	0.2
0.2	0.4	0.8	0.8	0.8	0.4	0.2
0.2	0.4	0.4	0.4	0.4	0.4	0.2
0.1	0.2	0.2	0.2	0.2	0.2	0.1

FIGURE 4.4 Hypothesis Probability Weights

In subsequent scans, hypotheses are generated, for association possibilities with all new observations within the gate **only** and for each of these, the hypothesis probability is weighted according to its pixel position.

Additionally, it is assumed that only one target track will exist in the detection gate. Thus the hypotheses are limited to two as follows;

H_1: Observation comes from a dim target
H_2: Observation is a FA

In some cases, it can be further assumed that the target is near stationary within the stabilisation accuracy of the IRST scanning system.

Limitations on currently available processing power prevent these processes from being applied over the whole of the IRST FOV.

4.2.3 *Filtering/Prediction Processing*
The practical application of these techniques requires the system to be cued as to the likely local regions within which distant targets might appear. This information may come from other on board sensors, or from external intelligence or from knowledge of likely threat trajectories.

In practice, the distant target location processing described in Section 4.2.2 would be carried out in conjunction with more conventional association and tracking techniques. The assumption of a near stationary target is only valid for the long range detection, sightline stabilised, scenario. It is necessary to have a tracking system that is also capable of dealing with manoeuvring targets and sightlines.

This is accomplished by more conventional techniques employing wider tracking gates and filtering/prediction processing. The problem of associating detections is different for wider gate sizes and manoeuvring targets. To obtain a high tracking accuracy with a manoeuvring target, it is necessary to make intelligent use of the past track history to optimise the parameters of the tracking system. This makes it easier to resolve conflict situations and reduce the association problems.

Filtering and prediction methods are used to estimate present and future target kinematic quantities such as position, velocity and acceleration. They are usually of the "fading memory" type, which can be implemented recursively. Data received in the past are included in the present estimates, and therefore all data are utilised, but forgotten at an exponential rate.

There are two common approaches, using fixed coefficients and Kalman filtering.

The Fixed Coefficient filters have the advantage of ease of implementation and use fixed parameters for the filter gains.

The most common of these is the α-ß Filter. This is used when only position measurements are available. α defines the position prediction measurement gain and ß defines the velocity prediction measurement gain.

An extension of this approach includes an estimate of acceleration. This filter hypothesises constant acceleration. Thus it will follow a ramp input with no steady state error.

The choice of the gain terms depends on the response required from the filter. Decreased values will lead to a less responsive filter. The choice of gains must reflect an overall compromise between noise and dynamic (manoeuvre) performance. Initially the coefficients would be chosen to ensure steady state noise performance assuming the target to be on a straight line trajectory. A movement detector could be used to determine when the target departs from the straight line trajectory and increase the tracking coefficients accordingly to follow the target manoeuvre.

In a complex multi-target situation, better performance can frequently be obtained by the use of the Kalman Filter [Ref 7]. This is the general solution to the recursive, minimised mean square estimation problem within the class of linear estimators.

The use of the Kalman Filter will minimise the mean squared error so long as the target dynamics and the measurement noise are accurately modelled. The problem with using the Kalman filter in an IRST system is that without range information, it is essentially a 2-Dimensional "bearings only" situation and problems can arise when using the Kalman Filter [Ref 8].

In operation, the Kalman Filter uses a sequence of observations of the co-ordinate position to initialise and maintain both the estimated state vector and covariance matrix, both of which are usually updated at each scan for the various tracks. The absence of range information can lead to conflicts with nearby or crossing targets, which can lead to erratic tracking, or in extreme cases, results in covariance collapse and solution divergence. Additionally, certain observation conditions can result in poor initialization performance.

These problems often lead to the Kalman Filter solution being rejected in favour of simpler, though less effective, tracking techniques.

The use of data from other sensors, particularly those with range information, can allow the reliable use of Kalman Filter techniques in IRST systems. This will result in increased track accuracy and reliability.

The predicted positions from the tracking filter are used to define the regions for association on the next scan. The size of these regions will depend on the anticipated target sightline motion. They may initially be quite large, with no knowledge of track statistics, but can usually be reduced, on subsequent scans as more track history becomes available. This process acts to reduce the residual FAR over a number of scans, allowing the initial detection operation to be performed at a low threshold level thus enabling more distant target detection, as follows;

$$FAR = N(Pfa)^n W_1 W_2 W_3 \ldots\ldots W_n$$

where; N = Pixels per scan
Pfa = Probability of FA
W_n = Window size at scan n

To further improve the detection range performance, techniques are being considered that involve processing prior to thresholding. Some of these techniques require prior knowledge as to likely target location to reduce the processing complexity. It is also possible to use them to verify an initial detection.

One possibility is to integrate the local pixel values from a number of scans before the thresholding operation. Clearly, for this to be successful, all pixels must be exactly spatially correlated, which, given the limits of stabilisation accuracy, is not likely to be achievable directly. However, the processing to remove the structured background false alarms (described in Section 4.3) generates a background scene "clutter map" which can be used to electrically correct for sightline drift. Since the clutter map consists of many features and pixels across the whole FOV, an integrated sightline stabilisation accuracy of better than one pixel, in the short term, is achievable.

However, the target may still vary marginally in position as a result of small movements and the fact that it may not exactly line up with the scanned pixel "grid". Its peak signal might vary between adjacent pixels. Thus is necessary to allow for a ± 1 pixel positional drift variation of target position. This is achieved by allowing for the integration of every possible combination of adjacent pixels during the integration operation.

It is assumed that the target can either stay on its current pixel, or move one pixel in any direction. Hence there are 9 possibilities at each scan.

The combinations increase rapidly with the number of scans of integration, at the rate 9^n, data rates and computational limitations restrict the area and number of scans over which this can be performed.

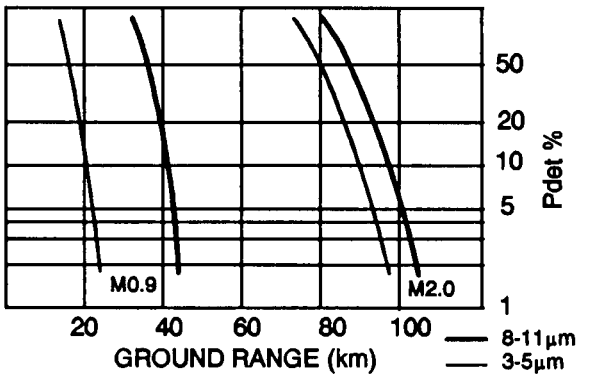


FIGURE 4.5 Long Range Target Detection

Figure 4.5 shows the calculated performance of the IRST system of Figure 4.2 assuming the use of MHT dim target processing, tracking over 4 scans and pre-threshold integration over 4 scans.

It can be seen that a significant increase in long range detection performance is obtained compared with that of the traditional processing method.

However, it does need both information as to the local area in which the target will lie and a considerable increase in processing power.

4.3 Medium/Short Range Target Detection and Tracking

The medium/short range detection and tracking problem is quite different from the longer range requirements, considered above. In particular, the target will usually present a high thermal signal level and the low observability processing will not be needed. Indeed the application of some the processing techniques described in Section 4.2 would actually **reject** a close-in target, subtending a number of pixels, in that it would be

regarded as an extended source and be removed by the 2-Dimensional matched filter. It is possible to employ a larger area matched filter which would have some performance in detecting bigger objects, but it would certainly reduce its capability at rejecting extended background clutter when looking for distant targets.

Thus for optimum performance it is necessary to employ a separate processing technique to deal with medium/close range targets.

Experimental results from the THORN EMI (now Pilkington Thorn Optronics) "Air Defence Alerting Device" (ADAD) [Ref 9], which is a ground based IRST system, currently in service with the British army, together with extensive simulations, have indicated that with a hypothetical system having an instantaneous FOV of about 0.25 mrad, good performance can be obtained, using the low observability matched filter processing, employing a 5x5 mask, down to ranges in the general region of 1-2km. Below these ranges, reliable detection requires specific processing algorithms directed at detecting resolved objects and obtaining information about them.

The major difference between acquiring and tracking medium/close range targets as opposed to long range, low observability ones, is that they are resolved and will comprise a cluster of pixels, often with some detail within, dependant upon range and scenario. There is thus a far wider range of possibilities for the form of the target video and it may change quite rapidly during tracking. The techniques to be adopted can be categorised in terms of range, or target size.

Consider the scenario of a target unmasking at medium range, and closing. At first detection, such a target might consist of a cluster of pixels with some contrast with respect to its immediate background. However, it's one distinctive feature is that it will comprise a closed object, newly entering the scene. A technique that has been found to be particularly effective in locating such targets is that of edge detection and closed contour processing. The aim is to extract "edge contours", or local regions of large temperature slope. The result of the edge detection process will show some discontinuities within the edges and edge contour logic is applied to "join up" related edges based on associating adjacent edges by means of comparing their direction, intensity and shape. The edge detection process defines edges as amplitude discontinuities between different local regions of the IR scanned scene. It makes use of 3x3 masks, convolved with the scene video. However, instead of using standard symmetric tap weights, which are equally responsive to any edge direction, directional mask operators are used.

For instance, consider the "x" and "y" gradient masks W_x and W_y below;

$$Wx = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \qquad Wy = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

W_x will give a maximum response to vertical intensity changes, whilst W_y will give a maximum response to horizontal intensity changes.

Prewitt [Ref 10] has shown that it is possible to generate a simple set of compass gradient masks by rotating Wx and Wy.

In practice they are rotated through the eight principle directions on the compass grid, as shown in Figure 4.6.

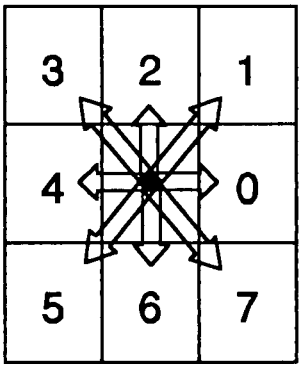


FIGURE 4.6 The Eight principle Compass Directions

The output from each of the eight direction gradient masks is evaluated for each edge point.

Comparison of the gradient mask giving the maximum output, allows any edge point to be "coded" in terms of its orientation.

In practice, a real IRST video signal will contain noise which can cause edge jitter when using these simple masks.

Work carried out on trials IR video signals has shown that better performance can be achieved, in a noisy signal environment, by the use of alternative masks.

The improved set of compass gradient masks can be obtained by rotating the functions M_x an M_y below;

$$Mx = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \qquad My = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

These masks approximate to the partial derivatives in the "x" and "y" directions respectively, and the zero weights in the centre assist in suppressing the jitter on the line where an edge might occur.

Thus, rotating the partial derivative mask functions will produce the eight directional gradient masks corresponding to the eight compass points. These are shown in Figure 4.7.

$$\begin{array}{cccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} & \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix} & \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} & \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} \\ \mathbf{4} & \mathbf{5} & \mathbf{6} & \mathbf{7} \\ \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} & \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} & \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \end{array}$$

FIGURE 4.7 8 Compass Gradient masks

In operation, the image is processed with the edge masks, only those points exceeding a local threshold are deemed to be edges and a gradient picture is produced by taking the maximum gradient magnitude at each point. The mask which yields the maximum gradient value determines the direction of the edge (coded 0-7). Thus a two dimensional array of numbers is generated, related to the edge direction.

This map is used to determine local connectivity of edge points to isolate closed regions.

The directions of edges are compared with succeeding and preceding points, allowing for edge vector variations of +1 or -1 to determine a connectivity relationship between edge points. A rule based logic system rejects extended, non connected edges, to leave the closed contour regions. This is shown diagrammatically in Figure 4.8.

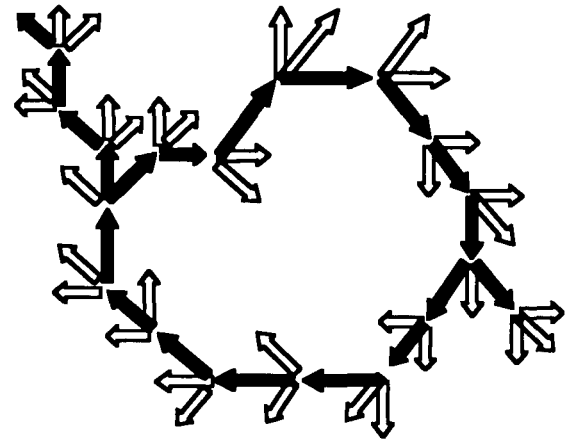


FIGURE 4.8 Closed Region Connectivity

As the target closes its image expands and may change aspect, as a result of manoeuvres. The closed contours in the edge map will expand and also change shape, but they will always remain as closed contours because of the natural contrast between the target and its background. Thus it is possible to maintain track very reliably, under these conditions. However, the situation described is quite benign, the real world presents a number of additional problems.

One major problem is that the background may also contain many closed contour features, such as clouds, the sun and terrain (if visible).

The general problem of eliminating background false alarms is one that is common to all autonomous detection systems. An IRST system has a particular problem resulting from its inherent advantages as a sensing system. The same factors causing the scene contrast that allows it to detect targets, day and night, also produces large signals from a wide variety of background sources. These signals are non linear and uncorrelated, thus traditional "noise reduction" processes are ineffective.

Over 500 man years of specialist processing and algorithm development have gone into the achievement of the low false alarm rate of the THORN ADAD passive ground based IRST system, and it has been necessary to conduct many hundreds of hours of trials under many varieties of background conditions, in all parts of the world, to achieve success.

The premise behind the success of background false alarm suppression is similar to the manner in which a human might operate. It is necessary to generate a "model of the world", in which the scene is broadly divided into objects, which are then analyzed,

categorised and tracked, whilst they exist in the field of view. This is necessarily a continuous process and occupies much of the processing power of the system. New objects entering the scene are placed in context with existing objects through rules of association. Streaming vectors are used to relate object distance and to predict object tracks. This model of the world is used to produce a continuously updated background "clutter map" of related objects that will not be regarded as targets.

All new detections, which are shown to form contiguous tracks, are compared with this background clutter map to establish whether they are related to background objects or are new, distinct entities, which might present threats.

Clearly, the model of the world and associated clutter map require some substantial processing power to set up and maintain, updated at each scan. If the host platform moves, it is necessary to change the axis co-ordinates of the model of the world, such that it remains correlated with the IRST FOV. It is also necessary to detect and analyze new objects entering the scene as the sightline shifts. In the past the available processing power has limited the ability of these algorithms to respond to rapidly changing scenes, thus applications have been limited to stationary ground based IRST systems, such as ADAD. However, with the advent of new, high speed, processing devices (DSP's) and improved architectures, it will be possible in the new future generation of IRST systems, currently being developed, to achieve both reliable long detection ranges and very low false alarm rates, with a moving host platform.

In operation, the new closed contour detections are compared with the clutter map and a rule base logic process is used to allocate the detection to either background or target categories.

An additional problem with tracking targets arises when they become obscured by local background objects. The usual technique employed in such circumstances is to "coast" the track, allowing it to continue on the same path at the same velocity until, hopefully, it is reacquired, at some point, when it emerges from behind the obscuration. This is a somewhat "hit and miss" process offering little confidence in its outcome. However, the existence of a model of the world allows for a better reaction to this situation. In the first instance, comparing the detection track with objects in the clutter map allows the anticipation of potential obscurations. Although the lack of precise range data limits "in front of" and "behind" inferences, it is known that when two objects become co-located one can be expected to "disappear" behind the other. Also, the outer closed contour of the obscuration will define the extent of the loss of track and allow prediction of the likely point of re-acquisition. A similar technique is used to rapidly categorise "pop-up" targets that may suddenly unmask at relatively close range, say from behind clouds.

Information from other sensor sources can be used to improve the operations described above. It particular they can be very valuable in developing and maintaining an accurate model of the world. In some cases it is possible to improve the model of the world, in the case of radar data being available, this would establish a third dimension (depth), which could be exploited not only by the IRST system, but by all on-board sensors, to validate and confirm targets.

Very close range targets present additional problems in that at some point the complete outline of the target may not be visible. Operational scenarios dictate that the IRST system should be able to accurately track targets down to less than 200m. At these ranges, the extra detail within the target renders the edge tracking system sub-optimal and the precise "aim point" of the tracker will "wander". A better method, in this situation, is to employ "image correlation tracking". The image correlation technique allows a chosen feature within the target (such as the cockpit) to be tracked (if visible) independently of the outline, or aspect, of the target. This is shown diagrammatically in Figure 4.9

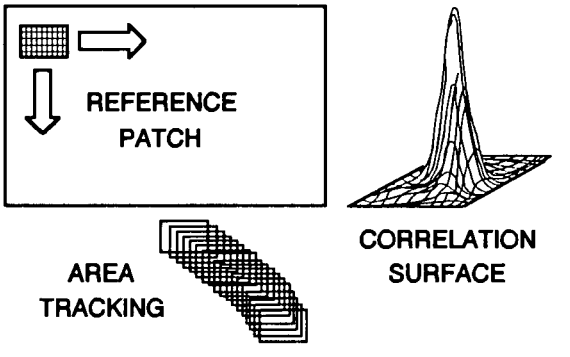


FIGURE 4.9 Image Correlation Tracking

The image correlation technique compares the stored image of a selected patch within the target with new video images from subsequent scans. The stored "reference" patch is correlated with a similar size patch of the new scan, at all points in the image, to find the best match. This position is used to update the track. Periodically the stored image of the aimpoint is updated by a new patch of video from the current scan. By tracking on a specified part of the target image a high aim point accuracy can be obtained. This can be useful in close combat situations where it can be used to cue weapons rapidly and accurately on to specific parts of the target.

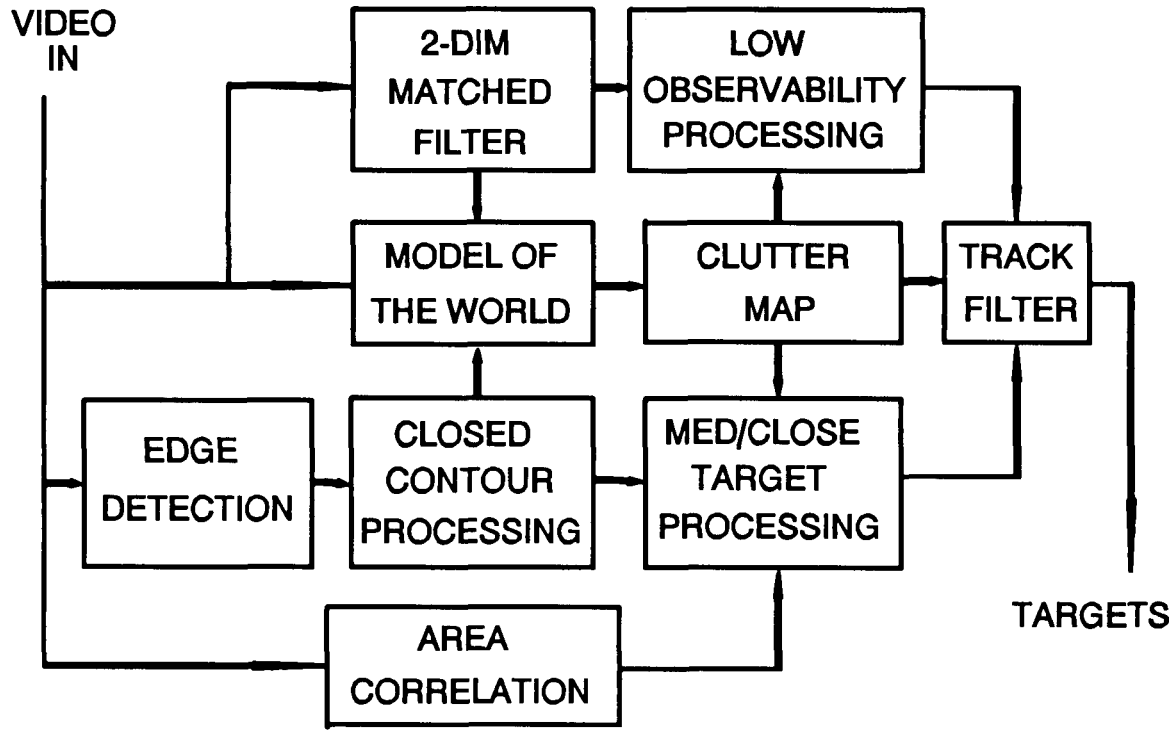


FIGURE 5.1 Generic Block Diagram of IRST Target Processing

The fusion of information from other sensors can be used to improve this process. In particular, for close in engagements, covertness is less important and the radar could provide both target confirmation and range information to the IRST system. This would then be used by the IRST to provide very accurate, confirmed, aiming co-ordinates for on-board weapon systems.

Some researchers have indicated that the availability of a resolved image of the target can be used to improve the motion estimates in tracking systems to enhance tracking accuracy [Ref 11]. In particular, the outline of the target from the edge contour processing can be used to derive an aspect angle measurement, which can then be used to provide a reduced "envelope" of likely future target positions. This would be based on maximum manoeuvrability along an axis perpendicular to the wing plane.

This technique could also be used to prime sudden target direction changes, which would be preceded by a drastic alteration in target aspect ratio, to further improve close-in combat performance.

5. CONCLUSIONS

This paper has presented some methods for the detection and tracking of airborne targets in an airborne IRST system. Techniques have been presented to deal with both long range, low observability, targets and shorter range, "pop-up" targets.

Current limitations on data rates and realistic aperture sizes mean that very long range targets will always be sub-resolution and specific low observability processing is required to achieve any useful performance gains.

The very high processing requirements needed to achieve the long range detection and tracking mean that, with current processor technology, it is not possible to carry out such processing over the whole of the scanned FOR of the IRST system in real time.

Thus the practical implementation of these methods requires some *a priori* knowledge as to the local region in which such targets might appear within the IRST FOR. This information could be gained from pre-mission intelligence, or it is possible that data obtained from other on-board sensors could be used to cue the IRST low observability processing search region.

Clearly, targets can appear at any range and those appearing locally at shorter ranges could well be of higher priority, being a more immediate potential threat. Such targets may well be spatially resolved and the processing specific to the low observability long range targets would be non-optimum in such a case. Thus, there is a need for further specific detection and track processing to cater for medium/short range targets. This paper has presented some processing options in this respect.

Ideally, both long and short range target processing must operate in unison within the same IRST system to achieve the desired all round performance.

Figure 5.1 is a generic block diagram of the processing that could be employed on a modern IRST system to permit the accurate and reliable detection and tracking of targets over a wide span of ranges. It contains a number of parallel processing paths, with associated decision logic to ensure optimum operation under all target engagement conditions. Matched spatial filters and low observability detection and tracking logic allow for very long range acquisition and tracking. Edge detection and area correlation processes allow for accurate and reliable acquisition and tracking of targets unmasking at shorter range. An analysis of the scene background is used to generate a Model Of the World (MOW) which is constantly updated with newly entering scene data. The reference axes of the MOW are constantly recalculated in accord with the changing line of sight of the IRST using both aircraft navigational data and, for fine corrections, by correlation of scene features from scan to scan. The MOW is used to generate a background clutter model, which contains the major objects in the background. This is used to categorise the detections and tracks obtained from the long and short range target processors to eliminate background related objects. In this way a very low false alarm rate is achieved under all background conditions.

The use of simple fusion of other sensor positional data has been described, with the purpose of both improving the performance of the IRST processing and to reduce its computational requirements.

However, although a useful level of performance improvement is possible, by the use of sensor fusion, the benefits that might be obtained are severely limited in application by the restrictions imposed by current aircraft data distribution systems and avionic architectures. It is considered that more emphasis will need to be placed on this aspect in future systems. Future avionic sensing systems must be designed with fusion in mind from the outset, in order to achieve optimum pointing and tracking performance.

For instance, it is certain that as weapon systems become more covert, faster and more accurate, human reaction time will become a very significant factor in the time to respond to an attack. In some situations it will only be possible to provide an effective countermeasure by reacting to a threat automatically. An operationally viable autonomous system will need to demonstrate a very high probability of accurate target detection, classification and tracking with a very low false alarm rate. The combination of information from a number of independent sensor sources may well be essential in order to achieve the requisite performance.

Additionally, a fully integrated multi-sensor suite would have the ability to optimise the available sensor deployment, by intelligent sensor allocation, to improve the overall system scope and performance in any given operational role or task. For instance, it may be important to maintain passivity, hence preference would be given to passive sensors such as IR. Under close battle conditions passivity may be of secondary importance to say range estimation when active sensors such as radar and laser would assume greater priority.

An IRST system will contain a powerful and highly sophisticated processing architecture which could well form the basis of a sensor data fusion "engine" providing significant performance improvements and resulting in increased operational benefits.

ACKNOWLEDGEMENTS

The work on ADAD was carried out with the support of the UK Ministry of Defence.

REFERENCES

Ref 1. Venables K. "Airborne IR Search and Track for EFA" 1st NATO-IRIS Joint Symposium, 14-16 July 1992.

Ref 2. Seyrafi K. "Electro-Optical Systems Analysis" Electro-Optical Research Company, 1973.

Ref 3. Coates P.V and Venables K. "Pirate, A State of-the-art Infrared Search and Track System for the EFA and its Potential for Data Fusion" AGARD Symposium on Integrated Target Acquisition and Fire Control Systems, 1991

Ref 4. Moravec K.S. "Integrating IRST Into Modern Aircraft" AFCEA Munich Chapter Technical Seminar, May 1990

Ref 5. Coates P.V. "Detection and Classification of Single Airborne Targets over a Wide Field of Regard" 54th AGARD AVP Symposium on Electro-Optical Systems and Image Analysis for Airborne Applications, Athens 1978.

Ref 6. Reid D.B. "An Algorithm for Tracking Multiple Targets" IEEE Transactions on Automatic Control, AC-24, Dec 1979.

Ref 7. Gelb A. "Applied Optimal Estimation" Cambridge, MA: MIT Press, 1974

Ref 8. Aidala V.J. "Kalman Filter Behaviour in Bearings-only Tracking Applications" IEEE Transactions on Aerospace and Electronic Systems AES-15, Jan 1979.

Ref 9. Venables K. "IR Search and Track - ADAD" 1st NATO-IRIS Joint Symposium, 14-16 July 1992.ADAD

Ref 10. Prewitt J.M.S. "Object Enhancement and Extraction" Picture Processing and Psychopictorics, Academic Press, New York,

Ref 11. Dayton D. et al. "Tracker Performance Improvement using Multiple Sensor Data" SPIEE Vol 1100 Sensor Fusion ll, 1989.

THE APPLICATION OF MMW/IR SENSOR FUSION TO TACTICAL U.S. AIR FORCE WEAPON SYSTEMS

J. W. Watson
Nichols Research Corporation
Shalimar Center, Suite C-1
One 11th Avenue
Shalimar, FL 32579, USA

Sengvieng Amphay
Bryce Sundstrom
Department of the Air Force
Wright Laboratory WL/MNGA
101 West Eglin Boulevard
Eglin Air Force Base, FL 32543

ABSTRACT

The precision guidance of autonomous weapons to fixed high value targets (HVTs) and critical mobile targets (CMTs) in high clutter backgrounds under adverse weather conditions represents one of the largest challenges facing the tactical Air Force weapons development community today. Such weapon systems will deny the enemy the cover of weather, just as infrared systems have eliminated the cover of night. One technology that is being explored as a means of addressing this mission is millimeter wave (MMW)/Infrared (IR) sensor fusion. The Armament Directorate of the U.S. Air Force Wright Laboratory has been sponsoring research in the area of MMW/IR sensor fusion for tactical weapon systems since the mid 1980's, much of which was accomplished under joint sponsorship with the U.S. Army Missile Command (MICOM). The purpose of this paper is to address the state of development of tactical Air Force MMW/IR sensor systems, and to demonstrate the results of current risk abatement efforts.

1.0 INTRODUCTION

The smart weapons employed so spectacularly against fixed high value targets during Operation Desert Storm actually represent late 1960's/early 1970's technology. The laser guided bomb, for example, had been successfully employed during the Vietnam war. For both of these conflicts, however, the friendly forces held complete air superiority. It is questionable whether an equivalent degree of unchallenged dominance of the air battle space will be so easily achieved in future conflicts. Additionally, improvements in enemy surface-to-air weaponry will require greater standoff capabilities for strike aircraft. The operational concept of having one aircraft designate a HVT while the other delivers the weapon will no longer be viable.

To many post Gulf Crisis evaluators, a particularly distressing determination was the frequent occurrence of inability to fly precision strike missions due to poor weather conditions. The number of sorties canceled was more than double the pre-Desert Storm estimates. An even clearer lesson learned was the extreme difficulties associated with finding and successfully attacking critical mobile targets. The coalition air forces flew some 5,500 sorties over Iraq without a single confirmed kill of a mobile SCUD [1]. This experience further demonstrated the need for an all-weather weapon system capable of acquiring and tracking CMTs.

It is very likely that the requirements specified for the next generation of smart air-to-surface weapons will be driven by the lessons learned during Operation Desert Storm. As a minimum these new weapons will incorporate a high degree of all-weather autonomous search, acquisition, and track-to-impact capability. However, achievement of high probabilities of target kill under all-weather conditions, even with newly

emerging weapon systems, is highly unlikely. All of these emerging systems still have limitations; their sensors are either blind at night, blinded by weather, or simply can't "sense" very well.

2.0 BACKGROUND

There has also always been a military need to find targets. This "sensing" can take many forms, some not quite so obvious. For example, sensing often takes the form of detecting and analyzing the signature of the object, whether it be seismic, acoustic, electromagnetic, or of some other origin. Even though the signature is not the object itself, if it can be uniquely attributed to the object with a high degree of confidence, then sensing the signature is equivalent to seeing the object. But, all such one-dimensional signature approaches to automatic target recognition have proven inadequate. Therefore, to be able to achieve an all-weather two-dimensional imaging capability, various schemes have been considered involving active micro or millimeter wave radar with electro-optical or imaging infrared sensors.

Millimeter Wave/Infrared sensor fused systems offer several advantages over their unitary sensor counterparts. They provide for an all-weather capability, allowing for operation in fog, rain, and smoke, while retaining the high resolution imaging capability of the IR sensor. This higher resolution is required for both improved target classification and aimpoint guidance accuracy thereby minimizing collateral damage. Countermeasure hardness is inherent to the dual-band design since unitary deceptive or denial methods are no longer effective. In addition, the two independent signal sources result in improved detection performance.

Fusion of the information from the two sensor types is accomplished in the signal processing algorithms. A common aperture can be utilized to ensure channel registration, but separate detection devices are required to sense the energy in the two spectral regions. Once collected, the two signals are then fused by the signal processing algorithm. In general, three distinct levels of sensor fusion are utilized: pixel, feature, and decision. At pixel-level fusion, the raw data is merged before being subjected to processing. This type of fusion is not plausible for MMW/IR weapon systems because the angular resolution differences between the two sensor types is commonly as high as 100:1. At the next level, feature-level fusion, preliminary processing is performed on the raw data in order to derive descriptive attributes (such as radar cross section or IR spatial extent) which are then subjected to further processing. A third form of fusion is decision-level in which each channel is processed independently and the results combined through some higher level logic to determine target presence.

Research has shown that feature-level fusion provides the greatest increase in performance for the sensor configuration considered. As previously mentioned, IR/MMW pixel level fusion is not possible, and decision level fusion has been found to mimic the performance of the worst sensor of the pair and offers no real improvement.

A number of efforts have been conducted over the past decade to characterize the detection and classification improvement provided by MMW/IR sensor fusion. For example, an algorithm which simply performs an AND on single channel detection has the effect of reducing false alarms but at the same time lowering detection performance. The probabilities are given by:

$P_{fa}=P_{fa1}*P_{fa2}$ where $P_{fa1,2}$ is the probability of false alarm of the individual channels.

$P_{d}=P_{d1}*P_{d2}$ where $P_{d1,2}$ is the probability of detection of the individual channels.

More complex fusion concepts have demonstrated the more desirable feature of increasing detection performance while reducing the number of false alarms through testing. However, a general characterization of these performance gains has yet to be

developed, largely because of the non-linearity of such devices.

2.1 History of Development

The fact that sensors from the electro-optical and radio frequency spectral regions are complementary was known early on, but it is difficult to determine just when the concept for sensor fusion was first postulated. An early example of sensor fusion where millimeter wave was first closely coupled with imaging infrared, albeit crude, can probably be traced to the 1970's development of target activated munitions (TAMs) such as the Sense and Destroy Armor Munition (SADARM).

A later and somewhat more sophisticated approach to hardware design and development of sensor fusion algorithms by General Dynamics (GD) Corporation stirred up considerable interest [2]. During this early 1980's effort, GD demonstrated a 140 GHz/3-5 μ MMW/IR seeker for the 4-inch diameter Terminally Guided Submunitions (TGSM). The pseudo-imaging IR sensor for this unpowered gliding mini-missile employed four fiber optic coupled detectors in a rosette scan mode. Impressive results were obtained during captive flight test demonstrations, particularly in terms of countermeasure resistance.

By 1984, a memorandum of agreement had been signed by the Directors of the U.S. Air Force Wright Laboratory Armament Directorate and the U.S. Army Missile Command to jointly develop Dual Mode MMW/IR technology. Shortly thereafter four study contract awards were let to Honeywell, General Dynamics, Hughes, and Texas Instruments. The objective of these studies was to define a seeker with the widest degree of target and delivery system applicability. If a common seeker, applicable to both Air Force High Value Targets and Army Ground Mobile Targets (GMTs) could not be achieved, then as much commonalty as possible was the new, admittedly vague, goal. Results from these studies were presented at a NATO Symposium on Terminally Guided Weapons [3].

A number of commonality issues were identified during these studies. Some of the issues of greatest concern were those associated with search area requirements. If a common search area could not be achieved, then it was believed that a variety of aperture sizes and gimbal designs would be necessary. Fortunately, these studies revealed that a maximum search width of approximately one kilometer was sufficient for the seeker applications considered. Downrange search requirements varied, but these did not appear to present any significant design difficulties. This meant that a common aperture size and gimbal could be achieved. Another, somewhat significant outcome of the studies was that the seeker target detection and acquisition algorithms for Air Force HVTs could be somewhat simpler in concept than that required for Army GMTs. As such, the Air Force requirements could be considered to be a subset of Army requirements. This latter result led to much better focused follow-on development effort oriented primarily toward the ground mobile suite of targets.

In June of 1986, two technology demonstration contracts were awarded, one to Westinghouse and one to Textron Defense Systems. The purpose of this program was to fabricate and integrate a MMW/IR test bed, collect coincident MMW/IR data with this test bed, and to develop and assess target acquisition algorithms using these data. As part of this effort, free flight designs of the DMS systems were developed and used as the basis for configuring the data collection test beds. A MMW/IR common aperture was developed and successfully demonstrated. A limited number of tower and captive flight tests were also conducted.

At the conclusion of the Dual Mode Seeker program in 1989, three contracts were awarded (Westinghouse, Boeing, and Alliant Techsystems) to perform additional algorithm development and data collection. Both tower and captive flight data were collected under these efforts. Several of the algorithms were also refined during this

period. The last of these efforts concluded in 1991.

The focus of sensor fusion research shifted at the conclusion of the seeker demonstration effort in 1991 from hardware demonstration to technology exploration. Some of these programs are discussed in the following section, along with the technology risk issues they are intended to address.

3.0 CURRENT RESEARCH EFFORTS

Since the conclusion of the Dual Mode Seeker program, several research efforts have been initiated aimed at addressing the technical issues identified during earlier hardware demonstration efforts. Table 3-1 summarizes some of these programs along with which specific issue they are intended to address. A brief summary of the objectives and status of some of these programs is also provided.

3.1 MACET Sensor Fusion Testbed

MACET is intended to be a user friendly software system to support the rapid prototyping and evaluation of air-to-surface acquisition and aimpoint selection algorithms. In order to facilitate this function, the following design requirements were specified for MACET: minimum training time, rapid prototyping capability, compatibility with existing data, capability to accept existing algorithm components, algorithm evaluation shell, multi-sensor algorithms, emulation of target acquisition and aimpoint selection algorithms, Sun Workstation platform, nominal development costs, generation of standard algorithm performance metrics, flexible graphical output, and data probe placement [5].

From these design requirements, it was determined that MACET would be constrained to execute on a Sun Workstation and must utilize existing software platform(s) in order to maintain nominal development costs. Other required capabilities include: a graphical user interface, the establishment of a standard data format, a computer aided algorithm design capability, and an ethernet interface.

The MACET concept is illustrated in Fig 3.1-1. A Sun SPARC Station 2 serves as the host platform for MACET. An ethernet is used to link the Sun with VAX systems located within the Radar Signal Processing Laboratory (RSPL) and Imaging Processing Laboratory (IPL) where existing data reduction and analysis utilities reside. Measured data from the TABILS database, laboratory tower and captive flight test exercises are used to exercise the prototype algorithm configurations. MMW and IR synthetic data from signature prediction codes, such as Irma, are also used for this purpose [6]. MACET will also contain a library of algorithm components generated during previous research efforts under 6.2, 6.3, and Small Business Innovative Research (SBIR) programs such as the Advanced Tactical LADAR Seeker (ATLAS), Low Cost Anti-Armor Submunitions (LOCAAS), Joint Adverse Weather Seeker (JAWS), Dual Mode Seeker (DMS), and others; thereby providing on-line access to over a decade of research in the area of autonomous acquisition algorithms for air-to-surface guidance. A graphical user interface is employed to minimize the system training time required of engineers and analysts to use the tool.

The MACET system is developed upon the Paragon Image Logic and Khoros 1.0 software platforms. The Paragon Image Logic was modified to support the top-level functionality of the MACET architecture including the user interface. Signal and image processing routines from Khoros are used to provide lower level library and functional routines for the purpose of algorithm prototyping and output display. This configuration was selected for MACET development because Paragon was found to have the better graphical interface and offered better user support at that time. However, given the similarity of the Khoros and Paragon environments, subsequent releases of these two software packages have been closely monitored in order to take advantage of any increased functionality.

Table 3-1. Research Programs Aimed at Addressing Dual Mode Seeker Technical Issues

ISSUE	RESEARCH
Acquisition Algorithm Performance	• Sensor Fusion Algorithm Performance Metrics Study • Modular Algorithm Concept Evaluation Tool (MACET)
MMW/IR Registration	• Superresolution Processing
MMW/IR Database	• Irma Multi-Sensor Model
Mission Planning	• Application of Model-Based Vision to Sensor Fusion
Tracking Algorithm Performance	• Multi-Sensor Tracking Filter

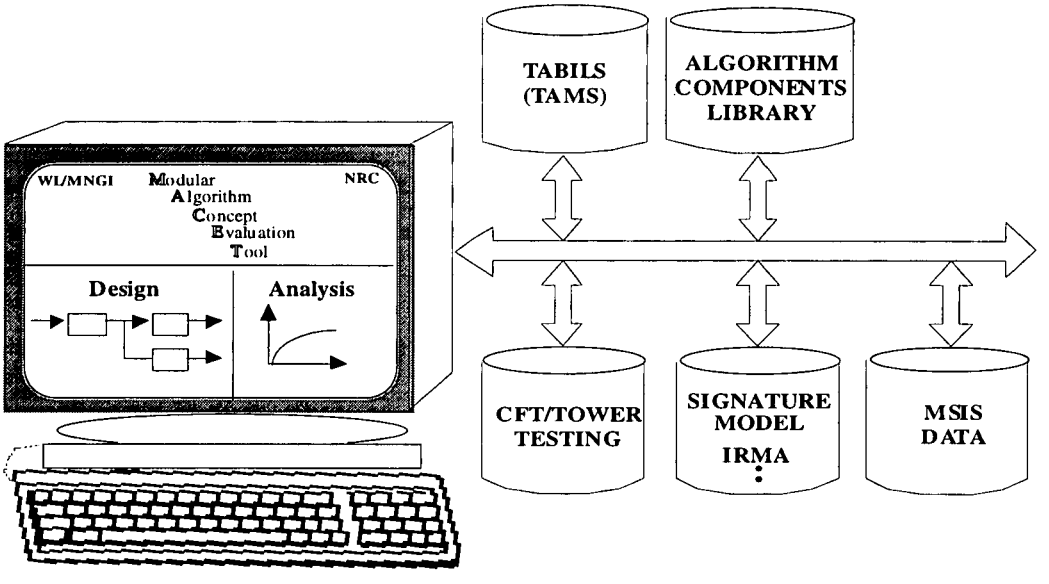


Figure 3.1-1. MACET Concept

The MACET architecture is illustrated in Fig 3.1-2. As shown in this graphic, MACET consists of six major functional areas. These functional areas include: ground truth editor, data selection/format conversion, algorithm selection, performance evaluation, output selection, and utilities. The ground truth editor provides the capability to tag targets in the data scene before injection into the signal processing algorithm in order to facilitate performance scoring. The data selection/format conversion function allows the user to select a data set for algorithm testing and will convert all selected data to a format compatible with other MACET functions. The algorithm selection function facilitates algorithm prototyping using either existing algorithms or algorithm components provided in an on-line library or by providing the capability to define new elements. Once the algorithm has been defined and the data selected, the performance evaluation function can be activated. The purpose of this function is to compute defined algorithm performance metrics (such as probability of detection/probability of false alarm) or user defined metrics via probe placement. Upon completion of algorithm testing, the resulting metrics can be viewed using the output selection function. Both graphical and text output are supported. Lastly, a utilities function is provided for workspace manipulation and accessing lower level MACET data processing routines.

Development of the MACET system is well underway with significant progress being realized in the areas of data conversion, ground truth editor, performance evaluation, and output selection. Incremental versions of the MACET will be delivered as additional channels (sensor types) are added. This incremental development approach will result in a comprehensive active/passive IR/MMW version, capable of emulating either unitary or sensor-fused systems, to be delivered at the conclusion of the Data Analysis and Modeling (DAAM) contract in November 1995. This simulation will provide the flexible non-realtime emulation by which to rapidly prototype and evaluate unitary and sensor fused algorithms.

A passive IR/active MMW detection and classification algorithm has been demonstrated using the MACET platform [4]. A block diagram of the dual-mode algorithm is shown in Fig 3.1-3, and the corresponding MACET implementation in Fig 3.1-4. The pictograph or glyph labeled "mmwcat" concatenates four dwells of MMW data which constitute a frame co-registered with the IR frame. The "input" glyph brings in the corresponding IR data. Characteristics of the MMW beam are specified in the glyph labeled "beamgen" which generates an approximate intensity beam pattern to be convolved with the MMW data. Normalization and convolution operations are performed by "prefilt" which outputs an enhanced version of the original image. Statistics of this image are computed including the location of the peak within the image. This information is used by "extractir" to determine a small (64x64) region of interest to be extracted from the IR image for further processing. Here the extracted IR image sub region is scaled and correlated with sets of target images at various orientations using multiple copies of "correlate", one for each target image set. In the present implementation, three targets were considered at twenty-four regularly spaced orientations. The results of the correlations could be input to decision logic, but for the example at hand are only output as plots of correlation versus orientation.

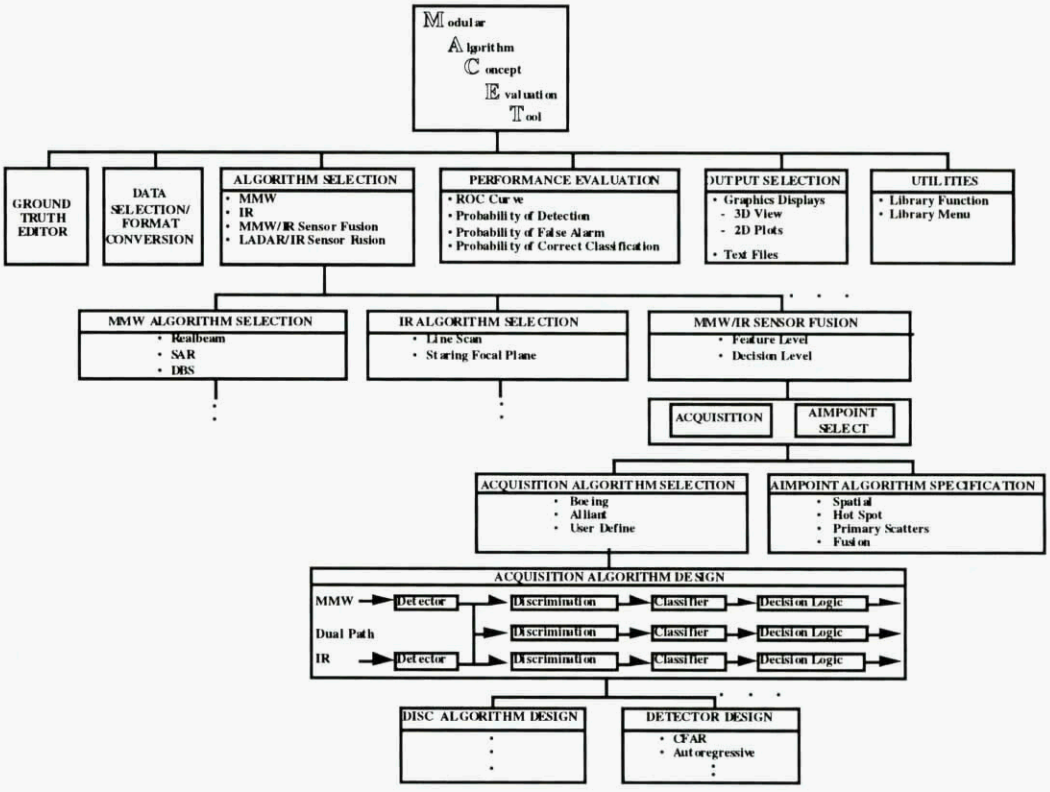


Figure 3.1-2. MACET Architecture

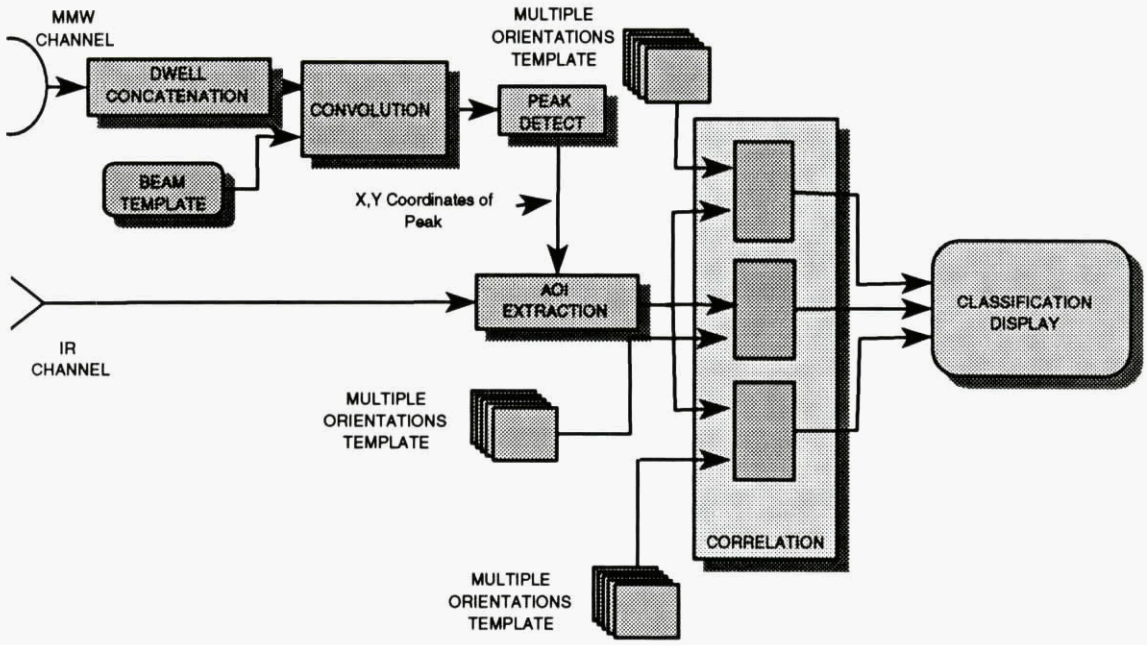


Figure 3.1-3. Dual Mode ATR Algorithm Block Diagram

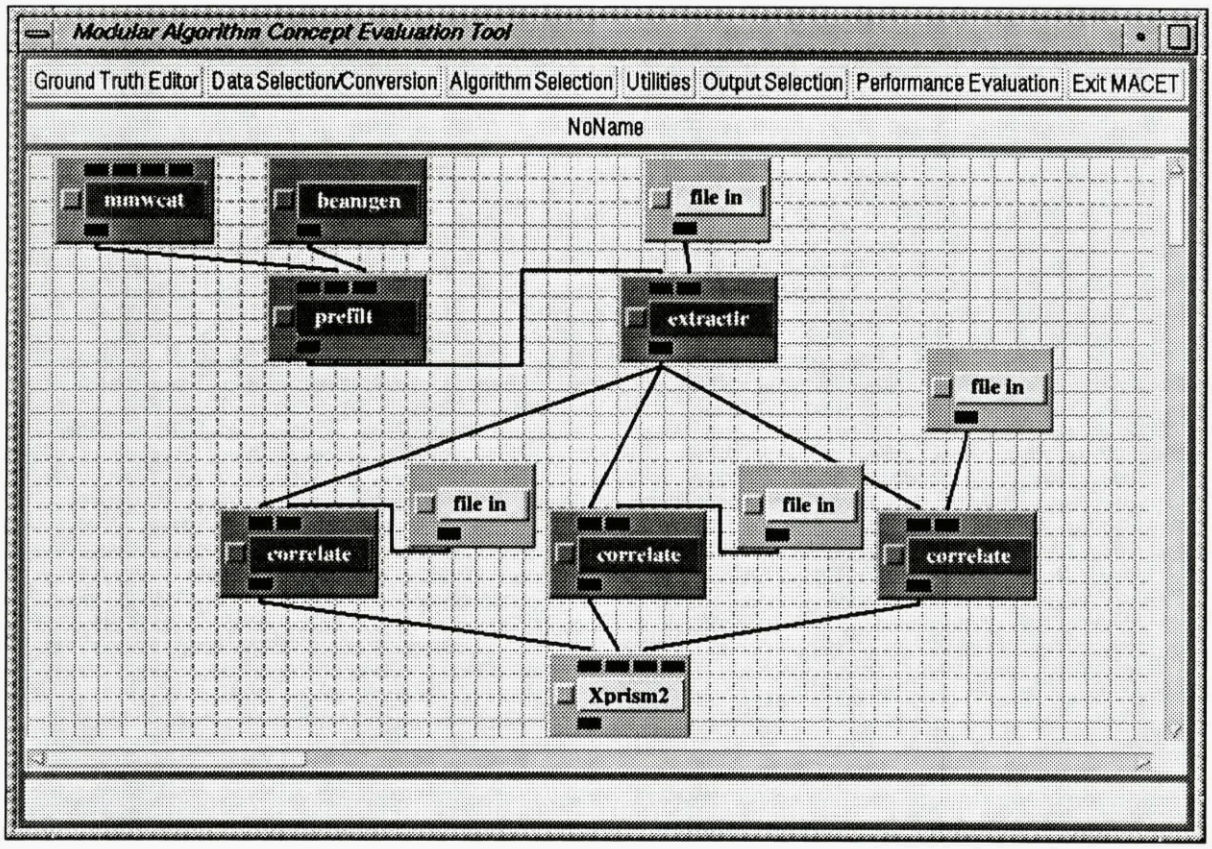


Figure 3.1-4. MACET Implementation of Dual Mode Algorithm

The three candidate targets are shown in Fig 3.1-5: a T62 tank, BMP, and ZIL. These are synthetic IR signatures generated by the Irma Signature modeling code [5,6]. The actual target is a T62 tank as seen by the IR detector described above. Fig 3.1-6 shows an example plot of the correlation of the actual target with candidate targets versus orientation.

This algorithm was developed entirely within the MACET environment without the necessity of writing any additional code.

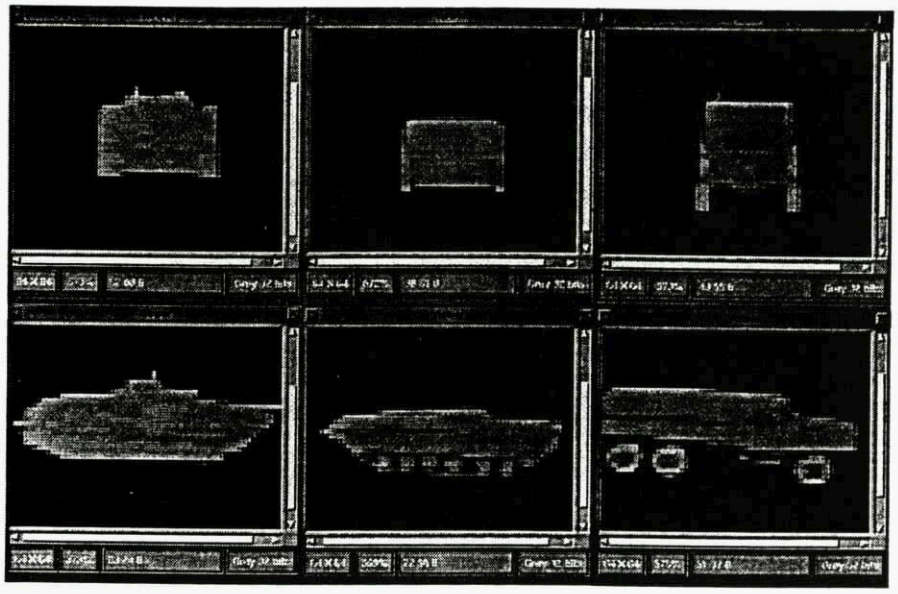


Figure 3.1-5. Candidate Targets: T62; BMP; ZIL

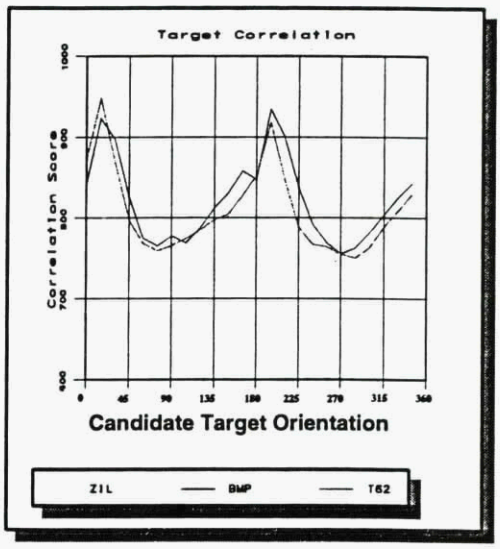


Figure 3.1-6. Plot of Correlation vs Orientation

3.2 Irma Multi-Sensor Predictive Model

The Irma code was originally developed during FY78-80 for the U.S. Air Force Wright Laboratory/Armament Directorate. This simulation represented one of the first high resolution synthetic IR target and background signature prediction models to be developed for tactical air-to-surface weapon scenarios.

During the 1980's, several improvements were made to the Irma code. These included the addition of models of Ground Mobile and High Value Targets, a higher fidelity, more realistic thermal response model, a sensor effects post processor, effects of heat flow through target surfaces, and a user interface. These enhancements culminated in the release of a new version of the code in 1989, designated Irma 2.1.

During the early 1990's, an active/passive IR version of the code was developed, and an enhanced scene generator was incorporated into Irma. To provide a target laser signature capability, models from the Defense Laser Target Signatures ($DELTAS_{NRC}$™) code were adapted to execute within Irma and to generate representative air-to-surface seeker outputs. This code was released in 1990 as Irma 3.0. The enhanced scene generator, the Grumman Scene and Sequence Generator (SSG), added the capability to model 3-D backgrounds, correlated frame-to-frame imagery, and generalized quadric surface descriptions. Released in 1992, the code was designated Irma 2.2. In 1994, Irma 3.2 was released which added a passive MMW channel, polarimetric IR, and enhanced treatment of solar and terrain reflectance.

Currently, efforts are underway to incorporate both active and passive channels for both IR and MMW into Irma 4.0. Approaches to the development of these channels have been addressed in several studies [6,7,9,10]. Implementation of the passive MMW channel was accomplished as part of the development of a passive MMW/IR version of Irma (Irma 3.2). A working version of the Irma active MMW channel has been demonstrated as part of the Irma 4.0 prototype. Irma 4.0, the active/passive MMW/IR version of the code, is slated for release in October 1995. A working prototype is currently undergoing testing.

A complete summary of the various versions of Irma is presented in Table 3-2.

The planned architecture for Irma 4.0 is illustrated in Fig 3.2.1. Separate signature channels will be implemented to generate images in the four spectral regions. External interfaces will be designed to be compatible with the WL/MNGA weapon system simulation standard to permit its use in larger simulations.

Table 3-2. Irma Model Versions

Irma Model/Version	Status	Capabilities
1.0	Out of distribution	Passive IR only. No interface.
2.1	In distribution	Passive IR. Improved thermal response routines. User interface. Vax-based.
2.2	In distribution	Passive IR. 3-D scene generator. Frame-to-frame correlation. Vax-based.
2.2(S)	In distribution	Sun-based version of Irma 2.2.
3.0	Out of distribution	Passive IR/Ladar channels. Vax-based.
3.1	In distribution	Improved Irma 3.0
3.2	In distribution	Passive MMW/IR. Sun-based.
4.0	Prototype development underway	Passive IR/Ladar channels. Active/passive MMW channels. Sun-based, motif interface.

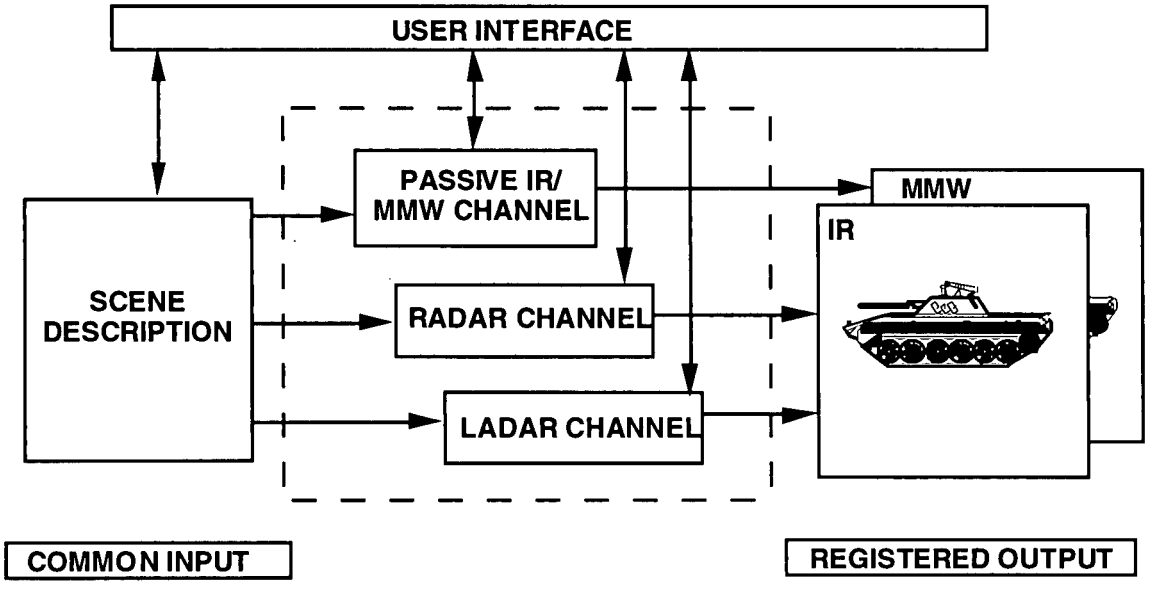


Figure 3.2-1. Irma 4.0 Model Configuration

In order to ensure upward compatibility of the Irma releases, existing elements of current model versions will be adapted to execute in the environment illustrated in Fig 3.2-1. Primarily these include the ENVIRO thermal response model (Irma 2.2), the SEEKER/DELTAS$_{NRC}$™ modules (Irma 3.1), and the Irma 3.2 passive MMW signature channel. The interface and scene generator functions represent the areas requiring the most change and are subject to replacement.

The basic architecture of existing Irma software is shown in Fig 3.3-2. The thermal response model, ENVIRO, computes the surface temperatures and thermal radiances as a function of time, accounting for such effects as environmental heating, material properties, and internal heating. ENVIRO incorporates a single dimensional heat flow model to perform these computations which represents the inner surface temperatures as a thermal reservoir coupled through a medium with a fixed R-factor. The time-dependent temperature at the surface and within the material is calculated using either a finite difference Crank-Nicholson technique, or for thermally thin material, a fast-running uniform temperature model.

The image generator module converts the three dimensional signature maps into two dimensional imagery. Sensor and atmospheric effects are introduced in the image generator. Examples of such effects include sensor spectral response, sensor optical blurring, fixed pattern noise, shot noise, and scan effects.

The Irma interface, IRMINT, is a user-friendly, menu-driven program which facilitates setting up, running, and examining the outputs of both ENVIRO and the image generator. IRMINT is machine specific since it makes use of numerous system utility programs to perform such tasks as present menus , spawn jobs, and read directories. The Irma interface is available for VAX systems and Sun workstations.

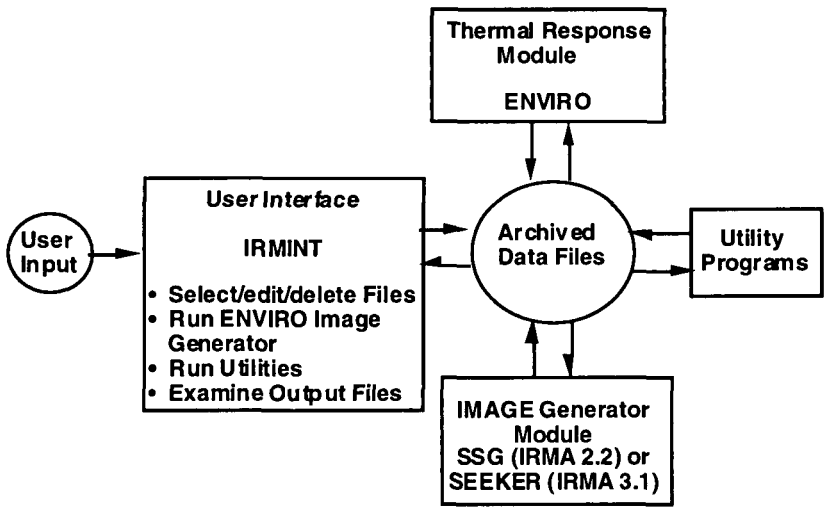


Figure 3.2-2. Existing Irma Simulation Architecture

The last released version of Irma, Irma 3.2, provides the user with the capability of modeling polarized, passive millimeter wave and infrared signatures. The Irma passive IR model, Irma 2.2, was modified to provide a passive MMW signature prediction capability. Upon examination of the phenomena associated with both reflected and self-emitted radiation in the MMW spectral region, it was concluded that these processes were very similar except in the following areas: (1) target/background emissivities, (2) sensor resolution, (3) polarimetrics, and (4) the contribution of reflected radiation in the imagery. An approach was developed to include these effects [7,8]. Irma 3.2 has been used to model signatures in the UV and visible regime where the signatures are all reflective, in the IR regime where the signatures are primarily emissive, and in the MMW regime where the signatures are both reflective and emmissive.

The Irma 4.0 prototype, the version to generate passive MMW, IR, Ladar, real-beam radar, and SAR, is currently undergoing testing and has been demonstrated to generate registered four channel imagery. Its design consists of tightly coupled passive channels, standalone active channels, and an X_windowing user interface. All four channels utilize the same geometrical representation, thereby ensuring channel registration. Once completed, it will serve as one of the cornerstones in multi-sensor precision guided munitions research.

3.3 Application of Model Based Vision to MMW/IR Sensor Fusion

To examine the feasibility of a multi-spectral model-based vision system, the test configuration depicted in Fig 3.3-1 was constructed. MMW/IR captive flight data collected during the Autonomous Acquisition Dual Mode Seeker (AADMS) program was reduced and used to test both classical and model-based vision classification approaches. The detection and discrimination stages of a prototype algorithm under development to demonstrate MACET were used to identify and register the MMW and IR Regions of Interest (ROIs) within these scenes. Candidate targets were then subjected to both classical (Quadratic Bayesian) and model-based vision classifiers. MMW and IR data from the Target And Background Information Library System (TABILS) were used to train the quadratic classifiers. Synthetic scenes were simulated using a prototype of the Infrared Modeling and Analysis (Irma) version 4.0 code. Both types of classifiers operated on the registered MMW and IR scenes independently. The resulting classification metrics were then fused using a weighted linear transformation. The elements of this test configuration are described in detail in Reference 4.

An example of the output from the MMW model-based classifier is shown in Fig 3.3-2. In this test case, the classifier was exercised against measured APC, tank, and truck signatures. The lowest scores represent the best fit for this example since a minimum distance algorithm is employed by the classifier. Fig 3.3-2 illustrates this distance as the test and training profiles were circularly convolved. These results demonstrate the difficulty the classifier had in distinguishing between the tank and the APC. This is consistent with the performance of classical MMW realbeam classifiers.

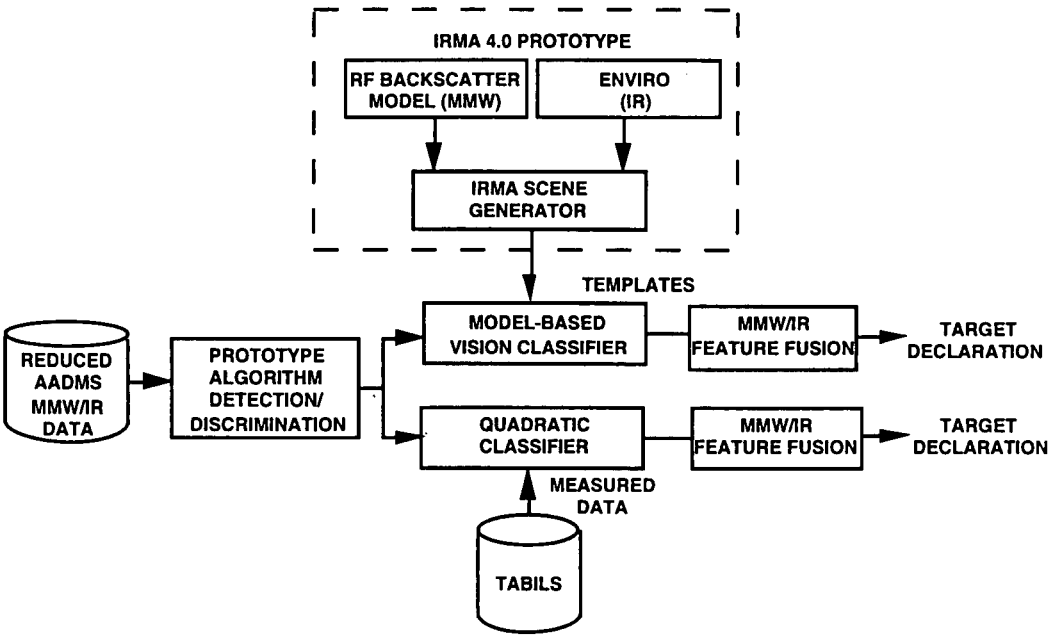


Figure 3.3-1. Model-Based Vision Classifier Test Configuration

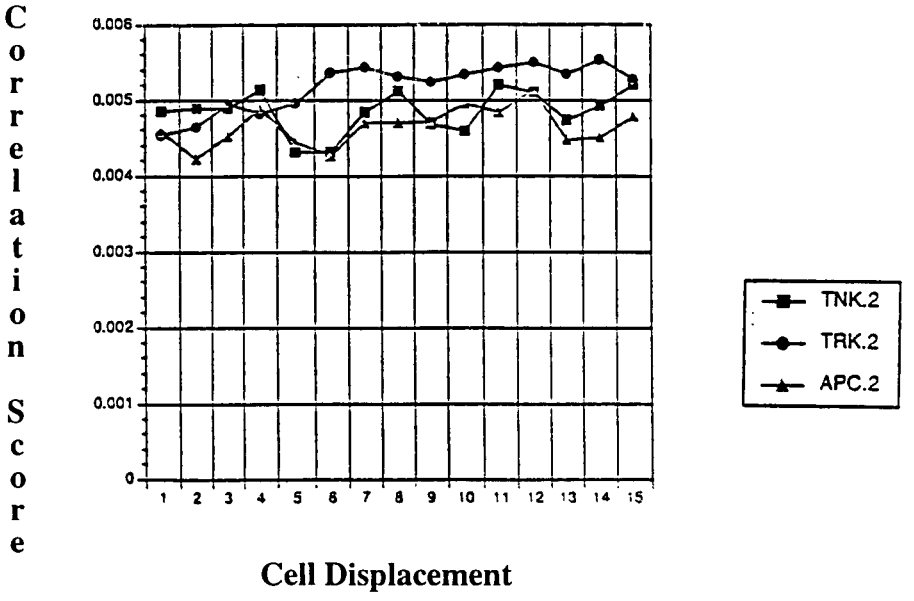


Figure 3.3-2. MMW Model-based Classifier Results

Output from the IR model-based classifier is shown in Fig 3.3-3. The target-to-background contrast was found to be relatively poor in all the scenes processed. Thus intensity resolution in the target regions of interest was not as fine as would be desired. Another source of error for the IR classifier occurred when the average intensity level in the image was not relatively constant, or when there was a non-target area of higher intensity than the target region. In that case, the classifier tends to operate on the region of highest intensity with a stronger weight. A technique, described previously, to normalize the correlation to help discount intensity levels on a region basis proved successful.

The results of the MMW/IR feature fusion are shown in Table 3.3-2. The scores shown in this table have been weighted by the contrast ratios. These scores were produced by exercising 3 test scenes (tank, truck, APC) against the two classifier types trained with tank data. Both classifiers, statistical and model-based, selected the APC over the tank. As mentioned previously, this result is consistent with other algorithms. The chart also indicates that the model-based classifier tended to favor the truck over the tank; however, examination of the unweighted scores demonstrated a higher score for the tank. Further examination uncovered that the MMW tank contrast ratio was low and tended to bias the dual-mode scores.

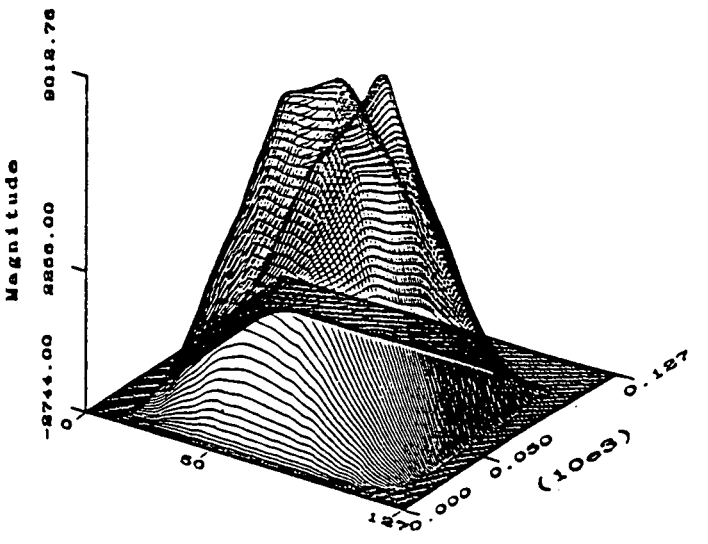


Figure 3.3-3. IR Model-based Classifier Performance for Tank

Table 3.3-2. Classifier Comparisons

Target Scene	Classification Method					
	IR Model-based	IR Quadratic	MMW Model-based	MMW Quadratic	Dual Mode Model-based	Dual Mode Quadratic
Tank	0.25	0.375	0.375	0.5	0.3125	0.4375
APC	1.0	1.0	0.5	0.75	0.75	0.875
Truck	0.5625	0.375	0.75	0.375	0.65625	0.375

Preliminary results indicated that the performance of such a system is comparable to that of classical methods. This would indicate that a model based approach has merit since it greatly reduces the complexity of the mission planning task. However, the data set on which these findings were based was too limited to conclusively support this statement.

3.4 Low Cost Dual Mode Seeker

With the changing emphasis from preparation for global warfare to involvement in regional conflict with the desire to inflict minimum collateral damage to civilian populace and property, precision guided weapons have taken on a new order of importance. Previously, Dual Mode MMW/IR seeker development was oriented towards thwarting Follow-on Force Attack. This was a many-on-many concept wherein long-range delivery vehicles, such as the Multiple Launch Rocket system, would dispense terminally guided submunitions to attack massed armor. This particular scenario is now much less likely. The more probable scenario is the one-on-one, close-in, aircraft delivered precision guided weapon to a specified target. This latter scenario invokes other practical considerations, such as aircraft survivability, which demands weapons delivery from stand-off ranges and autonomous guidance to permit launch and leave maneuvering. These and many other requirements are currently being addressed and implemented in the USAF Aeronautical Systems Division (ASD) Joint Direct Attack Munition (JDAM) program.

Various programs are developing improved guided bombs capable of all weather operation by integrating Global Positioning System (GPS) guidance with an improved Inertial Navigational System (INS). Even so, this union of GPS and INS will still not provide the desired degree of precision for select targets. Various terminal guidance schemes have been proposed for the silent target application such as Synthetic Aperture Radar (SAR), Dual Mode MMW/IR, Ladar, other multi-spectral, and stand-alone MMW or IR based solutions. However, since the GPS/INS guided bomb itself eliminates any requirement for wide area search, comparatively simple antenna scanning and tracking systems could be employed. For example, Fig 3.3-4 depicts a prototype of a low cost MMW/IR seeker developed specifically an earlier version, Inertially Aided Munition program.

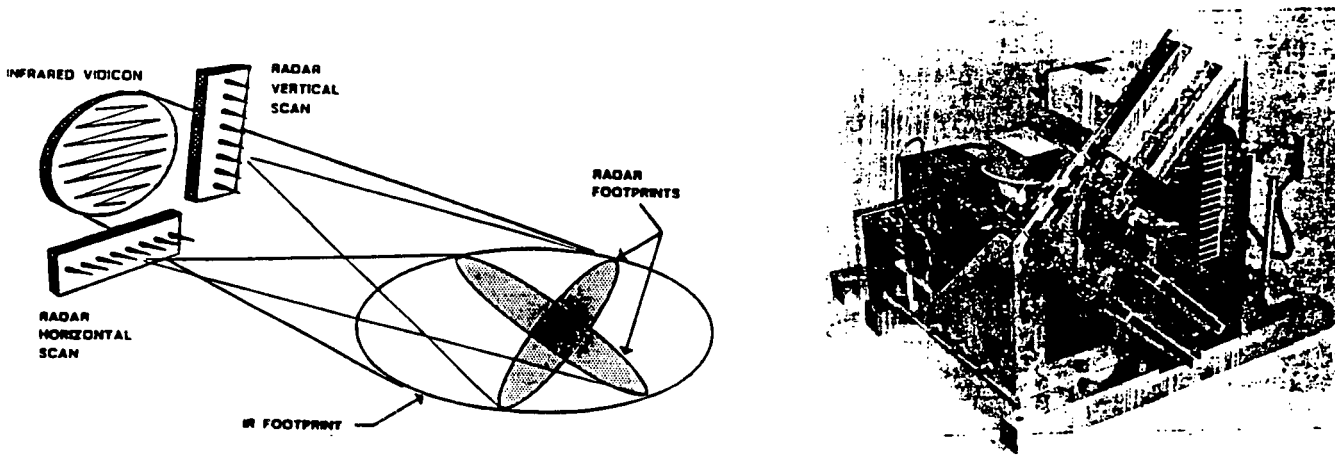


Figure 3.3-4. Low Cost Innovative Dual Mode Target Sensor

The MMW sensor utilized in this system is a frequency-modulated continuous wave (FMCW) high range resolution radar with a 94 GHz center frequency. It is configured of two antennas arranged in a symmetric "L-shaped" layout. Each of the orthogonal antennas generates a long, narrow fan beam. These beams may be frequency-steered to perform target search and two-axis angle track functions. The imaging IR sensor is an uncooled, commercially purchased, off-the-shelf pyroelectric vidicon-based camera that is sensitive to radiation in the 8-14 μm spectral region [11].

4.0 CONCLUSIONS

The U.S. Air Force Wright Laboratory/ Armament Directorate has been involved in the development of MMW/IR sensor fusion technology for tactical weapons application for over a decade. Significant resources were devoted, jointly with the Army, to hardware demonstrations in late 1980's. Under these efforts several milestones were demonstrated including the fabrication of a common MMW/IR aperture, and the flight test of a data collection instrument configured with this aperture. The hardware demonstration efforts also served to identify several issues associated with MMW/IR fused sensors: namely, the optimum design of the target acquisition algorithms and the lack of data and phenomenological models to support the development of these algorithms. Since 1990, significant resources have been devoted to resolving these issues through a number of programs with the most concentrated effort being performed under the Data Analysis and Modeling (DAAM) program. Under this effort, basic research is being performed in the area of MMW/IR signature modeling and data analysis, as well as algorithm evaluation tools and techniques.

The early indications of the benefit of Dual Mode MMW/IR seeker technology implied that there might be as much as a four-to-one improvement in detection performance per dollar cost over standalone MMW or IR systems. For the large production requirements anticipated prior to the end of the cold war, that degree of improvement represented savings of several billions of dollars. However, the end of the cold war and changing requirements resulting from Desert Storm led to down-scoping and, somewhat later, termination of a proposed $25 million dollar advanced development effort. As has been seen in this paper, the goal of 4:1 improvement will be extremely difficult to achieve. But, the requirements have changed and the area of warfare regions has been reduced by orders of magnitude. Additional technology, such as satellite based cueing, is available to assist with target location, and GPS can navigate the munition to the target. Much work remains, however, in the areas of target detection, classification, and acquisition. Current laboratory research programs will establish the base to minimize the performance risks associated with the acquisition and tracking algorithms. Combined with recent hardware advances, MMW/IR sensor fusion systems should be considered a low risk and affordable technology by the late 1990's.

References

1. Hacket, J., " Employ UAVs in Scud Hunt," Defense News, August 30-September 5, 1993, Page 19.

2. Barkett, F. T., et. al., "Guidance Apparatus With Dual Mode Sensor," Patent Application S.N. 634,548, July 26, 1984.

3. Sundstrom, B., et. al., "Joint USAF/Army Dual Mode IR/MMW Seeker Development Program," Proceedings of the NATO Defense Research Group Symposium on Terminally Guided Weapons, Brussels, Belgium, 15-17 September 1986.

4. J. S. Watson, D. H. Harrison, and S. Amphay, "The Application of Model Based Vision to MMW/IR Sensor Fusion," Symposium of Sensor Fusion, April 1992.

5. J. S. Watson, B. D. Williams, and N.A.D. Thompson, "MACET Sensor Fusion Algorithm Testbed," Symposium of Sensor Fusion, April 1993.

6. J. S. Watson, D. Harrison and G. Larson, "Irma Millimeter Wave Channel Upgrade Study," Nichols Research Corporation, Technical Report NRC-TR-91-053, Contract F08635-91-C-0110, 15 March 1991.

7. S. S. Steadman, "Recommendations for Passive Millimeter Wave Modeling," Technical Report NRC-TR-92-129, Contract F08635-91-C-0110, 12 June 1992.

8. D. Flynn, D. Harrison, and E. Martinez, "The Latest Advancements in Irma Model Development," 1993 Ground Target Modeling and Validation Conference, August 1993.

9. M. A. Richards, "Irma Active MMW Channel Upgrade Investigation," Technical Report NRC-TR-93-042, Contract F08635-91-C-0110, 1 May 1993.

10. M. Richards, S. Talele, and L. Love, "CCD Toolbox Radar Module Development Task," Technical Report NRC-TR-93-103, Contract F08635-91-C-0110, 14 July 1993.

11. Knox, Robert C., "Low Cost Dual Mode Seeker," Final Report WL-TR-92-7014, Wright Laboratory, Armament Directorate, AFMC, Eglin AFB, FL, December 1992.

SECTION II

DATA ASSOCIATION AND TRACKING TECHNIQUES

INTRODUCTION

by

Dr. David F. Liang
Head, Space Systems and Technology
Defence Research Establishment Ottawa
Department of National Defence
Shirley Bay, Ottawa
Canada K1A OZ4.

Tracking and fusion with multiple sensors deals with integration and correlation of data from diverse sources in order to arrive at the best possible situational assessment. The developments of data association and tracking algorithms have been driven by improved sensors, demanding mission requirements, target low observability, processor hardware limitations, severe clutter environment and challenging threat scenarios.

In general, there are two distinct approaches to the Data Association problem. The simpler approach is a deterministic one which includes nearest neighbor (NN) and global nearest neighbor data association. It takes the most likely of several possible "associations," and completely ignores the possibility that this selected "association" may be inappropriate. The alternative is the probabilistic approach based on Bayesian framework, which includes Probability Data Association (PDA), Joint Probabilistic Data Association (JPDA) and Multiple Hypothesis Tracking (MHT). S. S. Lim and D. F. Liang have provided a practical overview of multiple target tracking algorithms, including instruction for step by step implementation. The advantages and limitations of these techniques were summarized. For practical implementation of the algorithms, it is important to select them based on target dynamics, processor throughput, clutter environment, etc.

Then we deal with a track initiation problem of a ballistic missile using angle-only measurements. This problem is normally constrained by poor target-motion observability resulting in a very ill-conditioned estimation problem. M. Yeddanapudi, Y. Bar-Shalom, K. R. Pattipati and S. Deb have presented estimation algorithms that can handle far greater ill-conditioned problems than before and are more robust to errors in the initial estimates. Expressions for the Cramer-Rao lower bound (CRLB) on the covariance are presented with or without prior information. The use

of data from more than one satellite has been shown to significantly improve both target observability and trajectory estimates.

Air defence tracking problems using multi-radar defence systems have been known to be more complex than single sensor systems. Two separate procedures are presented. S.S. Blackman, R.J. Dempster and T.S. Nichols have presented a method for converting multi-radar plot data to a common stereographic coordinate system. They provided the mathematical basis and the implementation logic for the track-oriented MHT algorithms. This includes a discussion on clustering, pruning and merging methods that have made real-time processing implementable. The simulation results presented here based on the MHT approach, are shown to have superior track confirmation and maintenance capability than similar solutions using the conventional single hypothesis tracker.

On the other hand, K.C. Chang and Y. Bar-Shalom have proposed a track-oriented approach based on a form of "greedy" nearest-neighbor and multiple model algorithms. Data collected from multiple sensors are pooled together in a centralized fusion architecture. Tracks are initiated based on a single measurement and a probabilistic track score is calculated based upon the associated measurement history using a multiple model algorithm with an underlying Markov chain. Tracking results with multiple MTI radars are presented to demonstrate the feasibility of the algorithm.

In order to combine diverse data from sensors, weapons subsystems and other command and control sources to form a wide-area surveillance picture, it is crucial to have effective track correlation and management schemes. M.P. Dana and J.L. Dana have presented a track management method for a distributed, multiple sensor tracking systems. To avoid ambiguous or conflicting decisions at distinct C2 and weapon subsystems, it is important to maintain a common data base of high quality track information at multiple, distributed subsystems. They have provided a top-down system analysis approach to the design of track correlation and maintenance logic for multiple radar surveillance systems. This is illustrated by design problems derived from NATO air defence systems. The analytic technique permits statistical performance evaluation in terms of random, systematic and system errors.

A PRACTICAL OVERVIEW OF MULTIPLE TARGET TRACKING ALGORITHMS

S. S. Lim
System Concept Inc.
112 Grassy Plains Dr.
Kanata, Ontario
Canada, K2M 2M5

D. F. Liang
Defence Research Establishment
Ottawa
Department of National Defence
Ottawa, Ontario
Canada, K1A 0Z4

SUMMARY

Multiple Target Tracking (MTT) in cluttered environments has been addressed in numerous military aerospace defence and civilian air traffic control studies [1979-present]. Most MTT methods employ models based on simplified assumptions. In this paper, a practical overview of three algorithms for MTT is presented: these are the Nearest-Neighbor (NN), Multiple Hypothesis Tracking (MHT), and Joint Probabilistic Data Association (JPDA) methods. The advantages and limitations of each technique are summarized and suggestions presented for implementation considerations. It should be noted that an efficient Modified MHT applied to Air Defence Tracking is also presented in this AGARDOGRAPH.

1. INTRODUCTION

1.1 Background

Typical sensor systems, such as radar, are subject to noisy measurements from diverse sources, i.e., targets of interest and background noise sources such as radar ground clutter. The tracking problem involves the processing of measurements from targets of interest and produces, at each time step, an estimate of the target's position and velocity vectors. A common and versatile approach to such a problem involves an assumption that the target dynamics and measurements are both corrupted by additive, white, possibly Gaussian noise. The solution can be obtained from the Kalman filter, whereas the uncertainties in the target motion and measurement data (modeled as additive random noise) lead to corresponding uncertainties in the target state estimate.

There are additional uncertainties regarding the source of the measured data, which may include measurements from the target(s) of interest, interfering targets or random clutter. This leads to the problem of data association, where tracking performance depends not only upon the noise covariances, but also upon the amount of uncertainty in the measurement origin. Clearly, in this situation the tracking effort for N targets can be computationally more demanding than N times the effort for a single target, because it is not trivial to establish appropriate correspondence between targets and measurements. In [1], Blackman presents the state of the art algorithms that have been implemented in Multiple Target Tracking applications. Bar-Shalom and Fortmann's book [2] presents the mathematical tools underlying various algorithms for measurement association and Multiple Target Tracking. Two recent books edited by Bar-Shalom [3, 4], present various applications and current advances in data association and MTT as well as sensor fusion.

Although there are a number of algorithms that deal with the data association problem, two sets of algorithms have been derived based on fundamentally different models. The first is a deterministic model, i.e., one that takes the most likely of several "candidate" associations and treats them as if they were correct pairing, ignoring the fact that they may not be appropriate. The results of the deterministic association are then used in a standard state estimation algorithm. Examples include the Nearest Neighbor Standard Filter (NNSF) and the Track Splitting Filter (TSF) algorithms [5, 6]. The second model is a probabilistic model, utilizing a Bayesian framework in which the probabilities of individual associations are computed and then used in suitably modified state estimation algorithms. The examples are the Multiple Hypothesis Tracking (MHT)

approach and the Joint Probabilistic Data Association Filter (JPDAF). In the MHT, a number of candidate hypotheses are generated and evaluated until a correct decision can be made as more data become available.

The sequential nearest neighbor standard filter (NNSF) is the most simple and intuitive approach. It goes for unique pairings, so that, at most, one nearest observation can be used to update a given track. The assignment is made irrevocable after each scan of data is received. Although NNSF is simple, the major problem with choosing the nearest neighbor is that, with some probability, it is not the correct measurement while believing that the assignment has been correctly made. This is because the filter calculated error covariance does not account for the possibility of processing an incorrect measurement, and it can lead to the subsequent loss of track.

The MHT approach makes use of the assignment constraint and the formation and evaluation of alternative correlation hypotheses. In this approach, as each measurement is being received, data association hypotheses are generated together with the computation of their probabilities. By generating association hypotheses and deferring the decision until more measurements are available, the MHT algorithm can improve the problem of mis-correlation and loss of tracks. This leads to the more accurate and much more complex algorithm. Further, the MHT is capable of initiating tracks, accounting for missing or false reports, and of processing sets of dependent reports. In this regard, the MHT has been recognized as the "best" technique throughout the target tracking community. However, a major drawback to the MHT algorithm is its computational complexity due to the large number of hypotheses, which grow exponentially with time. Hence, an MHT implementation is limited by the extent of effective pruning of unlikely hypotheses [7-10]. It should be noted that in Section V of this AGARDOGRAPH, a modified MHT has been developed and shown to be effective in air defence radar surveillance system tracking assessments.

For the JPDA approach, all the reports in the vicinity of a given track are employed to form a weighted average and the result is used to update the track. The JPDA algorithm is a non-scan back approach, meaning that all hypotheses are combined after the computation of probabilities, for each target at each time step. Therefore, any given tracking filter will likely be assigned with a high weighting several "correct" measurements. However, it will also be assigned to the "incorrect" measurement with a low weighting. Since all reports are used in computing the track update, it implies that incorrect measurements are routinely used by the tracking filter, albeit with a lower weighting than that for a correct measurement. As well, the JPDA approach does not provide a track initiation procedure.

1.2 Description of Targets and Measurements

For a target being tracked, the discretized equations of motion may be modeled by

$$x(k+1) = F(k)x(k) + G(k)u(k) \qquad (1)$$

where $x(k)$ is the n dimensional state vector of the tracked targets at the k-th sample time, $F(k)$ is the transition matrix, and $u(k)$ is an m dimensional state excitation vector to account for both maneuvers and modeling errors and is generally assumed to be white Gaussian with zero mean and covariance $Q(k)$. In a track-while-scan system, the k-th sample will occur approximately at time kT, where T is the scan interval of the sensor. The measurement equation representing valid sensor observations of the targets being tracked has the form

$$y(k) = H(k)x(k) + v(k) \qquad (2)$$

where $y(k)$ is the m dimensional sensor measurement vector, H is the measurement matrix and $v(k)$ is white Gaussian measurement noise with zero mean and covariance $R(k)$. The measurement equation for extraneous sensor reports resulting from thermal false alarms, clutter, and other targets is assumed to satisfy

$$y(k) = H(k)\hat{x}(k,k-1) + w(k) \qquad (3)$$

where $w(k)$ is assumed white and uniformly distributed over some volume V of the measurement space centered about the predicted measurement

$$\hat{y}(k,k-1) = H(k)\hat{x}(k,k-1) \qquad (4)$$

The number of such extraneous reports in any volume V obeys a Poisson distribution with mean V, where is the unnormalized extraneous report density.

The tracking filter provides a state estimate $\hat{x}(k,k)$ and one - scan - predicted state

$\hat{x}(k+1,k)$ given all measurement data up to the time k. The basic filter equations are described in the following Sections.

2. THE NEAREST NEIGHBOR STANDARD FILTER ALGORITHM

2.1 Introduction

The nearest neighbor algorithm assumes that the best correlation is the one that associates reports and tracks closest to each other. A given observation can only be paired once, to either a previously established track or to initiate a new track. The report can be abandoned as noise if the measurement-to-track pairings cannot satisfy a preliminary gating test. The NNSF algorithm presented here is based on the modified Munkres optimal assignment method, as modified by Burgeois and Lassalle [8].

2.2 NNSF Algorithm

STEP 1: Start with $k = 1$ and initialize all the system parameters: F, H, R, Q.

STEP 2: Simulate target trajectories and sensor measurements

STEP 3: Measurement validation: A measurement is considered valid if

$$\{y(k)-\hat{y}(k,k-1)\}'S^{-1}(k)\{y(k)-\hat{y}(k,k-1)\} \le g^2,$$

where "g- ∂ " forms an elliptical gate. (5)

STEP 4: The elements of the assignment matrix are equal to the normalized distance function associated with the assignment of each of N_{ob} measurements assigned to each of N_{tr} tracks. If the measurements do not fall within the gate, the measurement-to-track pairing can be penalized by giving it a very large distance assignment.

STEP 5: The modified Munkres optimal assignment algorithm presented in the subsequent Section can be adopted to solve the Assignment Matrix by minimizing the normalized distance function.

STEP 6: Measurements are to be correlated to the tracks using the optimal assignment matrix.

STEP 7: Apply the Kalman filter to update and predict the state vectors (tracks) using :

$$A(k) = P(k,k-1)H'(k)[H(k)P(k,k-1)H'(k)+R(k)]^{-1} \tag{6}$$

$$\hat{x}(k,k) = \hat{x}(k,k-1)+A(k)[y(k)-H(k)\hat{x}(k,k-1)], \tag{7}$$

$$P(k,k) = (I-A(k)H(k))P(k,k-1), \tag{8}$$

$$\hat{x}(k+1,k) = F(k)\hat{x}(k,k), \tag{9}$$

$$P(k+1,k) = F(k)P(k,k)F'(k)+G(k)Q(k)G'(k). \tag{10}$$

STEP 8: If $k = k_f$ tracking to be stopped. Otherwise, set $k = k + 1$, and output the results and go to STEP 2.

2.3 The Modified Munkres Optimal Assignment Algorithm.

The practical advantage of this algorithm is the fact that the assignment matrix need not be square. For the convenience of presentation of the algorithm, the rows and columns of the matrix may be distinguished by being starred (*) or primed(').

The Optimal Assignment Algorithm [8]:

STEP 1: Let v = min. { No. of columns, No. of rows}. Initially, no lines are covered; no zeros are starred or primed.

STEP 2: If the number of rows is greater than the number of columns, go to STEP 5.

STEP 3: For each row of the matrix (a_{ij}), subtract the value of the smallest element from each element in the row.

STEP 4: If the number of columns is greater than the number of rows, go to STEP 6.

STEP 5: For each column in the matrix, subtract the smallest element of the column from each component of the column. Then go to STEP 6.

STEP 6: 1) Find a zero, Z, of the matrix.
2) If there is no starred zero in its column nor its row, star Z.
3) Repeat for each zero of the matrix. Go to STEP 7.

STEP 7: 1) Cover every column containing a 0*.
2) If all columns are covered, the starred zeros form the desired independent set. The algorithm is now completed. Otherwise, continue to the next step.

STEP 8: 1) Choose a noncovered zero and prime it; then consider the row containing it.
2) If there is no starred zero Z in this row, go to STEP 9.
3) If there is a starred zero Z in this row, cover this row and uncover the column of Z.
4) Repeat until all zeros are covered and then go to STEP 10.

STEP 9: There is a sequence of alternating starred and primed zeros constructed as follows:
1) Let Z_0 denote the uncovered 0'.
2) Let Z_1 denote the 0* in the column of Z_0 (if any).
3) Let Z_2 denote the 0' in the row of Z_1.
4) Continue in a similar way until the sequence stops at a 0', Z_{2k}, which has no 0* in its column.
5) Unstar each starred zero of the sequence.
6) Star each primed zero of the sequence.
7) Erase all primes and uncover every line.
8) Go to STEP 7.

STEP 10: 1) Let h denote the smallest uncovered element of the matrix; and it will be positive.
2) Add h to each covered row.
3) Subtract h from each uncovered column.
4) Go to STEP 8 without altering any stars, primes, or covered lines.

3. MHT ALGORITHM FOR MULTIPLE TARGET TRACKING

3.1 Introduction

The multiple hypothesis tracking algorithm was first proposed by Reid in 1979 [9]. In the MHT, instead of trying to resolve a difficult association immediately or sequentially, all possibilities are enumerated as hypotheses. The probability of each hypothesis is computed, and ideally, all hypotheses are maintained until there is enough data to make a decision as to the correct one. This algorithm contains a number of advantages over the sequential techniques discussed above. Some of these advantages are that: (1) track initiation is an integral part of the algorithm, (2) maneuver detection can be easily handled within the basic algorithm, and (3) one of the hypotheses must be the correct one. Its main drawback is that it does not generate a single best solution which can be returned to the operator directly but rather recommends a set of alternative solutions with different quality which must be evaluated outside of the algorithm. Another problem is that the algorithm is extremely demanding in computer resources because it must maintain all hypotheses to arrive at an optimal solution. Most often, a sub-optimal implementation which includes a mechanism for pruning unlikely hypotheses is considered [2, 3, 7, 9-10]. Recently Lim, Liang and Blanchette [11] proposed a modified MHT (MMHT) algorithm which has been extensively tested against numerous sets of real radar data. In this Section an algorithm based on the MMHT is presented.

3.2 Flow Chart.

The main part of the algorithm consists of clustering, hypothesis generation, pruning by dynamic thresholding and N scan back approach. The cluster (CLUST) subroutine associates measurements with the previous clusters. A cluster is a group of hypotheses containing associated tracks that do not interact with any other group of hypotheses within other clusters. The hypotheses within a cluster will not share measurements with the hypotheses of any other clusters. The basic goal of clustering is to divide the large tracking problem into a number of smaller ones that can be solved independently. The hypothesis generation (HGEN) subroutine creates new data association hypotheses for the set of validated measurements of each cluster. The probability of individual hypothesis is then computed in the PROB subroutine. Both the clustering and hypothesis generation procedures use the pruning subroutine to reduce the number of hypotheses which grow rapidly with time. The hypotheses satisfying certain qualifications will remain and all the unlikely hypotheses are eliminated in the PRUN subroutine. Every N scan, only the most likely hypotheses are selected and the other hypotheses are all eliminated from the subsequent considerations, as explained in the above. As part of the pruning process similar hypotheses may be combined into one. The subroutine FILT computes the estimate of each

track based on the previous estimate and a new measurement using a standard Kalman filter. The major distinction of the MMHT from Reid's MHT is with the two pruning schemes, i.e., dynamic thresholding and N-scan back pruning (the subroutine PRUN in Fig. 1).

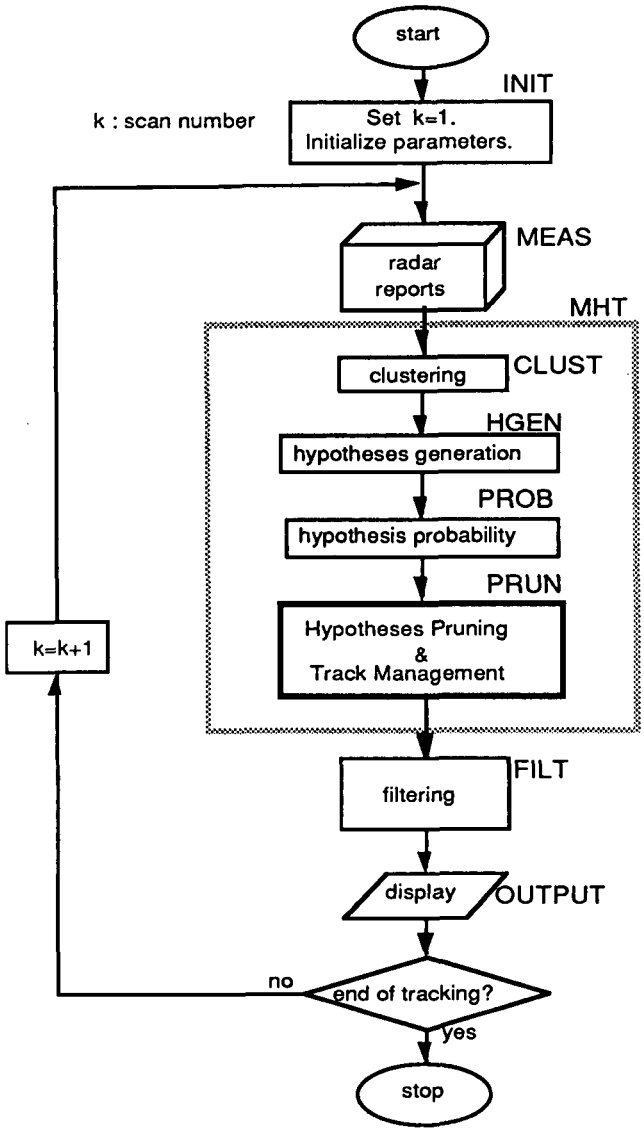


Fig. 1 The Multiple Hypothesis Test algorithm.

3.3 Detailed MHT Algorithm

STEP 1: Set $k = 1$ and initialize all the system parameters: F, Q, H, R.

STEP 2: Simulate target trajectories.

STEP 3: Simulate measurements. (when it is implemented, the data are received from the sensor.)

STEP 4: For each hypothesis h_ℓ, $1 \le h_\ell \le L(k-1)$, compute (from STEP 10): $P_\ell(k,k-1)$, $K_\ell(k-1)$, $\hat{x}(k,k-1)$, $P(k,k-1)$, $P_r(k-1)$.

STEP 5: Validate measurement using the elliptical gating test of Equation (5) to select N_k sensor reports for use in filter update.

STEP 6: 1) Set new hypotheses h_ℓ, for $1 \le h_\ell \le (1+N_k)L(k-1)$.

2) Compute $P_r(k)$ of the hypothesis h_ℓ with measurement $Y^k = \{Y(1), Y(2), \ldots, Y(k)\}$ using

$$P_r^\ell(k) \equiv P_r(h_\ell(k)|Y^k)$$
$$= \frac{1}{c} P_D^{N_{DT}} (1-P_D)^{(N_{TGT}-N_{DT})} \beta_{FT}^{N_{FT}} \beta_{NT}^{N_{NT}}$$
$$\left[\prod_{m=1}^{N_{DT}} N(y_m - H\hat{x}, S)\right] P_r^\ell(k-1), \tag{11}$$

The c is a normalization constant and $P_r^\ell(k-1)$ is the probability of the hypothesis $h_\ell(k-1)$. $N(x,S)$ denotes the normal distribution given by

$$N(x,S) \equiv \exp\left(-0.5x' S^{-1} x / \sqrt{(2\pi)^n |S|}\right),$$
$$S = H\bar{P}H' + R \tag{12}$$

where $\bar{P}$ is the error covariance of the target estimate for the prior hypothesis $h_\ell(k-1)$ and R is the measurement noise covariance.

STEP 7: For each hypothesis h_ℓ, compute

$$P_h(k,k-1) = P_\ell(k,k-1) - P_\ell(k,k-1) * H'(k) S^{-1}(k) H(k) P_\ell(k,k-1) \tag{13}$$

$$K_h(k,k-1) = K_\ell(k,k-1) + P_\ell(k,k-1) * H'(k) S^{-1}(k,k-1) * [y(k) - \hat{y}(k,k-1) - H(k) K_\ell(k,k-1)] \tag{14}$$

or if no measurement of the target was received
in the gate,

$$K_h(k,k-1) = K_\ell(k,k-1), \quad (15)$$

$$P_h(k,k-1) = P_\ell(k,k-1). \quad (16)$$

STEP 8: For each hypothesis h_ℓ,

Compute $\mathbf{A}(k) = \sum_{h=1}^{L(k)} \left[P_r^h(k) K_h(k,k-1) \right]$. (17)

Compute $P_h(k,k) = P_h(k,k-1)$ (18)

Compute $K_h(k,k) = K_h(k,k-1) - \mathbf{A}(k)$. (19)

STEP 9: Calculate the new estimates:

$$\hat{x}(k,k) = \hat{x}(k,k-1) + \mathbf{A}(k) \quad (20)$$

Compute error covariance:

$$P(k,k) = \sum_{h=1}^{L(k)} [P_r^h(k)\{P_h(k,k-1) + K_h(k,k-1)$$

$$K_\ell(k,k-1)\}] - \mathrm{A}(k)\mathrm{A}'(k) \quad (21)$$

STEP 10: Prediction for the next update:

Compute $K_h(k+1,k) = F(k)K_h(k,k)$ (22)

$$P_h(k+1,k) = F(k)P_h(k,k)F'(k) + G(k)Q(k)G'(k) \quad (23)$$

Compute $\hat{x}(k+1,k) = F(k)\hat{x}(k,k)$ (24)

Compute

$$P(k+1,k) = F(k)P(k,k)F'(k) + G(k)Q(k)G'(k) \quad (25)$$

STEP 11: If $k = k_f$ (end of tracking), stop. Otherwise, set $k = k + 1$, output the results of tracking, and go to STEP 3.

Remark: Equation (11) was derived by Reid [9] under the assumption that the number of such extraneous reports in any volume C obeys a Poisson distribution with mean bC, where b is the normalized extraneous report density. Further, the formula is valid for type 1 sensors such as radar. If type 2 sensors are assumed, some modifications have to be made as given in [9]. Similar expression for the computation of hypothesis probabilities in the nonrecursive form can be found in Bar-Shalom [2].

4. JOINT PROBABILISTIC DATA ASSOCIATION FILTER

4.1 Introduction

The JPDA algorithm belongs to all-neighbor class of target tracking techniques, since all measurements reasonably close to a given track position are incorporated to form a weighted average, and the result is used to update the tracks. The weights are based on the probability that the report originating from the target is a statistical function of report-to-track distance for all observations within the validation region. In fact, the JPDA algorithm can be interpreted as a one-scan memory, track oriented with no new target assignments and non-measurement oriented merging to one hypothesis. Further, this is a target-oriented approach, in the sense that a set of established targets is used to form gates in the measurement space and to compute posterior probabilities, in contrast to the measurement-oriented algorithm such as the MHT.

4.2 JPDAF Algorithm

The JPDAF algorithm [2] is presented as follow:

STEP 1: Set $k = 1$ (time index) and initialize all the system parameters: F, H, R, Q.

STEP 2: Simulate target trajectories.

STEP 3: Simulate measurements. (when it is implemented, receive the data from the sensor.)

STEP 4: Validate the measurements using "g- ∂ ellipsoid" gating test (Equation (5)).
Set the validation matrix $\mathbf{W}(\eta)$ [13]:

$$\mathbf{W}(\eta) = \{w(j,t;\eta);$$

$$\text{for } j=1,2,\ldots,N_m;\ t=1,2,\ldots,N_t\} \quad (26)$$

$w(j,t;\eta) = 1$ if the event "η" occurs (if measurement j falls within the gate of target t),
$= 0$ otherwise.

STEP 5: Compute the residual for each $t=1,2,\ldots,N_t$:

$$\tilde{y}_j^t(k) = y_j(k) - \hat{y}^t(k) \quad (27)$$

where $\hat{y}^t(k) = H(k)\hat{x}_t(k,k-1)$, corresponding to the target t. (28)

STEP 6: Compute the covariance of $\tilde{y}_j^t(k)$:

$$S_j^t(k) = H(k)P(k,k-1)H'(k) + R(k). \quad (29)$$

STEP 7: Compute the joint event probabilities using Bayes' formula:

$$P_r(\eta) = prob\{\eta | Y(k)\}$$
$$= C^e / c \prod_{j=1}^{N_m(j)} [-0.5(\tilde{y}_j^t)'(S_j^t)^{-1}(\tilde{y}_j^t) / \{(2\pi)^{\frac{N_m}{2}}$$
$$|S_j^t|^{0.5}\}] \prod_{t=1}^{N_{to}(t)} [1 - P_D^t] \quad (30)$$

where S^{-1} is the inverse of $S_j^t(k)$ at scan k.

STEP 8: For t=0, compute:

$$\theta_0^t(k) = 1 - \sum_{j=1}^{N_m} [\theta_j^t(k)]. \quad (31)$$

For each target t=1,2,....,N_t, compute

$$\theta_j^t(k) = \sum_{i=1}^{N_m} [P_r(j) w(j,t;i)] \quad (32)$$

STEP 9: For each t = 1,2,....,N_t, compute

$$\tilde{y}^t(k) = \sum_{j=1}^{N_m} [\theta_j^t(k) \tilde{y}_j^t(k)] \quad (33)$$

STEP 10: Compute the filter gain $\mathbf{A}(k)$:

$$\mathbf{A}(k) = P(k,k-1) H'(k) S^{-1}(k) \quad (34)$$

STEP 11: The state and covariance update for each t = 1, 2, ..., N_t,

$$\hat{x}(k,k) = \hat{x}(k,k-1) + \mathbf{A}(k)\tilde{y}^t(k) \quad (35)$$
$$P(k,k) = P(k,k-1) - \mathbf{A}(k) S_j^t(k) \mathbf{A}'(k). \quad (36)$$

STEP 12: State and covariance prediction:

$$\hat{x}(k+1,k) = F(k)\hat{x}(k,k), \quad (37)$$
$$P(k+1,k) = F(k)P(k,k)F'(k) + G(k)Q(k)G'(k) \quad (38)$$

STEP 13: If $k = k_f$ (end of tracking), stop. Otherwise, set $k = k + 1$, output the results, and go to STEP 3.

Remark: The joint event probabilities are computed from Equation (30), derived from the assumption that the probability mass function (PMF) of the number of false measurements (clutter points) is given by the Poisson density [2, p. 168]:

$$\mu(\varepsilon) = \exp(-\lambda V) \frac{(\lambda V)^{\varepsilon}}{\varepsilon!} \quad (39)$$

where ε is the number of false measurements, λ is the spatial density of false measurements (i.e., the average number per unit volume), and $V(k)$ is the volume of the validation region. $\lambda * V(k)$ is the expected number of false measurements within the gate. Thus If the PMF is uniformly distributed, then the probability should be modified accordingly ([2], p.228).

5. SUGGESTIONS

This paper has presented three algorithms based on NNSF, MHT, and JPDA methods for multiple target tracking in clutter environments. For practical implementation of the algorithm, it is important to compare these algorithms in view of the target dynamic models and clutters:

1) In a dense target environment, the NNSF algorithm suffers severe performance degradation, since it forces miscorrelation by choosing (or pairing) a nearest report as the correct track for a target. The greatest advantage is its ease of implementation and its computational efficiency.

2) One of the greatest advantages of the MHT algorithm, is its intrinsic ability to initiate tracks, and to account for missing or false reports. Its major drawback is its computational complexity due to the number of hypotheses that grow exponentially with time. Hence, the MHT implementation depends on effective pruning of unlikely hypotheses [2, 3, 7, 9-12].

3) The greatest drawback of the JPDA algorithm is its inability to initiate tracks automatically. This is a non back-scan (or zero-scan) algorithm, in which all hypotheses are combined into one in every scan. This approach is more cost-effective than an n-scan algorithm, especially in a heavy clutter environment.

Therefore, for practical implementaion involving targets in a dense environment, one should look for an effective way of combining both JPDA and MHT.

6. REFERENCES

[1] Blackman, S.S., *Multiple Target Tracking with Radar Applications*, Artech House, Dedham, MA, 1986.

[2] Bar-Shalom, Y. and Fortmann, T., *Tracking and Data Association*, Artech House, Norwood, MA, 1990.

[3] Bar-Shalom, Y. (ed.), *Multitarget-Multisensor Tracking: Advanced*

Applications, Artech House, Norwood, MA, 1990.

[4] Bar-Shalom, Y. (ed.), *Multitarget-Multisensor Tracking: Applications and Advances*, Vol. II, Artec House, 1992

[5] Kountzeris, A., *A Track Splitting Algorithm for Multiple Target Tracking*, University of Sussex, School of Engineering and Applied Sciences, Report CE/S/43, Oct. 1989.

[6] Smith, A. and Buechler, G., "A Branching Algorithm for Discriminating and Tracking Multiple Objects," IEEE Trans. on Automatic Control, vol. AC-20, pp. 101-104, Feb. 1975.

[7] Kovacich, M.,etc., "An Application of MHT to Group to Object Tracking," SPIE Vol. 1481, Signal and Data Processing of Small Targets, pp. 357-370, 1991.

[8] Burgeois, F. and Lassalle, J.C., " An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices," Communications of the ACM, Vol.14, pp.802-806, Dec. 1971.

[9] Reid, D.B., "An Algorithm for Tracking Multiple Targets," IEEE Trans. Automat. Contr., Vol. AC-24, pp.843-854, Dec.1979.

[10] Danchick, R. and Newmam, G.E., "A Fast Method for Finding the Exact N-best Hypotheses for Multitarget Tracking," IEEE Trans. on Aerospace and Electronic Systems, vol. AES-29, pp. 555-560, April 1993.

[11] Lim, S.S, Liang, D.F. and Blanchette, M., "Air Defence Surveillance System Tracking Assessment," Proc. of the AGARD Conf. on Guidance and Control for Future Air-Defence Systems, Copenhagen, Denmark, May 17-20, 1994, AGARD-CP-555 (Supplement: Classified papers), pp.24.1-24.14, July 1995.

[12] Singer, R.A. and Sea, R.G., "New Results in Optimizing Surveillance System Tracking and Data Correlation Performance in Dense Multitarget Environments," IEEE Trans. Automat. Contr., Vol. AC-18, pp.571-581, Dec. 1973.

APPENDIX: NOTATIONS

k; time or scan index.

k_f ; final time of tracking

$N(x,S) \equiv \exp\left(-0.5x' S^{-1}x / \sqrt{(2\pi)^n |S|}\right)$

(see Equation (12)

n dimension of the state vector x.

m; dimension of the measurement vector y

N_{DT}; the number of measurements associated with the prior targets.

N_{FT}; the number of measurements associated with false targets.

N_{NT}; the number of measurements associated with new targets.

N_{TGT}; the number of previously known targets within the area of coverage of the sensor.

P_D; probability of detection.

β_{FT}; density of false targets.

β_{NT}; density of previously known targets that have been detected.

c; the normalization constant and given by summation of all the hypothesis probabilities.

$P_r(k)$; probability of a hypothesis at time k given probability of the prior hypothesis.

$L(k)$; total number of hypotheses at k and given by $(1+N_k)L(k-1)$ for N_k the number of sensor reports that fall within the gate at scan k.

$P_\ell(k,k-1)$ covariance of the estimation error given observations through scan k-1 and given track hypothesis "ℓ".

$K_\ell(k,k-1)$ mean of the estimation error given observations through scan k-1 and given track hypothesis "ℓ".

$P_h(k,k-1), K_h(k-1)$; the $P_\ell(k,k-1)$ and $K_\ell(k-1)$ analogs corresponding to the hypothesis "h".

$\hat{x}(k,k-1)$ estimated state at scan k given hypothesis up to time k-1.

$P(k,k)$ covariance of the estimation error given observations up to scan k.

$\mathrm{A}(k)$; the optimal tracking filter correction vector for MHT filter.

$A(k)$; the optimal tracking filter gain for NNSF.

η; the event that an observation j belongs to a target t.

Y^k; the set of measurements up to time k.

N_m number of measurements.

N_t; number of targets.

$N_{ta}(j)$;number of targets associated with the observation j except for the false alarms (false targets).

$N_{td}(t)$; number of observations associated with the target t

$N_{to}(t)$; number of observations not associated with the target t.

N_{cr}; (No. of columns of $\mathbf{W}(\eta)$) x (No. of rows of $\mathbf{W}(\eta)$), where $\mathbf{W}(\eta)$ is validation matrix given by Equation (26).

Ballistic Missile Track Initiation from Satellite Observations with Extrapolation to Impact*

Murali Yeddanapudi, Yaakov Bar-Shalom, Krishna R. Pattipati and Somnath Deb
Department of Electrical & Systems Engineering, University of Connecticut, Storrs, CT 06269-3157
e-mail: ybs@ee.uconn.edu

1. SUMMARY

This chapter presents an algorithm to initiate tracks of a ballistic missile in the initial exoatmospheric phase, using line of sight measurements from one or more moving platforms (typically satellites). The major feature of this problem is the poor target motion observability which results in a very ill-conditioned estimation problem.

The Gauss-Newton iterative least squares minimization algorithm for estimating the state of a nonlinear deterministic system with nonlinear noisy measurements has been previously applied to the problem of angles-only orbit determination using more than three observations. A major shortcoming of this approach is that convergence of the algorithm depends strongly on the initial guess. By using the more sophisticated Levenberg-Marquardt method in place of the simpler Gauss-Newton algorithm and by developing robust new methods for obtaining the initial guess in both single and multiple satellite scenarios, the above mentioned difficulties have been overcome. In addition, an expression for the Cramer-Rao lower bound on the error covariance matrix of the estimate is derived.

We also incorporate additional partial information as an extra pseudo-measurement and determine a modified maximum likelihood estimate of the target state and the associated bound on the covariance matrix. In most practical situations, probabilistic models of the target altitude and/or speed at the initial point constitute the most useful additional information.

Monte Carlo simulation studies on some typical scenarios were performed, and the results indicate that the estimation errors are commensurate with the theoretical lower bounds, thus illustrating that the proposed estimators are *efficient*.

2. INTRODUCTION

In this chapter we consider the estimation of the state of a target — a ballistic missile — at a specified time instant during the initial phase of its exoatmospheric flight. During this (ballistic) phase of the target's flight, the predominant force acting on the target is Earth's gravity, while the effect of atmospheric drag and the gravitational forces due to other celestial bodies can be ignored. Under these assumptions, the equations of motion of the target are given by the laws of Keppler [1]; hence, the target dynamics are described by a zero input deterministic system. The state of the target is a six dimensional vector comprised of the three position and the three velocity components. Since the target dynamics are completely deterministic, the estimation of the target's exoatmospheric trajectory is equivalent to the estimation of the state of the target at some reference time instant t_0.

The measurements originate from sensors which are strategically located so as to detect the target in the early stages of its exoatmospheric flight. These sensors could be ground-based observatories, airborne early warning systems or low-altitude Earth satellites. The sensors could be either passive, which can measure only the line of sight direction, or active, which also measure the distance. In this chapter we focus on line of sight (LOS) measurements made from passive sensors located on low-altitude Earth satellites. A line of sight measurement is a unit vector which can be represented using two angles: azimuth and elevation of the target as seen from the sensor. We need at least three LOS measurements to estimate the six dimensional state of the target[1].

An important consideration in this problem is the observability of the target through the LOS measurements. In general, the sensors are at quite a long distance from the target such that a significant change in the target position (over a period of time) is reflected only as a slight variation in the LOS measurements. This results in poor observability of the target motion. Consequently, the Maximum Likelihood (ML) trajectory estimation problem is very ill-conditioned [2]. While an initial state estimate might be obtained from the boost phase, this is not assumed to be available in the present problem formulation.

Previous work by Chang [3] proposed the use of a Gauss-Newton iterative least squares minimization algorithm for the problem of angles-only orbit determination using more than three observations. A major shortcoming of this approach (which is an in-

*Research supported by AFOSR Grant F49620-95-1-0229 and ONR/BMDO Grant N00014-91-J-1950.

[1] Of the three (or more) LOS measurements at least two should be at different time instants; otherwise, the target velocity would remain unobserved and, hence cannot be estimated.

herent limitation of the Gauss-Newton algorithm) is that the convergence of the algorithm depends strongly on the initial guess. In addition, the method for obtaining the initial guess suggested by Chang is applicable only for the single satellite scenario. A more recent paper by Danis [4] deals with the approximate evaluation of errors associated with the estimation of the launch point of a target from two LOS measurements and other trajectory information.

The main contributions of this chapter are: 1) The trajectory estimation algorithms can handle far more ill-conditioned problems than before and is more robust to errors in the initial estimate, 2) Trajectory estimation with *partial prior information*, 3) Expressions for the Cramer-Rao lower bound (CRLB) on the covariance of the target state with or without prior information, and 4) The actual covariance is shown to be commensurate with the CRLB.

The problem is formulated in Section **3.** Section **4** deals with the ML trajectory estimation. The problem of trajectory estimation with probabilistic prior altitude and speed information is considered in Section **5.** Results of computer simulations performed for some typical scenarios are presented in Section **6.**

3. PROBLEM FORMULATION

It is assumed that Earth's gravity is the predominant force acting on the target. In addition, if the non-spherical nature of the Earth is neglected, then the dynamics of the target are described by the following equation:

$$\ddot{\xi}(t) = \xi(t)\left[\frac{-\mu}{\|\xi(t)\|^3}\right] \tag{1}$$

where $\|\cdot\|$ denotes the Euclidean norm and $\xi(t) = [\xi_x(t)\ \xi_y(t)\ \xi_z(t)]'$ is the target position vector at time t in the inertial geocentric equatorial (GCE) coordinate system and μ is the Earth's gravitational constant. Let $\mathbf{x}(t)' = [\xi(t)'\ \dot{\xi}(t)']$ be the 6-dimensional target state vector at time t. The state propagation equation can be represented, using the state $\mathbf{x}_0' = [\xi_0'\ \dot{\xi}_0']$ at the reference time t_0 (i.e., $\mathbf{x}(t_0) = \mathbf{x}_0$) as

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{x}_0, t_0, t) \tag{2}$$

An algorithm for evaluating $\mathbf{f}(\cdot)$, (i.e., state propagation) is given in Appendix A. A LOS measurement made at time t_k by sensor $s(k)$ can be represented by a 2-dimensional vector $\mathbf{z}(k)$. Let N_s be the total number of sensors, indexed 1 through N_s and the number of LOS measurements be M. Then the measurement equation at time t_k (discrete time k) can be written as

$$\mathbf{z}(k) = \mathbf{h}\left(t_k, t_0, \mathbf{x}_0, \eta_{s(k)}(t_k)\right) + \mathbf{w}_{s(k)}(k)$$
$$k = 1, \ldots, M \quad 1 \le s(k) \le N_s \tag{3}$$

where $\mathbf{w}_{s(k)}(k)$ is the measurement noise vector in sensor $s(k)$. These noises are modeled as zero mean, white Gaussian random variables with known covariance matrices $R_{s(k)}$. The two components of the vector $\mathbf{h}\left(t_k, t_0, \mathbf{x}_0, \eta_{s(k)}(t_k)\right)$, which subsequently will be denoted by $\mathbf{h}_k(\mathbf{x}_0)$ for simplicity, are the azimuth $\phi_k(\mathbf{x}_0)$ and the elevation $\theta_k(\mathbf{x}_0)$ of the target as seen from sensor $s(k)$, namely,

$$\begin{aligned}\mathbf{h}_k(\mathbf{x}_0) &= [\phi_k(\mathbf{x}_0)\ \theta_k(\mathbf{x}_0)]' \\ &= \begin{bmatrix} \arctan\left\{\frac{\xi_{0_y}(t_k)-\eta_{s(k)_y}(t_k)}{\xi_{0_x}(t_k)-\eta_{s(k)_x}(t_k)}\right\} \\ \arcsin\left\{\frac{\xi_{0_z}(t_k)-\eta_{s(k)_z}(t_k)}{\rho_k(\mathbf{x}_0)}\right\}\end{bmatrix}\end{aligned} \tag{4}$$

In the above equation, the vector $\eta_{s(k)}(t_k) = [\eta_{s(k)_x}(t_k)\ \eta_{s(k)_y}(t_k)\ \eta_{s(k)_z}(t_k)]'$ is the position of sensor $s(k)$ at time t_k and is assumed to be known for all k; $\rho_k(\mathbf{x}_0)$ is the distance between the target and sensor $s(k)$ at time t_k, i.e.,

$$\rho_k(\mathbf{x}_0) = \|\xi_0(t_k) - \eta_{s(k)}(t_k)\| \tag{5}$$

The problem of ballistic trajectory estimation is then the estimation of the parameter $\mathbf{x}_0$ given the set of measurements $Z_M = \{\mathbf{z}(k),\ k = 1, \ldots, M\}$.

4. MAXIMUM LIKELIHOOD ESTIMATION

The conditional probability density of the measurement $\mathbf{z}(k)$, given that the target state at time t_0 is $\mathbf{x}_0' = [\xi_0'\ \dot{\xi}_0']$, is

$$p_k\left[\mathbf{z}(k)|\mathbf{x}_0\right] = \left|2\pi R_{s(k)}\right|^{-\frac{1}{2}}$$
$$\exp\left\{-\frac{1}{2}\left[\mathbf{z}(k) - \mathbf{h}_k(\mathbf{x}_0)\right]' R_{s(k)}^{-1}\left[\mathbf{z}(k) - \mathbf{h}_k(\mathbf{x}_0)\right]\right\} \tag{6}$$

The measurement noises of a sensor at different times as well as those of distinct sensors are all assumed to be mutually independent. The likelihood function $\Lambda_{Z_M}(\mathbf{x}_0)$ of $\mathbf{x}_0$ based on the measurement set Z_M is then the product of the individual probability density functions

$$\Lambda_{Z_M}(\mathbf{x}_0) = \left[\prod_{k=1}^{M}\left|2\pi R_{s(k)}\right|^{-\frac{1}{2}}\right]\cdot\exp\Bigg\{$$
$$-\frac{1}{2}\sum_{k=1}^{M}\left[\mathbf{z}(k) - \mathbf{h}_k(\mathbf{x}_0)\right]' R_{s(k)}^{-1}\left[\mathbf{z}(k) - \mathbf{h}_k(\mathbf{x}_0)\right]\Bigg\} \tag{7}$$

Let $\nu(\mathbf{x}_0)$ be the $2M \times 1$ vector of normalized measurement residuals, i.e.,

$$\nu(\mathbf{x}_0) = \begin{bmatrix} R_{s(1)}^{-\frac{1}{2}}\left[\mathbf{z}(1) - \mathbf{h}_1(\mathbf{x}_0)\right] \\ \vdots \\ R_{s(M)}^{-\frac{1}{2}}\left[\mathbf{z}(M) - \mathbf{h}_M(\mathbf{x}_0)\right]\end{bmatrix} \tag{8}$$

Using $\nu(\mathbf{x}_0)$ in the expression for $\Lambda_{Z_M}(\mathbf{x}_0)$, and replacing the constant terms in Eq. (7) by c, we obtain

$$\Lambda_{Z_M}(\mathbf{x}_0) = c\cdot\exp\left\{-\frac{1}{2}\nu(\mathbf{x}_0)'\nu(\mathbf{x}_0)\right\} \tag{9}$$

The maximum likelihood estimate of the target state $\mathbf{x}_0^{\mathrm{ML}}$ is obtained by minimizing the squared norm of $\nu(\mathbf{x}_0)$ over all possible values of $\mathbf{x}_0 \in \Re^6$, i.e.,

$$\mathbf{x}_0^{\mathrm{ML}} = \arg\max_{\mathbf{x}_0\in\Re^6}\Lambda_{Z_M}(\mathbf{x}_0) = \arg\min_{\mathbf{x}_0\in\Re^6} f(\mathbf{x}_0) \tag{10}$$

where $f(\mathbf{x}_0)$ is the squared norm of $\boldsymbol{\nu}(\mathbf{x}_0)$

$$f(\mathbf{x}_0) = \boldsymbol{\nu}(\mathbf{x}_0)'\boldsymbol{\nu}(\mathbf{x}_0) \tag{11}$$

4.1 Covariance of the Estimates

In addition to the estimate $\mathbf{x}_0^{\text{ML}}$, we also need to obtain the covariance associated with it. An expression for the covariance is obtained as follows. Let the $2M \times 6$ matrix $\Gamma(\mathbf{x}_0)$ be the Jacobian[2] of the vector $\boldsymbol{\nu}(\mathbf{x}_0)$

$$\Gamma(\mathbf{x}_0) = [\nabla_{\mathbf{x}_0}\boldsymbol{\nu}(\mathbf{x}_0)']' = \begin{bmatrix} -R_{s(1)}^{-\frac{1}{2}}\,[\nabla_{\mathbf{x}_0}\mathbf{h}_1(\mathbf{x}_0)']' \\ \vdots \\ -R_{s(M)}^{-\frac{1}{2}}\,[\nabla_{\mathbf{x}_0}\mathbf{h}_M(\mathbf{x}_0)']' \end{bmatrix} \tag{12}$$

An algorithmic procedure for evaluating the Jacobian $\Gamma(\cdot)$ is given in Appendix B. Note that $\Gamma(\mathbf{x}_0)$ is independent of the actual measurements $\mathbf{z}(k)$. The CRLB on the covariance, P_{CRLB}, for an unbiased estimator (evaluated at the true value, $\mathbf{x}_0^{\text{TRUE}}$, of the parameter being estimated, i.e., the target state at t_0) is

$$P_{\text{CRLB}}^{-1} = E\{[\nabla_{\mathbf{x}_0}\lambda_{Z_M}(\mathbf{x}_0)][\nabla_{\mathbf{x}_0}\lambda_{Z_M}(\mathbf{x}_0)]'\}_{\mathbf{x}_0=\mathbf{x}_0^{\text{TRUE}}} \tag{13}$$

where the expectation $E\{\cdot\}$ is over the measurement set Z_M, and $\lambda_{Z_M}(\mathbf{x}_0)$ is the log-likelihood function

$$\lambda_{Z_M}(\mathbf{x}_0) = \log \Lambda_{Z_M}(\mathbf{x}_0) = -\frac{1}{2}\boldsymbol{\nu}(\mathbf{x}_0)'\boldsymbol{\nu}(\mathbf{x}_0) + \log c \tag{14}$$

The r.h.s. of Eq. (13) can be simplified to yield

$$P_{\text{CRLB}}^{-1} = [\Gamma(\mathbf{x}_0)'\Gamma(\mathbf{x}_0)]_{\mathbf{x}_0=\mathbf{x}_0^{\text{TRUE}}} \tag{15}$$

The above result can be obtained in a straightforward manner by substituting Eq. (14) into Eq. (13) and using

$$E\{\boldsymbol{\nu}(\mathbf{x}_0)\}_{\mathbf{x}_0=\mathbf{x}_0^{\text{TRUE}}} = 0_{2M\times 1} \tag{16}$$

$$E\{\boldsymbol{\nu}(\mathbf{x}_0)\boldsymbol{\nu}(\mathbf{x}_0)'\}_{\mathbf{x}_0=\mathbf{x}_0^{\text{TRUE}}} = I_{2M\times 2M} \tag{17}$$

The actual covariance P_{ML} associated with the ML estimate $\mathbf{x}_0^{\text{ML}}$ is quite difficult to evaluate. In the linear case, it can be shown that the ML estimator is both unbiased (i.e., $E\{\mathbf{x}_0^{\text{ML}}\} = \mathbf{x}_0^{\text{TRUE}}$) and efficient (i.e., $P_{\text{ML}} = P_{\text{CRLB}}$) [5]. In the present nonlinear situation, we may approximately assume that these properties of the ML estimator still hold. Furthermore, P_{CRLB} requires the knowledge of $\mathbf{x}_0^{\text{TRUE}}$, which is available only in simulations. Otherwise, we can evaluate an approximate P_{ML} via

$$P_{\text{ML}}^{-1} \approx [\Gamma(\mathbf{x}_0)'\Gamma(\mathbf{x}_0)]_{\mathbf{x}_0=\mathbf{x}_0^{\text{ML}}} \tag{18}$$

4.2 The Minimization Algorithm

The ML estimate is the solution of a nonlinear least squares (NLS) minimization problem as indicated by the r.h.s. of Eq. (10). This problem can be solved using a variety of iterative minimization algorithms. The performance of these algorithms is affected by two major factors. The first factor is the condition number of the Hessian $\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}' f(\mathbf{x})$, of the cost function $f(\mathbf{x})$, at the minimum and at points in its neighborhood. The second factor is the choice of the initial point used to start the minimization algorithm.

For well conditioned problems[3], the performance of most algorithms is very good, in the sense that they converge rapidly to the minimum and the rate of convergence is not sensitive to the initial point used to start the algorithm. On the other hand, for ill-conditioned problems, the performance of almost all the minimization algorithms is critically dependent on the initial starting point. As stated earlier, the minimization of the cost function $f(\mathbf{x}_0)$ in Eq. (11), associated with the ML target state estimation from LOS measurements, is very ill-conditioned because of poor target motion observability. Hence, a robust minimization algorithm is required for this problem. The Levenberg-Marquardt algorithm, which is specially designed for the minimization of NLS functions, has been found to perform well for the present mimization problem and is described below.

Consider the cost function $f(\mathbf{x}_0)$ given Eq. (11) which is to be minimized. Denoting the estimated minimum at the nth iteration by $\mathbf{x}_{0,n}$, the vector of normalized measurement residuals by $\boldsymbol{\nu}(\mathbf{x}_{0,n}) = \boldsymbol{\nu}_n$ and the Jacobian by $\Gamma(\mathbf{x}_{0,n}) = \Gamma_n$, we can form an affine approximation $\boldsymbol{\omega}_n(\mathbf{x}_0)$ to $\boldsymbol{\nu}(\mathbf{x}_0)$ as

$$\boldsymbol{\omega}_n(\mathbf{x}_0) = \boldsymbol{\nu}_n + \Gamma_n\,(\mathbf{x}_0 - \mathbf{x}_{0,n}) \tag{19}$$

Using the above equation, we can form a local quadratic model $q_n(\mathbf{x}_0)$ for the cost function as

$$\begin{aligned} q_n(\mathbf{x}_0) &= \boldsymbol{\omega}_n(\mathbf{x}_0)'\boldsymbol{\omega}_n(\mathbf{x}_0) \\ &= \boldsymbol{\nu}_n'\boldsymbol{\nu}_n + 2\boldsymbol{\nu}_n'\Gamma_n(\mathbf{x}_0 - \mathbf{x}_{0,n}) \\ &\quad +(\mathbf{x}_0 - \mathbf{x}_{0,n})'\Gamma_n'\Gamma_n(\mathbf{x}_0 - \mathbf{x}_{0,n}) \end{aligned} \tag{20}$$

Under the assumptions that this quadratic model $q_n(\mathbf{x}_0)$ is a good approximation to the actual cost function $f(\mathbf{x}_0)$, the Gauss-Newton algorithm would update the estimate to the minimizer of this local quadratic approximation, i.e.,

$$\mathbf{x}_{0,n}^{\text{GN}} = \arg\min_{\mathbf{x}_0\in\Re^6} q_n(\mathbf{x}_0) \tag{21}$$

[2] Denoting the gradient of a scalar as a column vector, the gradient of a scalar z with respect to a n dimensional vector $\mathbf{x}$ is defined as the $n\times 1$ vector $\nabla_x z$ whose j^{th} component is $\frac{\partial z}{\partial x_j}$. Extending this notation to the derivative of a vector, we define the Jacobian of an m dimensional vector $\mathbf{y}$ with respect to an n dimensional vector $\mathbf{x}$ as the $m\times n$ matrix $[\nabla_x \mathbf{y}']'$ whose $(i,j)^{\text{th}}$ term is $\frac{\partial y_i}{\partial x_j}$.

[3] The condition number of a matrix is the ratio of the largest and the smallest singular values. If the condition number of the Hessian of $f(\mathbf{x})$ is less than 10^3, then both the Hessian and the associated cost function $f(\mathbf{x})$ can be said to be well-conditioned.

It can shown that this Gauss-Newton update, $\mathbf{x}_{0,n}^{\mathrm{GN}}$, is obtained as the least squares solution of a set of $2M$ linear equations, i.e.,

$$\Gamma_n \left(\mathbf{x}_{0,n}^{\mathrm{GN}} - \mathbf{x}_{0,n}\right) = -\boldsymbol{\nu}_n \tag{22}$$

The Gauss Newton iterative algorithm, though very straightforward to implement, is not globally convergent. Convergence is assured only if the initial starting point $\mathbf{x}_{0,0}$ is quite close to the global minimum. Quite a few modifications to this algorithm have been proposed to make it globally convergent. A class of algorithms that fall into this category are based on a modification to the Gauss-Newton algorithm suggested by Levenberg [6] and Marquardt [7]. This algorithm (i.e., the Levenberg-Marquardt algorithm) also updates the estimate at the nth iteration by trying to minimize the local quadratic model $q_n(\mathbf{x}_0)$ at each iteration, but this updated estimate, $\mathbf{x}_{0,n+1}$, is constrained to lie within a hyper-ellipsoid centered around $\mathbf{x}_{0,n}$, i.e.,

$$\mathbf{x}_{0,n+1} = \arg \min_{\mathbf{x}_0 \in \Omega} q_n(\mathbf{x}_0) \tag{23}$$

$$\Omega = \{\mathbf{x}_0 \ : \ \mathbf{x}_0 \in \Re^6, \ \ \|D_n \left(\mathbf{x}_0 - \mathbf{x}_{0,n}\right)\| \leq \delta_n\} \tag{24}$$

where D_n is a 6×6 diagonal scaling matrix and δ_n is the radius of the "scaled model trust region". It is easy to see that if the Gauss-Newton update is feasible, i.e., $\left\|D_n \left(\mathbf{x}_{0,n}^{\mathrm{GN}} - \mathbf{x}_{0,n}\right)\right\| \leq \delta_n$ then it is accepted as the estimate $\mathbf{x}_{0,n+1} = \mathbf{x}_{0,n}^{\mathrm{GN}}$. In general, however, $\mathbf{x}_{0,n}^{\mathrm{GN}}$ may lie outside the trust region; in such a case, the updated estimate is obtained by solving the following min-max problem:

$$\mathbf{x}_{0,n+1} = \arg \min_{\mathbf{x}_0 \in \Re^6} \max_{\lambda \in \Re} \left[q_n(\mathbf{x}_0) + \lambda \left(\|D_n \left(\mathbf{x}_0 - \mathbf{x}_{0,n}\right)\|^2 - \delta_n^2\right)\right] \tag{25}$$

where λ is the Lagrangian multiplier, which in the present context is called the Levenberg parameter. It can be shown that given a value of λ, we can solve for a $\mathbf{x}_{0,n}(\lambda)$, (i.e., solve the outer minimization problem) as the least squares solution of the following $(2M + 6)$ set of linear equations:

$$\begin{bmatrix} \Gamma_n \\ \lambda^{\frac{1}{2}} D_n \end{bmatrix} \left(\mathbf{x}_{0,n}(\lambda) - \mathbf{x}_{0,n}\right) = \begin{bmatrix} -\boldsymbol{\nu}_n \\ 0_{6\times 1} \end{bmatrix} \tag{26}$$

Note that $\mathbf{x}_{0,n}(0) = \mathbf{x}_{0,n}^{\mathrm{GN}}$, i.e., the Gauss-Newton update. This solution $\mathbf{x}_{0,n}(0)$ may not feasible, but in general, it can be shown that there exists a $\lambda_n \in [0, \infty)$ such that for all $\lambda \geq \lambda_n$, $\mathbf{x}_{0,n}(\lambda)$ is feasible. The updated estimate is then $\mathbf{x}_{0,n+1} = \mathbf{x}_{0,n}(\lambda_n)$. This value, λ_n, of the Levenberg parameter is obtained by solving $\|D_n \left(\mathbf{x}_0(\lambda) - \mathbf{x}_{0,n}\right)\| = \delta_n$. Some of the important aspects of this algorithm are summarized in the following remarks.

- Note that at each iteration, the Levenberg parameter λ_n has to obtained via an inner iterative loop. In each of these inner iterations for obtaining λ_n, we need to solve Eq. (26) for $\mathbf{x}_{0,n}(\lambda)$. In the implementation of this algorithm by Moré [8], [9], a two stage QR procedure is used to solve for the Levenberg parameter. In addition, this procedure factors the the Jacobian matrix so that the intermediate parameters required for updating δ_n can be computed very efficiently.

- Another important aspect in the evaluation of λ_n is the choice of the iterative method used in solving the nonlinear equation in λ. Newton's method for solving nonlinear equations can be used, but a much more efficient iteration has been suggested by Hebden [10] that exploits the special structure of the problem.

- The ii element of the diagonal scaling matrix D_n is updated at each iteration in the following manner:

$$[D_n]_{ii} = \max\left\{\|\mathrm{col}_i\left(\Gamma_n\right)\|, [D_n]_{ii}\right\}, n > 0 \tag{27}$$

  At the first iteration, we have $[D_0]_{ii} = \|\mathrm{col}_i\left(\Gamma_0\right)\|$.

- The radius of the model trust region δ_n is updated based on how well the local quadratic approximation predicts the change in the function value. The ratio ϑ_n of the actual to the predicted reduction in the function value is given by

$$\vartheta_n = \frac{f\left(\mathbf{x}_{0,n}\right) - f\left(\mathbf{x}_{0,n}\left(\lambda_n\right)\right)}{f\left(\mathbf{x}_{0,n}\right) - q_n\left(\mathbf{x}_{0,n}\left(\lambda_n\right)\right)} \tag{28}$$

  The denominator of the expression on the right hand side of the above equation can be easily shown to be non-negative. Depending on the value of ϑ_n, the following update strategy is adopted:

  - If ϑ_n is negative, it implies that the value of the cost function has increased instead of going down. This happens if the local quadratic model is totally inadequate. In such a case, δ_n is drastically reduced (typically by a factor of 10) and the value of the Levenberg parameter is recomputed. The estimate is updated, i.e., $\mathbf{x}_{0,n+1} = \mathbf{x}_{0,n}\left(\lambda_n\right)$ only if $\vartheta_n \geq 0.0001$ (a small positive number).

  - If $0 < \vartheta_n \leq 0.25$, the estimate is accepted, but δ_n is reduced for the next iteration, since values of ϑ_n considerably less than one indicate that the local quadratic model does not accurately reflect the cost function behavior. The new trust region radius δ_{n+1} is computed using, $\delta_{n+1} = \mu\delta_n$ where a value of $\mu \in [0.1 \ 0.5]$ is chosen using a quadratic fit in the direction of the updated estimate.

- If $\vartheta > 0.75$, then we choose $\delta_{n+1} = \delta_n$ if $\lambda_n = 0$, and $\delta_{n+1} \approx 2\delta_n$ if $\lambda_n > 0$. Since, in these cases, the local quadratic model seems to be a reasonably good approximation to the actual cost function, the Gauss-Newton update may be appropriate. To allow this update to be feasible, δ_{n+1} is increased.

A detailed theoretical development of this algorithm can be found in [8], [11]. The algorithm is started using an initial point $\mathbf{x}_{0,0}$ obtained using the initialization procedure described below. The algorithm has been found to converge to the solution, i.e., $\mathbf{x}_{0,N} \approx \mathbf{x}_0^{\mathrm{ML}}$ after a reasonably small number of iterations N (approximately 5 to 15 iterations, depending on the scenario and also on the particular realization of the LOS measurements). Furthermore, the algorithm performs optimally for this application, since the minimized value of the cost function, $f(\mathbf{x}_{0,N})$, is shown to satisfy a certain statistical test that is equivalent to the optimality of the algorithm. In addition, the ML estimates are shown to be efficient, i.e., their covariance meets the CRLB.

4.3 Algorithm Initialization

In this section, a method for obtaining an initial estimate $\mathbf{x}_{0,0} = [\boldsymbol{\xi}_{0,0}\ \dot{\boldsymbol{\xi}}_{0,0}]'$ is presented. A classical method attributed to Laplace [1] can be used if all the LOS measurements originate from the same sensor, i.e., in a single sensor scenario. The method described below is more general in the sense that it can be used both for *single* as well as for *multiple sensor* scenarios.

Let the reference time t_0 coincide with the rth measurement time instant t_r, $r \in \{1, \ldots, M\}$, and let $\boldsymbol{\rho}_r$ be the unit LOS vector, which is formed using the measured angles $\mathbf{z}(r)$.

$$\boldsymbol{\xi}_{0,0} = \boldsymbol{\eta}_{s(r)}(t_r) + \rho_r \boldsymbol{\rho}_r \tag{29}$$

where ρ_r is the (unknown) distance between the sensor and the target at t_r. In order to obtain an initial value for the velocity vector $\dot{\boldsymbol{\xi}}_{0,0}$, we use a simple constant velocity model to describe the target motion. Using the LOS vectors $\boldsymbol{\rho}_k$, obtained from $\mathbf{z}(k)$ we can write

$$\boldsymbol{\xi}_{0,0} + \dot{\boldsymbol{\xi}}_{0,0}(t_k - t_r) = \boldsymbol{\eta}_{s(k)}(t_k) + \boldsymbol{\rho}_k \rho_k \qquad k = 1, \ldots, M, \ k \neq r \tag{30}$$

Substituting $\boldsymbol{\xi}_{0,0}$ from Eq. (29) and eliminating the unknowns ρ_k by using the vector cross product $\otimes$, we obtain the following linear equations for the four unknowns ρ_r and $\dot{\boldsymbol{\xi}}_{0,0}$:

$$[\boldsymbol{\rho}_k(t_k - t_r)] \otimes \dot{\boldsymbol{\xi}}_{0,0} + [\boldsymbol{\rho}_k \otimes \boldsymbol{\rho}_r]\rho_r = \boldsymbol{\rho}_k \otimes [\boldsymbol{\eta}_{s(k)}(t_k) - \boldsymbol{\eta}_{s(r)}(t_r)] \qquad k = 1, \ldots, M, \ k \neq r \tag{31}$$

The above $2(M-1)$ linear equations can be solved (in the least squared error sense, since $M \geq 3$) for the four unknowns $\dot{\boldsymbol{\xi}}_{0,0}$ and ρ_r. Finally, $\boldsymbol{\xi}_{0,0}$ can be determined by substituting ρ_r back in Eq. (29).

5. ESTIMATION WITH PRIOR INFORMATION

In the previous section the set of measurements Z_M was the only information available. In most practical scenarios the first scan (or measurement), made at time t_1, is obtained approximately when the target comes out of the atmosphere. In this case the altitude of the target at the first scan, i.e., at time t_1 is known to be within a certain range of values. In addition, it is assumed that bounds on the speed of the target at t_1 are also available.

In this section, we consider the estimation of the target state at the reference time t_0 given the measurement set Z_M *and* additional information about some functions of the target state. The prior information is not a full prior probability density function on the state; thus, a standard MAP approach is not possible. The approach [12] will be to model the prior information as "pseudo-measurements" of some functions of the state $\mathbf{x}_0$. This keeps the problem within the ML estimation framework, but makes use of the (partial) prior information that is (reasonable to assume to be) available, i.e., it will yield a modified ML estimate with incomplete prior (MLwp). Using this approach, we shall first derive a new lower bound on the covariance matrix, taking into account the availability of prior information. Let $\mathbf{h}_0(\mathbf{x})$ be a $p \times 1$ vector of parameters about which additional information is available. This information is in the form of a probability density function $p(\mathbf{z}_0)$, where $\mathbf{z}_0$ is a pseudo-measurement. This can be modeled as the sum of the function $\mathbf{h}_0(\mathbf{x}_0^{\mathrm{TRUE}})$ and some random noise. The prior information is independent of the measurement set Z_M. The pseudo-measurement is combined with Z_M to obtain the augmented measurement set $Z_a = Z_M \cup \{\mathbf{z}_0\}$. The conditional pdf of the measurement set Z_a, which is also the new likelihood function of $\mathbf{x}_0$, is

$$\Lambda_{Z_a}(\mathbf{x}_0) = p(Z_a|\mathbf{x}_0) = p(\mathbf{z}_0|\mathbf{x}_0)p(Z_M|\mathbf{x}_0) \tag{32}$$

The mutual independence of the measurements also implies that the Fisher information due to the different measurements is additive. Let P_{CRLBwp} be the new Cramer-Rao lower bound on the covariance. The inverse of this matrix is the sum of the individual information matrices

$$P_{\mathrm{CRLBwp}}^{-1} = F_0(\mathbf{x}_0^{\mathrm{TRUE}}) + P_{\mathrm{CRLB}}^{-1} \tag{33}$$

where $F_0(\mathbf{x}_0^{\mathrm{TRUE}})$ is the Fisher information matrix associated with the measurement $\mathbf{z}_0$, given by

$$F_0(\mathbf{x}_0^{\mathrm{TRUE}}) = E\{[\nabla_{\mathbf{x}_0}\lambda_{z_0}(\mathbf{x}_0)]\,[\nabla_{\mathbf{x}_0}\lambda_{z_0}(\mathbf{x}_0)]'\}_{\mathbf{x}_0=\mathbf{x}_0^{\mathrm{TRUE}}} \tag{34}$$

The expectation $E\{\cdot\}$ in the above equation is over the pseudo-measurement $\mathbf{z}_0$ and $\lambda_{z_0}(\mathbf{x}_0)$ is the log-likelihood function associated with it

$$\lambda_{z_0}(\mathbf{x}_0) = \log p(\mathbf{z}_0|\mathbf{x}_0) \tag{35}$$

Thus,

$$F_0(\mathbf{x}_0) = \int_{-\infty}^{\infty} [\nabla_{\mathbf{x}_0}\lambda_{z_0}(\mathbf{x}_0)]\,[\nabla_{\mathbf{x}_0}\lambda_{z_0}(\mathbf{x}_0)]'\; p(\mathbf{z}_0|\mathbf{x}_0)d\mathbf{z}_0 \quad (36)$$

Substituting the expressions for $F_0(\mathbf{x}_0)$ and P_{CRLB} from Eq. (15) in Eq. (33), P_{CRLBwp} is given by

$$P^{-1}_{\text{CRLBwp}} = \{\; F_0(\mathbf{x}_0) + \Gamma(\mathbf{x}_0)'\Gamma(\mathbf{x}_0)\;\}_{\mathbf{x}_0=\mathbf{x}_0^{\text{TRUE}}} \quad (37)$$

The existence of $F_0(\mathbf{x}_0)$ depends on the integral on the r.h.s. of Eq. (36). By using a differentiable likelihood function (e.g., a uniform pdf can be approximated by a Butterworth function of sufficiently high order), $F_0(\mathbf{x}_0)$ is guaranteed to exist.

5.1 Constrained Estimation

In the above analysis, no assumptions have been made on the components of the vector $\mathbf{h}_0(\mathbf{x}_0)$ or on the associated pdf $p(\mathbf{z}_0)$. In this section, we present an estimation algorithm for the case when the target altitude and speed at the reference time t_0, are constrained to be within known bounds. The components of the parameter vector $\mathbf{h}_0(\mathbf{x})$ are the target altitude and speed, i.e.,

$$\mathbf{h}_0(\mathbf{x}_0) = \begin{bmatrix} \|\xi_0\| \\ \|\dot{\xi}_0\| \end{bmatrix} \quad (38)$$

The target altitude[4] is bounded by a known interval $r_{\min} \le \|\xi_0\| \le r_{\max}$ and the target speed is also assumed to be similarly bounded $v_{\min} \le \|\dot{\xi}_0\| \le v_{\max}$. The expression for the state estimate is obtained in a much simpler manner, without the necessity of using a pseudo-measurement[5], by forming a prior pdf directly on the state estimate $\mathbf{x}_0$ as:

$$p(\mathbf{x}_0) = \begin{cases} c_0 & \text{if} \begin{cases} \|\xi_0\| \in [r_{\min}, r_{\max}] \\ \quad\text{and} \\ \|\dot{\xi}_0\| \in [v_{\min}, v_{\max}] \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (39)$$

where c_0 is the normalization constant. The posterior pdf of $\mathbf{x}_0$ given the measurement set Z_M is

$$p(\mathbf{x}_0|Z_M) = c_1 p(\mathbf{x}_0)p(Z_M|\mathbf{x}_0) \quad (40)$$

where c_1 is another normalization constant. The desired estimate $\mathbf{x}_0^{\text{MLwp}}$ is the one that maximizes the above posterior pdf, or minimizes $\lambda_{Z_a}(\mathbf{x}_0) = -\log p(\mathbf{x}_0|Z_M)$, the negative logarithm of the posterior pdf. Using Eqs. (14) and (39) in Eq. (40), we can write

$$\lambda_{Z_a}(\mathbf{x}_0) = \begin{cases} \frac{1}{2}\boldsymbol{\nu}(\mathbf{x}_0)'\boldsymbol{\nu}(\mathbf{x}_0) - \log\,(c\,c_1) & \mathbf{x}_0 \in \Omega \\ \infty & \text{else} \end{cases} \quad (41)$$

where

$$\Omega = \left\{\mathbf{x}_0 \in \Re^6 \;:\; \|\xi_0\| \in [r_{\min}, r_{\max}],\; \|\dot{\xi}_0\| \in [v_{\min}, v_{\max}]\right\} \quad (42)$$

Hence, we can write

$$\begin{aligned} \mathbf{x}_0^{\text{MLwp}} &= \arg\min_{\mathbf{x}_0\in\Omega} f(\mathbf{x}_0) \\ &= \arg\min_{\mathbf{x}_0\in\Omega} [\;\boldsymbol{\nu}(\mathbf{x}_0)'\boldsymbol{\nu}(\mathbf{x}_0)\;] \end{aligned} \quad (43)$$

5.2 Covariance

The lower bound on the covariance matrix, P_{CRLBwp}, is obtained by computing the the Fisher information matrix $F_0(\mathbf{x}_0^{\text{TRUE}})$ and using it in Eq. (37). The integral on the r.h.s. of Eq. (36) is evaluated (see Appendix C) using a Butterworth approximation for the uniform prior pdfs of the altitude and speed (assumed independent), yielding

$$F_0(\mathbf{x}_0) \approx \begin{bmatrix} \frac{\xi_0}{\|\xi_0\|}\frac{1}{\sigma_r^2}\frac{\xi_0'}{\|\xi_0\|} & 0_{3\times3} \\ 0_{3\times3} & \frac{\dot{\xi}_0}{\|\dot{\xi}_0\|}\frac{1}{\sigma_v^2}\frac{\dot{\xi}_0'}{\|\dot{\xi}_0\|} \end{bmatrix} \quad (44)$$

where σ_r and σ_v are

$$\sigma_r = \lim_{N\to\infty} \frac{r_{\max} - r_{\min}}{2\sqrt{2-\frac{1}{N}}} = \frac{r_{\max} - r_{\min}}{2\sqrt{2}} \quad (45)$$

$$\sigma_v = \lim_{N\to\infty} \frac{v_{\max} - v_{\min}}{2\sqrt{2-\frac{1}{N}}} = \frac{v_{\max} - v_{\min}}{2\sqrt{2}} \quad (46)$$

Substituting $F_0(\mathbf{x}_0)$ back into Eq. (37), P_{CRLBwp} is obtained. In a practical scenario, where the true target state is not known, P_{MLwp}, the covariance in the MLwp estimate, is approximated via:

$$P^{-1}_{\text{MLwp}} \approx \{\; F_0(\mathbf{x}_0) + \Gamma(\mathbf{x}_0)'\Gamma(\mathbf{x}_0)\;\}_{\mathbf{x}_0=\mathbf{x}_0^{\text{MLwp}}} \quad (47)$$

5.3 Constrained Minimization Algorithm

The MLwp state estimate is obtained as the solution of a constrained minimization problem as indicated by Eqs. (43) and (42). In theory, general constrained minimization algorithms (e.g., Lagrangian multiplier method) can be used to solve this problem. A major difficulty in the practical implementation of these methods lies in the fact that, at each iteration, they require the solution of a set of nonlinear equations (i.e., the constraints). The nonlinear equations pertaining to the problem at hand (i.e., Eq. (42)) do not have a closed form solution and hence an iterative method (e.g., multidimensional Newton's method [13]) has to be used. There are

[4] The altitude is actually $\|\xi_0\|$−Earth Radius, i.e., the distance above the surface of the Earth. For the sake of convenience, we consider the distance from the center of the Earth as the altitude.

[5] The use of the pseudo-measurement is useful in obtaining the CRLB; however, in the course of the optimization the use of an equivalent constrained search is advantageous.

quite a few problems associated with these iterative methods for solving multidimensional nonlinear equations, the major one being that all of them have convergence problems and are numerically very expensive. The algorithm presented below avoids these difficulties by taking advantage of the special structure of the constraint equations.

Let $\mathbf{x}'_{0,n} = [\boldsymbol{\xi}'_{0,n}\ \dot{\boldsymbol{\xi}}'_{0,n}]$ be a feasible state estimate and $\boldsymbol{\nu}_n$ be the vector of measurement residuals at the nth iteration. We partition the Jacobian Γ_n evaluated at $\mathbf{x}_{0,n}$ as $\Gamma_n = [\Gamma_{pn}\ \Gamma_{vn}]$, where

$$\Gamma_{pn} = [\nabla_{\boldsymbol{\xi}_0} \boldsymbol{\nu}(\mathbf{x}_0)']'_{\boldsymbol{\xi}_0=\boldsymbol{\xi}_{0,n}} \tag{48}$$

$$\Gamma_{vn} = [\nabla_{\dot{\boldsymbol{\xi}}_0} \boldsymbol{\nu}(\mathbf{x}_0)']'_{\dot{\boldsymbol{\xi}}_0=\dot{\boldsymbol{\xi}}_{0,n}} \tag{49}$$

i.e., Γ_p and Γ_v (both $2M \times 3$ matrices) are the Jacobians of $\boldsymbol{\nu}(\mathbf{x}_0)$ w.r.t. $\boldsymbol{\xi}_0$ and $\dot{\boldsymbol{\xi}}_0$ respectively. The algorithmic steps involved in obtaining the updated estimate $\mathbf{x}'_{0,n+1} = [\boldsymbol{\xi}'_{0,n+1}\ \dot{\boldsymbol{\xi}}'_{0,n+1}]$ are explained below.

In the unconstrained problem, the updated estimate is obtained as the minimum of the local quadratic approximation to the cost function $f(\mathbf{x}_0)$. This local quadratic model $q_n(\mathbf{x}_0)$ is given by

$$\begin{aligned} q_n(\mathbf{x}_0) &= \boldsymbol{\nu}'_n\boldsymbol{\nu}_n + 2\boldsymbol{\nu}'_n\Gamma_n(\mathbf{x}_0 - \mathbf{x}_{0,n}) \\ &\quad + (\mathbf{x}_0 - \mathbf{x}_{0,n})'\Gamma'_n\Gamma_n(\mathbf{x}_0 - \mathbf{x}_{0,n}) \end{aligned} \tag{50}$$

It can be easily shown that the unconstrained minimizer $\mathbf{y}'_{0,n} = [\boldsymbol{\zeta}'_{0,n}\ \dot{\boldsymbol{\zeta}}'_{0,n}]$ of this local quadratic model $q_n(\mathbf{x}_0)$ is obtained by solving the linear least squares problem

$$\Gamma_n\, \mathbf{y}_{0,n} = \boldsymbol{\nu}_n \tag{51}$$

In the present case, we minimize $q_n(\mathbf{x}_0)$ subject to altitude and speed constraints. Minimizing $q_n(\mathbf{x}_0)$ with both the constraints applied simultaneously is extremely difficult, as has been indicated at the beginning of this subsection. Hence, we start by applying only the altitude constraint and using the speed constraint at a later stage, i.e., we first solve the following problem:

$$\begin{aligned} &\text{minimize} && q_n(\mathbf{x}_0) \\ &\text{such that} && r^2_{\min} \le \|\boldsymbol{\xi}_0\|^2 \le r^2_{\max} \end{aligned} \tag{52}$$

Note that if the unconstrained position estimate obtained by solving Eq. (51) is feasible (i.e., if $r_{\min} \le \|\boldsymbol{\zeta}_{0,n}\| \le r_{\max}$), we can directly accept it as the updated position estimate. In general, however, $\boldsymbol{\zeta}_{0,n}$ is not feasible, and hence a state estimate satisfying the altitude constraint is obtained by minimizing the following augmented cost function

$$g_n(\mathbf{x}_0, \alpha) = q_n(\mathbf{x}_0) + \alpha\left(\|\boldsymbol{\xi}_0\|^2 - r^2\right) \tag{53}$$

where α is the Lagrangian multiplier and

$$r = \begin{cases} r_{\min} & \text{if } \left\|\boldsymbol{\zeta}_{0,n}\right\| < r_{\min} \\ r_{\max} & \text{if } \left\|\boldsymbol{\zeta}_{0,n}\right\| > r_{\max} \end{cases} \tag{54}$$

The solution to Eq. (53) is obtained by solving both $\nabla_{\mathbf{x}_0} g_n(\mathbf{x}_0, \alpha) = 0$ and $\nabla_\alpha g_n(\mathbf{x}_0, \alpha) = 0$. For a given value of α, $\nabla_{\mathbf{x}_0} g_n(\mathbf{x}_0, \alpha) = 0$ can be solved to obtain an estimate, say $\mathbf{y}_0(\alpha)' = [\boldsymbol{\zeta}_0(\alpha)'\ \dot{\boldsymbol{\zeta}}_0(\alpha)']$. It can be shown [14] that this estimate $\mathbf{y}_0(\alpha)$ is obtained by solving the following linear equation

$$\begin{bmatrix} (\Gamma'_{pn}\Gamma_{pn} + \alpha I_3) & \Gamma'_{pn}\Gamma_{vn} \\ \Gamma'_{vn}\Gamma_{pn} & \Gamma'_{vn}\Gamma_{vn} \end{bmatrix} \begin{bmatrix} \boldsymbol{\zeta}_0(\alpha) \\ \dot{\boldsymbol{\zeta}}_0(\alpha) \end{bmatrix} = \Gamma'_n\left(\Gamma_n \mathbf{x}_{0,n} - \boldsymbol{\nu}_n\right) \tag{55}$$

Note that $\mathbf{y}_0(0) = \mathbf{y}_{0,n}$, i.e., the unconstrained estimate. Starting with a tentative value of α, we can iteratively solve the constraint equation $\|\boldsymbol{\zeta}_0(\alpha)\| = r$ (e.g., using Newton's method) to obtain the solution α_n. The updated position estimate is then given by

$$\boldsymbol{\xi}_{0,n+1} = \boldsymbol{\zeta}_0(\alpha_n) \tag{56}$$

If in addition, $\dot{\boldsymbol{\zeta}}_0(\alpha_n)$ satisfies the speed constraint, i.e., if $\left\|\dot{\boldsymbol{\zeta}}_0(\alpha_n)\right\| \in [v_{\min}, v_{\max}]$, then it can be taken as the updated velocity estimate: $\dot{\boldsymbol{\xi}}_{0,n+1} = \dot{\boldsymbol{\zeta}}_0(\alpha_n)$. In general, however, this need not be the case. In order to keep the updated velocity in the feasible region, we use a damping factor β_n as follows:

$$\dot{\boldsymbol{\xi}}_{0,n+1} = \dot{\boldsymbol{\xi}}_{0,n} + \beta_n\left(\dot{\boldsymbol{\zeta}}_0(\alpha_n) - \dot{\boldsymbol{\xi}}_{0,n}\right) \tag{57}$$

Since $\dot{\boldsymbol{\xi}}_{0,n}$ is always feasible, there exists a value of $\beta_n \in (0, 1]$ such that $\dot{\boldsymbol{\xi}}_{0,n+1}$ is also feasible. The value of β_n is obtained as the root (in $[0, 1]$) of a quadratic equation.

$$\beta_n = \begin{cases} \text{Root of } \left\|\dot{\boldsymbol{\xi}}_{0,n+1}\right\|^2 - v^2_{\min} = 0, \\ \qquad \text{if } \left\|\dot{\boldsymbol{\zeta}}_0(\alpha_n)\right\| < v_{\min} \\ 1, \text{ if } v_{\min} \le \left\|\dot{\boldsymbol{\zeta}}_0(\alpha_n)\right\| \le v_{\max} \\ \text{Root of } \left\|\dot{\boldsymbol{\xi}}_{0,n+1}\right\|^2 - v^2_{\max} = 0, \\ \qquad \text{if } \left\|\dot{\boldsymbol{\zeta}}_0(\alpha_n)\right\| > v_{\max} \end{cases} \tag{58}$$

It can be seen that in this algorithm the altitude constraint plays a dominant role in that it affects both the position and the velocity updates while the speed constraint is used only in updating the velocity components. The algorithm is designed in this manner due to the fact that the altitude constraint has a far more significant effect on the estimates than the speed constraint. In addition, it avoids the difficulty of applying both the constraints simultaneously. Since the search space is confined to the feasible region, the initial starting point of the algorithm $\mathbf{x}_{0,0}$ must also be feasible. Finally, since this algorithm is based on the damped Gauss-Newton method [11], it has similar convergence characteristics in that convergence is assured if the starting point is sufficiently close to the unconstrained minimum.

5.4 Algorithm Initialization

The initialization procedure presented in Section 3.3 can be improved using the prior altitude and speed information. Following the same notation as in Section 3.3, we first note that $\|\xi_{0,0}\| \in [r_{min}, r_{\max}]$. Hence, a tentative value of $\|\xi_{0,0}\| \approx 0.5(r_{min} + r_{\max})$ can be used to determine ρ_r as one of the positive roots of the following quadratic equation

$$\rho_r^2 + 2[\rho_r' \cdot \eta_{s(r)}(t_r)]\,\rho_r + \|\eta_{s(r)}(t_r)\|^2 = \|\xi_{0,0}\|^2 \quad (59)$$

Substituting ρ_r in Eq. (29) directly yields $\xi_{0,0}$. In addition Eq. (31) for $\dot{\xi}_{0,0}$ can be modified (since $\xi_{0,0}$ is known) as follows

$$[\rho_k(t_k - t_r)] \otimes \dot{\xi}_{0,0} = \rho_k \otimes [\eta_{s(k)}(t_k) - \xi_{0,0}] \qquad k = 1, \ldots, M, \quad k \neq r \quad (60)$$

The above set of $2(M-1)$ linear equations is solved (i.e., in the least squared error sense) for $\dot{\xi}_{0,0}$. Since we require that the speed constraint be satisfied, the components of the velocity vector $\dot{\xi}_{0,0}$ obtained are scaled such that $\|\dot{\xi}_{0,0}\| \approx 0.5(v_{min} + v_{\max})$.

6. RESULTS

In this section we consider an example target-sensor(s) scenario and investigate the performance of the ML and the MLwp state estimators. The main goal of this numerical example is to illustrate the usefulness of the theoretical bounds and also establish the fact that the estimation algorithms are indeed optimal. Before we present the results the ground-truth scenario is described first.

At the reference time $t = 0$ s, The ballistic object (target) is located at 30°N latitude 45°E longitude and at an altitude of 50 km, which yields $\xi_0^{\text{TRUE}} = [3936.42 \;\; 3936.42 \;\; 3214.07]'$ km as the true target position in the GCE coordinate frame. The target has a flight range of 1000 km and is heading North-East at the reference time. This range and heading specifications yield $\dot{\xi}_0^{\text{TRUE}} = [-0.4028 \; 0.1792 \; 0.2360]'$ km/s as the true target velocity in the GCE coordinate frame. The LOS measurements are obtained from two satellites (S_1 and S_2), both of which are in circular orbits in the equatorial plane rotating in the westward direction, at an altitude of 1000 km above mean sea level. At time $t = 0$, the satellite S_1 is located at 0° latitude and 0° longitude and satellite S_2 is located at 0° latitude and 90°E longitude. The generation of the target trajectory and measurements from the moving satellites used the software in [15].

The measurement covariance $R_{s(k)}$ for the sensor can be written in terms of σ_{LOS} as follows:

$$R_{s(k)} = \begin{bmatrix} \frac{\sigma_{\text{LOS}}^2}{\cos^2\theta(k)} & 0 \\ 0 & \sigma_{\text{LOS}}^2 \end{bmatrix} \quad (61)$$

where $\theta(k)$ is the target elevation as seen by the sensor. We present results for $\sigma_{\text{LOS}} = 5$, 10 and 25μ rad. For simplicity, we assume that both sensors operate at the same σ_{LOS}.

The extent of target observability depends not only on the sensor-target geometry, but also on the total time of observation. Hence, in our simulation we fixed the total time window of observation (T_{win}) at 60 seconds and varied the sampling time interval: T_s. For example, with $T_s = 1$ s we have $M = 61$ LOS measurements while with $T_s = 20$ s we get $M = 4$ LOS measurements.

We consider the following two measurement scenarios: 1) a single sensor scenario where all the LOS measurements originate from sensor S_1 only; 2) a two sensor scenario where the LOS measurements come from the two sensors S_1 and S_2 in an alternating fashion, with the sensor S_1 making the first measurement. All the simulated results are based on 100 Monte Carlo runs.

6.1 Residual Measurement Error

Given a target state estimate $\mathbf{x}_0$, the value of the normalized residual measurement error $f(\mathbf{x}_0)$ is a measure of the "goodness of fit" between the actual measurements and the predicted measurements. It can be easily shown that the residual measurement error evaluated at the true target state, $f_{\text{TRUE}} = f(\mathbf{x}_0^{\text{TRUE}})$ is chi-squared distributed with mn_z degrees of freedom ($n_z = 2$ and m = number of LOS measurements, in the present problem).

$$f_{\text{TRUE}} \sim \chi^2_{mn_z} \quad (62)$$

The ML estimate, $\mathbf{x}_0^{\text{ML}}$, is the unconstrained minimizer of the cost function. In a completely linear setting, it can be shown (e.g., [5]) that this minimum value, $f_{\text{ML}} = f(\mathbf{x}_0^{\text{ML}})$ is chi-squared distributed with $(mn_z - n_x)$ degrees of freedom ($n_x = 6$ in this case). This latter property is very useful both in simulations and in real time applications for evaluating the optimality of estimators. In the present nonlinear context, we can expect this property

$$f_{\text{ML}} \sim \chi^2_{mn_z - n_x} \quad (63)$$

to hold approximately.

The constrained estimate, $\mathbf{x}_0^{\text{MLwp}}$, which is obtained using the prior information, may not always be the global minimum. Hence, the associated residual measurement error $f_{\text{MLwp}} = f(\mathbf{x}_0^{\text{MLwp}})$ is in general slightly greater than f_{ML}. Using Eq. (63) we can obtain the 95% two-sided probability region for the average value of f_{ML} over N Monte Carlo runs as follows[6]

$$\frac{\chi^2_{(2M-6)N}(0.975)}{N} \leq \bar{f}_{\text{ML}} = \frac{1}{N}\sum_{i=1}^{N} f_{\text{ML}}^{(i)} \leq \frac{\chi^2_{(2M-6)N}(0.025)}{N} \quad (64)$$

[6] The notation used indicates that the probability mass to the left of the point $\chi^2_n(\alpha)$ is equal to α.

M	T_s in sec	$\frac{\chi^2_{(2M-6)N}(0.975)}{N}$	$\overline{f}_{ML}$ ($\overline{f}_{MLwp}$)			$\frac{\chi^2_{(2M-6)N}(0.025)}{N}$
			$\sigma_{LOS} = 5\mu rad$	$\sigma_{LOS} = 10\mu rad$	$\sigma_{LOS} = 25\mu rad$	
61	1.0	113.03	116.18 (116.39)	116.26 (116.71)	113.32 (113.88)	119.00
31	2.0	53.94	55.62 (55.90)	54.74 (55.52)	56.59 (57.31)	58.09
21	3.0	34.35	33.98 (34.40)	36.32 (36.97)	35.78 (36.78)	37.68
16	4.0	24.60	26.55 (27.01)	26.34 (26.90)	26.15 (27.04)	27.43
13	5.0	18.77	20.73 (21.08)	19.80 (20.35)	20.02 (20.99)	21.25
11	6.0	14.91	16.32 (16.94)	16.09 (16.62)	15.44 (16.19)	17.12
7	10.0	7.23	8.62 (9.14)	8.46 (8.99)	7.62 (8.80)	8.80
6	12.0	5.34	6.24 (7.06)	6.28 (7.18)	6.46 (7.34)	6.69
5	15.0	3.46	3.81 (4.41)	3.69 (4.29)	3.88 (5.04)	4.57
4	20.0	1.62	1.75 (2.21)	1.76 (2.45)	1.98 (3.07)	2.41

$\overline{f}_{ML}$ and $\overline{f}_{MLwp}$ averaged over 100 Monte Carlo runs.
Time interval between the first and the last measurements is 60 s.

Table 1: Residual measurement error for the single sensor scenario with 60 s observation window.

M	T_s in sec	$\frac{\chi^2_{(2M-6)N}(0.975)}{N}$	$\overline{f}_{ML}$ ($\overline{f}_{MLwp}$)			$\frac{\chi^2_{(2M-6)N}(0.025)}{N}$
			$\sigma_{LOS} = 5\mu rad$	$\sigma_{LOS} = 10\mu rad$	$\sigma_{LOS} = 25\mu rad$	
61	1.0	113.03	116.48 (116.48)	115.96 (115.96)	113.56 (113.56)	119.00
31	2.0	53.94	55.56 (55.56)	55.39 (55.39)	55.64 (55.64)	58.09
21	3.0	34.35	33.90 (33.90)	36.02 (36.02)	35.79 (35.79)	37.68
16	4.0	24.60	26.78 (26.78)	26.11 (26.11)	26.28 (26.28)	27.43
13	5.0	18.77	19.56 (19.56)	20.32 (20.32)	19.48 (19.48)	21.25
11	6.0	14.91	16.77 (16.77)	16.25 (16.25)	15.16 (15.16)	17.12
7	10.0	7.23	8.50 (8.50)	7.82 (7.82)	8.57 (8.57)	8.80
6	12.0	5.34	5.81 (5.81)	5.97 (5.97)	6.40 (6.40)	6.69
5	15.0	3.46	3.90 (3.90)	3.79 (3.79)	3.61 (3.61)	4.57
4	20.0	1.62	1.71 (1.71)	1.79 (1.79)	1.78 (1.78)	2.41

$\overline{f}_{ML}$ and $\overline{f}_{MLwp}$ averaged over 100 Monte Carlo runs.
Time interval between the first and the last measurements is 60 s.
The first LOS measurement is from sensor S_1, the next one is from S_2, and so on in an alternating fashion.

Table 2: Residual measurement error for the two sensor scenario with 60 s observation window.

where $f_{ML}^{(i)}$ is the residual error in the ith Monte Carlo run.

The values of $\overline{f}_{ML}$ (and $\overline{f}_{MLwp}$) averaged over 100 monte Carlo runs, and the corresponding 95% confidence bounds, are listed in Tables 1 and 2. Table 1 is for the single sensor scenario and Table 2 is for the two sensor scenario. It can be seen that $\overline{f}_{ML}$ lies within its 95% bounds in both the scenarios.

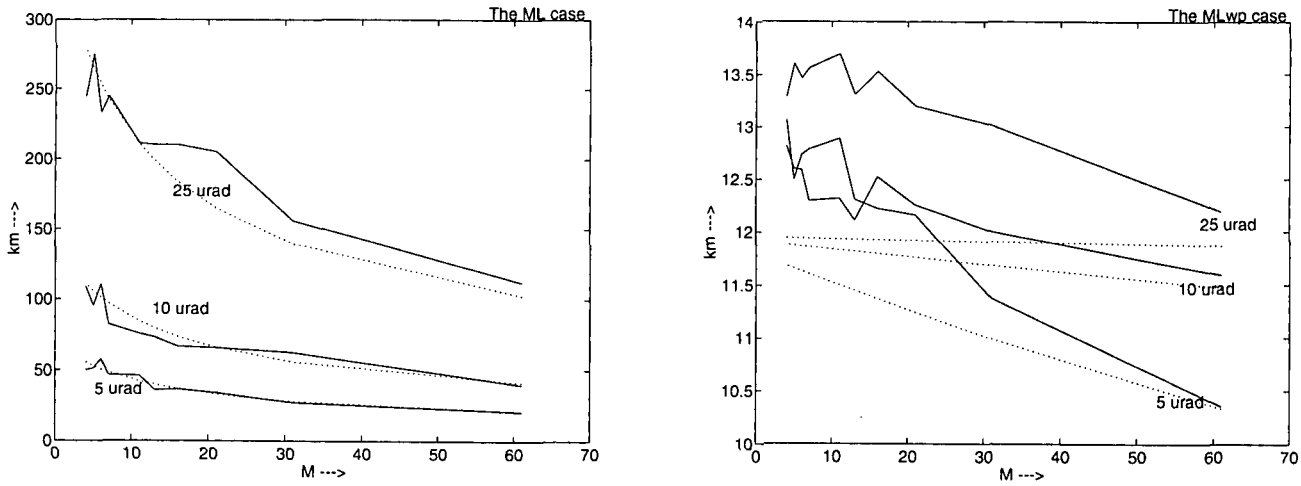
In the single satellite scenario (see Table 1), $\overline{f}_{MLwp}$ is slightly higher than $\overline{f}_{ML}$ for $M < 10$. This is due to the fact that we are trading off residual measurement error for less estimation errors. Interestingly, in the two sensor scenario (see Table 2) $\overline{f}_{ML}$ and $\overline{f}_{MLwp}$ are almost identical. The use of two satellites yields a large improvement in the target observability ensuring very low estimation errors. Hence, in this case, the additional target information is redundant.

6.2 RMS Position and Velocity Errors

The position and velocity RMS errors for the single sensor scenario are presented in Figures 1 and 2. The RMS errors of the ML and the MLwp are plotted separately due to the large difference in their magnitudes. The Cramer-Rao bounds on these errors are also indicated for both the ML and the MLwp cases. The simulated values of position and velocity errors are obtained from the diagonal elements of the sample Mean-Square Error matrix over 100 Monte Carlo runs.

Note the large difference between the position errors in the ML and the MLwp estimates in Figure 1. This error reduction is achieved by utilizing the available partial prior information which is very valuable in poor target observability situations. As the target observability improves (either due to increased number measurements, or reduced sensor noise levels) the prior information becomes less important and there is not much difference in the errors in the ML and the MLwp estimates. Also, note a similar difference (though to a much lesser extent) between the velocity errors in the ML and the MLwp estimates shown in Figure 2. In addition, note that the actual errors are commensurate with the theoretical values, an indication of the efficiency of the estimates.

The RMS position and velocity errors for the two satellite scenario are shown in Figures 3 and 4. First, note the dramatic reduction in the magnitudes of the errors (as compared to the single satellite case),



····· Cramer-Rao Bound; —— Simulated Result.

Figure 1: Normalized RMS position error for the single sensor scenario with 60 s observation window.

which is a direct result of the enhancement in target observability. Secondly, the prior information is of negligible value, since the errors in the ML and the MLwp estimates are identical.

6.3 Efficiency and Consistency

The RMS errors indicate the efficiency of the estimator by considering only the diagonal terms of the covariance matrix. A more rigorous test of the efficiency of the estimator is obtained by taking into consideration the entire covariance matrix. The *normalized estimation error squared* (NEES) [5] is defined as

$$\alpha_{\text{ML}} = [\mathbf{x}_0^{\text{ML}} - \mathbf{x}_0^{\text{TRUE}}]' P_{\text{CRLB}}^{-1} [\mathbf{x}_0^{\text{ML}} - \mathbf{x}_0^{\text{TRUE}}] \qquad (65)$$

If the estimator is unbiased and if its covariance is equal to P_{CRLB}, (i.e., it is *efficient*), then α_{ML} defined above is chi-squared distributed with n_x (i.e., 6 in the present problem) degrees of freedom. Taking the average value over N Monte Carlo runs, the 95% probability bound on $\overline{\alpha}_{\text{ML}}$ is

$$\frac{\chi^2_{6N}(0.975)}{N} \leq \overline{\alpha}_{\text{ML}} = \frac{1}{N}\sum_{i=1}^{N} \alpha_{\text{ML}}^{(i)} \leq \frac{\chi^2_{6N}(0.025)}{N} \qquad (66)$$

where $\alpha_{\text{ML}}^{(i)}$ is the NEES calculated in the ith Monte Carlo run. Both estimator bias and inefficiency would tend to make $\overline{\alpha}_{\text{ML}}$ lie outside these statistical bounds.

The estimator-calculated covariance P_{ML}, has to be consistent with its actual covariance. This consistency of the estimator is very important because in practical applications only the estimator calculated covariance is available, and any inconsistency (optimistic or pessimistic) would make the estimator perform suboptimally. The same NEES as defined above can be used, with the estimator calculated covariance P_{ML} in place of P_{CRLB}, as a measure of the estimator consistency

$$\beta_{\text{ML}} = [\mathbf{x}_0^{\text{ML}} - \mathbf{x}_0^{\text{TRUE}}]' P_{\text{ML}}^{-1} [\mathbf{x}_0^{\text{ML}} - \mathbf{x}_0^{\text{TRUE}}] \qquad (67)$$

Taking the average value of β_{ML} over N Monte Carlo runs, we get

$$\frac{\chi^2_{6N}(0.975)}{N} \leq \overline{\beta}_{\text{ML}} = \frac{1}{N}\sum_{i=1}^{N} \beta_{\text{ML}}^{(i)} \leq \frac{\chi^2_{6N}(0.025)}{N} \qquad (68)$$

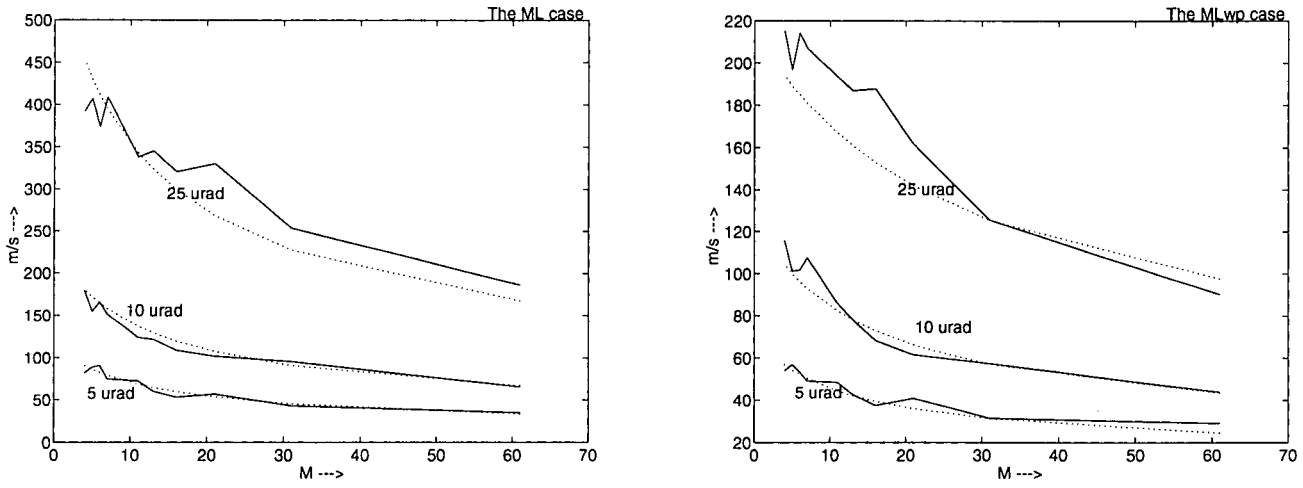
where $\beta_{\text{ML}}^{(i)}$ is the NEES calculated in the ith Monte Carlo run.

The measures of efficiency $\overline{\alpha}_{\text{MLwp}}$ and consistency $\overline{\beta}_{\text{MLwp}}$ for the MLwp estimator are defined as above with $\mathbf{x}_0^{\text{MLwp}}$, P_{CRLBwp} and P_{MLwp} used in place of $\mathbf{x}_0^{\text{ML}}$, P_{CRLB} and P_{ML} respectively.

The NEES measures of efficiency and consistency for the ML and MLwp estimators are shown in Figures 5 and 6 for the single sensor scenario in Figures 7 and 8 for the two sensor scenario. With a few exceptions (i.e., for $M < 10$), it can be seen that for both scenarios the estimators are both efficient and consistent. This result establishes the claim that the estimator achieves the CRLB.

6.4 Altitude and Speed Errors

In this subsection we present the 2×2 matrices H_{CRLB} and H_{CRLBwp}, which are the approximate lower bounds on the covariance of the estimated altitude, speed for the ML and MLwp estimators respectively. Using this covariance matrix we plot the 2σ error ellipse on the altitude-speed axes along with a scatter plot of the altitude-speed estimates from the individual Monte Carlo runs.



····· Cramer-Rao Bound; —— Simulated Result.

Figure 2: Normalized RMS velocity error for the single sensor scenario with 60 s observation window.

We define the 2×1 vector $\mathbf{h}_0(\mathbf{x}_0)$ whose components are the target altitude and speed as

$$\mathbf{h}_0(\mathbf{x}_0) = \begin{bmatrix} \|\xi_0\| \\ \|\dot{\xi}_0\| \end{bmatrix} \tag{69}$$

The desired covariance matrices H_{CRLB} and H_{CRLBwp} can be approximated using the Jacobian of $\mathbf{h}_0(\mathbf{x}_0)$ with respect to $\mathbf{x}_0$ evaluated at the true target state $\mathbf{x}_0^{\text{TRUE}} =$ as follows

$$H_{\text{CRLB}} \approx \left[\nabla_{\mathbf{x}_0}\mathbf{h}_0\left(\mathbf{x}_0^{\text{TRUE}}\right)'\right]' P_{\text{CRLB}} \left[\nabla_{\mathbf{x}_0}\mathbf{h}_0\left(\mathbf{x}_0^{\text{TRUE}}\right)'\right] \tag{70}$$

$$H_{\text{CRLBwp}} \approx \left[\nabla_{\mathbf{x}_0}\mathbf{h}_0\left(\mathbf{x}_0^{\text{TRUE}}\right)'\right]' P_{\text{CRLBwp}} \left[\nabla_{\mathbf{x}_0}\mathbf{h}_0\left(\mathbf{x}_0^{\text{TRUE}}\right)'\right] \tag{71}$$

The expression for $\nabla_{\mathbf{x}_0}\mathbf{h}_0(\mathbf{x}_0)'$ can be easily shown to be given by the following equation

$$\left[\nabla_{\mathbf{x}_0}\mathbf{h}_0(\mathbf{x}_0)'\right] = \begin{bmatrix} \frac{\xi_0}{\|\xi_0\|} & 0_{3\times1} \\ 0_{3\times1} & \frac{\dot{\xi}_0}{\|\dot{\xi}_0\|} \end{bmatrix} \tag{72}$$

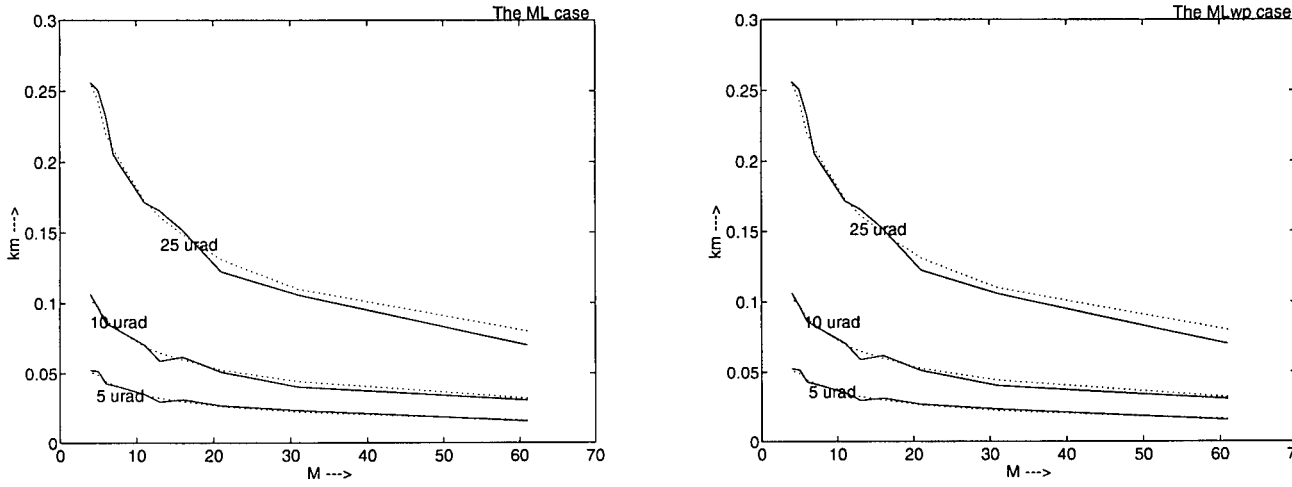
In Figures 9 and 10 the scatter plot of altitude and speed estimates for the single and two sensor scenarios is shown. In this case, the sampling time is fixed at 3 sec (i.e., 21 scans) and $\sigma_{\text{LOS}} = 10\mu$rad. Note that in the MLwp case for the single sensor scenario shown in Figure 9 all the altitude estimates are constrained within ± 25 km from the true altitude (50 km) and the speed bound of ± 0.5 km/s is not tight enough to make a similar impact on the speed estimates. Once again note the rather large difference between the ML and MLwp altitude errors, for the single sensor scenario (shown in Figure 9).

In the two sensor scenario (Figure 10) the dramatic reduction in the magnitudes of the errors (as compared to the single satellite case) is once again a direct result of the enhanced observability. As has been seen before in the case of position errors, the prior information is of negligible value, since the errors in the ML and the MLwp estimates are identical.

6.5 Launch and Impact Point Errors

The exoatmospheric trajectory of the target is completely determined by the target state $\mathbf{x}_0$ at the reference time t_0, because the target dynamics are completely determined by Keppler's laws. But, during the boost and re-entry phases, the effect of the Earth's atmosphere and the thrust (in the boost phase) play a major role in determining the motion of the target. An accurate determination of the launch and impact points of the target would require a completely different motion model which is beyond the scope of this chapter, hence we shall estimate the launch and impact point errors under the simplifying assumption that the entire target trajectory (from launch to impact) is governed only by Keppler's laws. This reflects the contribution of the estimation errors on the accuracy of the trajectory extrapolation. Additional errors not accounted for here, would be from imperfect thrust/drag modeling.

In Figures 11 and 12 the launch point estimates for the single and two sensor scenarios are shown, for $T_s = 3$ sec (i.e., 21 scans) and $\sigma_{\text{LOS}} = 10\mu$rad. The 2σ error ellipses are plotted using the sample covariance matrices. In the single sensor scenario (Figure 11), the launch point errors are considerably less in the ML wp case as compared to the ML case. Also, the clustering of the sample points in the MLwp case is due to the constraints on the target altitude. In sharp contrast, the two sensor scenario (Figure 12) yields very accurate launch point estimates. As indicated above, these errors reflect only the contribution of the estimation errors at t_0; The use of a thrust profile and motion equations with drag with appropriate errors would be necessary to complete



····· Cramer-Rao Bound; —— Simulated Result.

Figure 3: Normalized RMS position error for the two sensor scenario with 60 s observation window

the error picture.

The impact point errors for the single and two sensor scenarios are shown in Figures 13 and 14 respectively. Note that the the impact point is at a far longer distance away (compared to the launch point) from the estimated state at the reference time. Hence, the impact point errors are higher than the launch point errors. In the single sensor scenario (Figure 13), the MLwp estimator still yields considerably less error than the ML estimator. It can be seen from Figure 14 that in the two sensor scenario the impact point can be determined fairly accurately, even though the state estimate was at a far earlier point in time.

Finally, we would like to reiterate that both the launch and impact point errors would in general be higher if the thrust of the missile during the boost phase and the effects of the Earth's atmosphere (e.g., atmospheric drag during the re-entry phase) and the associated errors are not neglected. Nonetheless, this example illustrates the possible application of the methods presented in this chapter in determining the origin and the destination of a ballistic missile.

7. CONCLUSIONS

We have presented a robust algorithm to estimate the trajectory of a ballistic object from angle-only measurements and also incorporating additional trajectory information. Furthermore, the Cramer-Rao lower bounds on the covariance matrix for both the ML and MLwp cases have been derived. The results indicate that even with poor observability, the algorithm provides a reliable estimate of the target trajectory. The incorporation of additional target motion information yields significant improvement in the estimation errors in single satellite scenarios, particularly when the number of LOS measurements is small. The use of data from more than one satellite improves target observability and yields estimates of much higher accuracy.

There are still some unresolved issues that are related to the problem of ballistic trajectory estimation. There do not exist any analytical results relating the observability of the target to the position of the sensor(s). Such a result is very useful in strategically locating sensor satellites so as to maximize the observability of a target launched from some particular region on the Earth's surface. Incorporation of probabilistic prior information on attributes of target trajectory, such as the target flight angle and heading angle, into the framework of the modified ML approach is currently under investigation.

8. ACKNOWLEDGEMENTS

The first author would like to thank Vijaya Raghavan for all the insightful discussions that helped in considerably enhancing the quality of this work. Also thanks to B. Fridling for the SAAM software of [15].

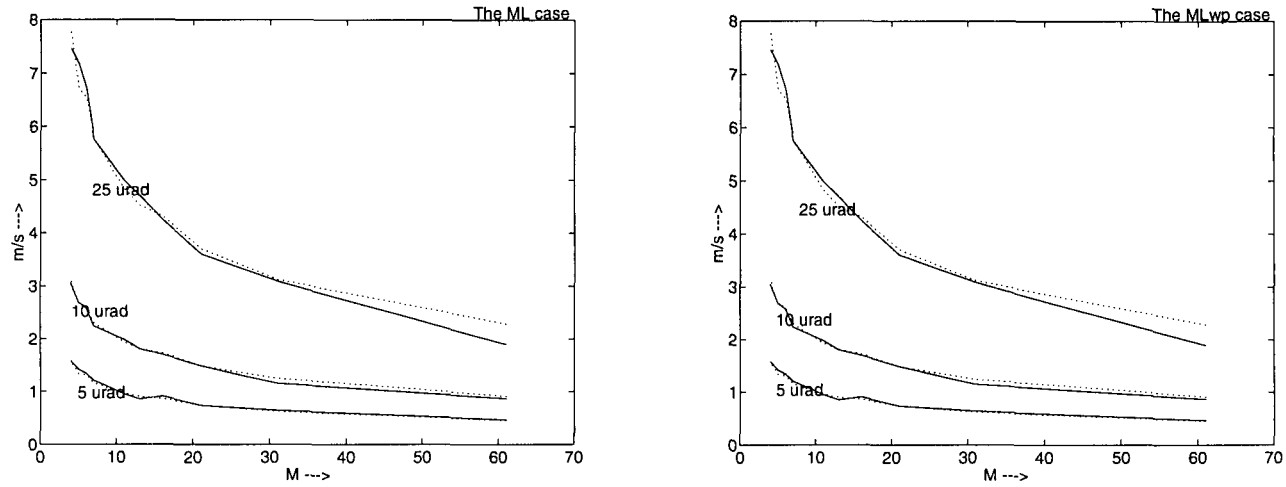
9. REFERENCES

[1] Bate, R., et al. *Fundamentals of Astrodynamics.* New York, NY: Dover, 1971.

[2] Becker K. "Simple linear theory approach to TMA observability", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 29, No. 2, pp. 575-578, April 1993.

[3] Chang, C. B. "Ballistic Trajectory Estimation with Angle-Only Measurements", *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 3, pp. 474-480, June 1980.

[4] Danis N. J. "Space-Based Tactical Ballistic Missile Launch Parameter Estimation", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 29, No. 2, pp. 412-424, April 1993.



····· Cramer-Rao Bound; —— Simulated Result.

Figure 4: Normalized RMS velocity error for the two sensor scenario with 60 s observation window

[5] Bar-Shalom, Y. and Li, X. R. *Estimation and Tracking: Principles, Techniques and Software.* Norwood, MA: Artech House 1993.

[6] Levenberg, K. "A method for the solution of certain nonlinear problems in least squares", *Quart. Appl. Math.* **2**, pp. 164-168, 1944.

[7] Marquardt, D. "An algorithm for least-squares estimation of nonlinear parameters", *SIAM J. Appl. Math.* **11**, pp. 431-441, 1963.

[8] Moré, J. J. "The Levenberg-Marquardt algorithm: implementation and theory", in *Numerical Analysis* G. A Watson, ed., Lecture Notes in Math. 630. Springer Verlag, Berlin, 1977.

[9] Moré, J. J., Garbow, B. S. and Hillstrom, K. E. "Fortran subroutines for testing unconstrained optimization software", *IEEE Transactions on Mathematical Software*, Vol. 7, Jan. 1981.

[10] Hebden, M. D. "An algorithm for minimization using exact second derivatives", Rept. TP515, Atomic Energy Research Establishment, Harwell, UK, 1973.

[11] Dennis, J. E. and Schnabel, R. B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Englewood Cliffs: Prentice-Hall, 1983.

[12] Taylor, J. H. "The Cramer-Rao Estimation error lower bound computation for deterministic nonlinear systems", *IEEE Transactions on Automatic Control*, Vol. AC-24, No. 2, pp. 343-344, April 1979.

[13] Press, W. H., et al. *Numerical Recipes in C.* Cambridge, UK: Cambridge University Press, 1990.

[14] Golub, G. H. and Van Loan, C. F. *Matrix Computations.* Baltimore, MD: Johns Hopkins Press, 1989.

[15] Garretson, W. D, and Paterson, T. S., "Sensor architecture analysis model (SAAM)", IDA Document D-748, Institute for Defense Analysis, Alexandria, VA 22311-1772, April 1990.

[16] Gradshteyn, I. S. and Ryzhik, I. M. *Table of Integrals, Series, and Products.* New York, NY: Academic Press, 1980.

A. Algorithm for State Propagation

Presented below is an algorithm that propagates the state of an object in a ballistic trajectory around the earth. Let $\mathbf{x}(t)' = [\boldsymbol{\xi}(t)' \ \dot{\boldsymbol{\xi}}(t)'] = \mathbf{f}(\mathbf{x}_0, t_0, t)$ be the unknown state to be computed given the time t, a reference state $\mathbf{x}_0' = \mathbf{x}(t_0)' = [\boldsymbol{\xi}_0' \ \dot{\boldsymbol{\xi}}_0']$ and the reference time t_0. The underlying theoretical concepts and the derivation of the equations used in this algorithm can be found in [1]. The gravitational parameter $\mu = 3.986012 \times 10^5 \, \mathrm{km}^3/\mathrm{sec}^2$ and the convergence check parameter $TOL = 10^{-10}$ are used.

STEP 1

$r_0 := \|\boldsymbol{\xi}_0\|;$

$v_0 := \left\|\dot{\boldsymbol{\xi}}_0\right\|;$

$q_0 := \frac{1}{\mu} \boldsymbol{\xi}_0' \dot{\boldsymbol{\xi}}_0;$

$a_0 := \frac{2}{r_0} - \frac{v_0^2}{\mu};$

$p_0 := \frac{1 - a_0 r_0}{\sqrt{\mu}};$

STEP 2

if $(a_0 \geq 0)$

$\alpha := \frac{a_0(t - t_0)}{\sqrt{\mu}};$

else

$$\alpha := \frac{\text{sign}(t-t_0)}{\sqrt{-a_0}} \log\left(\frac{2(\sqrt{-a_0})^3(t-t_0)}{q_0\sqrt{-a_0}+p_0\text{sign}(t-t_0)}\right);$$

STEP 3

$\beta := a_0\alpha^2;$

if $(a \geq 0)$

$c := \frac{1-\cos(\sqrt{\beta})}{\beta};$

$s := \frac{\sqrt{\beta}-\sin(\sqrt{\beta})}{\beta\sqrt{\beta}};$

else

$c := \frac{1-\cosh(\sqrt{-\beta})}{\beta};$

$s := \frac{\sinh(\sqrt{-\beta})-\sqrt{-\beta}}{\beta\sqrt{-\beta}};$

STEP 4

$\tau := p_0\alpha^3 s + q_0\alpha^2 c + \frac{r_0}{\sqrt{\mu}}\alpha;$

$\frac{d\tau}{d\alpha} := p_0\alpha^2 c + q_0\alpha(1-s\beta) + \frac{r_0}{\sqrt{\mu}};$

$\alpha := \alpha + \left[\frac{d\tau}{d\alpha}\right]^{-1} [(t-t_0)-\tau];$

STEP 5

if $([(t-t_0)-\tau] > TOL)$

goto STEP 3

STEP 6

$f := 1 - \frac{\alpha^2 c}{r_0};$

$g := (t-t_0) - \frac{\alpha^3 s}{\sqrt{\mu}};$

$\boldsymbol{\xi}(t) := f\,\boldsymbol{\xi}_0 + g\,\dot{\boldsymbol{\xi}}_0;$

$r := \|\boldsymbol{\xi}(t)\|;$

STEP 7

$\dot{f} := \left(\frac{\sqrt{mu}}{r_0}\right)(s\beta-1)\left(\frac{\alpha}{r}\right);$

$\dot{g} := 1 - \frac{\alpha^2 c}{r};$

$\dot{\boldsymbol{\xi}}(t) := \dot{f}\,\boldsymbol{\xi}_0 + \dot{g}\,\dot{\boldsymbol{\xi}}_0;$

The above steps yield the required state $\mathbf{x}(t)' = [\boldsymbol{\xi}(t)'\ \dot{\boldsymbol{\xi}}(t)']$ at time t. For computing the Jacobian $\Gamma(\mathbf{x}_0)$ we require the 6×3 matrix $\nabla_{\mathbf{x}_0}\boldsymbol{\xi}(t)$. The computation of this matrix involves the following additional steps.

STEP 8

$$\nabla_{\mathbf{x}_0} r_0 := \begin{bmatrix} \boldsymbol{\xi}_0 \\ 0_{3\times 1} \end{bmatrix}\left(\frac{1}{r_0}\right);$$

$$\nabla_{\mathbf{x}_0} v_0 := \begin{bmatrix} 0_{3\times 1} \\ \dot{\boldsymbol{\xi}}_0 \end{bmatrix}\left(\frac{1}{v_0}\right);$$

$$\nabla_{\mathbf{x}_0} q_0 := (\nabla_{\mathbf{x}_0} r_0)\left(\frac{r_0}{\mu}\right) + (\nabla_{\mathbf{x}_0} v_0)\left(\frac{v_0}{\mu}\right);$$

$$\nabla_{\mathbf{x}_0} a_0 := (\nabla_{\mathbf{x}_0} r_0)\left(\frac{-2}{r_0^2}\right) + (\nabla_{\mathbf{x}_0} v_0)\left(\frac{-2v_0}{\mu}\right);$$

$$\nabla_{\mathbf{x}_0} p_0 := (\nabla_{\mathbf{x}_0} r_0)\left(\frac{-a_0}{\sqrt{\mu}}\right) + (\nabla_{\mathbf{x}_0} v_0)\left(\frac{-r_0}{\sqrt{\mu}}\right);$$

STEP 9

$\frac{dc}{d\beta} := \frac{c-3s}{2\beta};$

$\frac{ds}{d\beta} := \frac{1-s\beta-2c}{2\beta};$

STEP 10

$A :=$

$$\begin{bmatrix} 3p_0\alpha^2 s + 2q_0\alpha c + \frac{r_0}{\sqrt{\mu}} & p_0\alpha^3\frac{ds}{d\beta} + q_0\alpha^2\frac{dc}{d\beta} \\ 2a_0\alpha & -1 \end{bmatrix};$$

$b_1 := (\nabla_{\mathbf{x}_0} q_0)(-\alpha^2 c) + (\nabla_{\mathbf{x}_0} p_0)(-\alpha^3 s)$
$\quad + (\nabla_{\mathbf{x}_0} r_0)\left(\frac{-\alpha}{\sqrt{\mu}}\right);$

$b_2 := (\nabla_{\mathbf{x}_0} a_0)(-\alpha^2);$

$[(\nabla_{\mathbf{x}_0}\alpha)\ (\nabla_{\mathbf{x}_0}\beta)] := [b_1\ b_2]\,A^{-1};$

STEP 11

$$\nabla_{\mathbf{x}_0} f := \left[(\nabla_{\mathbf{x}_0} r_0)\left(\frac{\alpha c}{r_0}\right) - (\nabla_{\mathbf{x}_0}\alpha)(2c) - (\nabla_{\mathbf{x}_0}\beta)\left(\alpha\frac{dc}{d\beta}\right)\right]\left[\frac{\alpha}{r_0}\right];$$

$$\nabla_{\mathbf{x}_0} g := \left[(\nabla_{\mathbf{x}_0}\alpha)(3s) - (\nabla_{\mathbf{x}_0}\beta)\left(\alpha\frac{ds}{d\beta}\right)\right]\left[\frac{-\alpha^2}{\sqrt{\mu}}\right];$$

STEP 12

$$\nabla_{\mathbf{x}_0}\boldsymbol{\xi}(t) = \begin{bmatrix} fI_3 \\ gI_3 \end{bmatrix} + (\nabla_{\mathbf{x}_0} f)\,\boldsymbol{\xi}_0' + (\nabla_{\mathbf{x}_0} g)\,\dot{\boldsymbol{\xi}}_0';$$

B. Computation of the Jacobian

The procedure to evaluate the Jacobian $\Gamma(\mathbf{x}_0)$ given the target state $\mathbf{x}_0$ at time t_0 is presented below. It can be seen from Eq. (12) that the Jacobian can be divided into M sub-blocks as:

$$\Gamma(\mathbf{x}_0) = \begin{bmatrix} \Gamma_1(\mathbf{x}_0) \\ \vdots \\ \Gamma_k(\mathbf{x}_0) \\ \vdots \\ \Gamma_M(\mathbf{x}_0) \end{bmatrix} = \begin{bmatrix} -R_{s(1)}^{-\frac{1}{2}}\,[\nabla_{\mathbf{x}_0}\mathbf{h}_1(\mathbf{x}_0)']' \\ \vdots \\ -R_{s(k)}^{-\frac{1}{2}}\,[\nabla_{\mathbf{x}_0}\mathbf{h}_k(\mathbf{x}_0)']' \\ \vdots \\ -R_{s(M)}^{-\frac{1}{2}}\,[\nabla_{\mathbf{x}_0}\mathbf{h}_M(\mathbf{x}_0)']' \end{bmatrix} \quad (73)$$

Each of these $\Gamma_k(\mathbf{x}_0)$ $k = 1\cdots M$ sub-blocks is a 2×6 matrix. Since the Jacobian is formed by stacking up these individual sub-blocks, we shall only consider the computation of the kth sub-block, $\Gamma_k(\mathbf{x}_0)$. Differentiating $\mathbf{h}_k(\mathbf{x}_0)$ in Eq. (4) with respect to $\mathbf{x}_0$ we obtain

$$\Gamma_k(\mathbf{x}_0) = -R_{s(k)}^{-\frac{1}{2}}\,[\nabla_{\mathbf{x}_0}\mathbf{h}_k(\mathbf{x}_0)']' = -R_{s(k)}^{-\frac{1}{2}}\cdot \begin{bmatrix} \sec\theta_k(\mathbf{x}_0) \\ \rho_k(\mathbf{x}_0) \end{bmatrix} \begin{bmatrix} [\nabla_{\mathbf{x}_0}\xi_x(t_k)\sin\phi_k(\mathbf{x}_0)]' - [\nabla_{\mathbf{x}_0}\xi_y(t_k)\cos\phi_k(\mathbf{x}_0)]' \\ [\nabla_{\mathbf{x}_0}\rho_k(\mathbf{x}_0)\sin\theta_k(\mathbf{x}_0)]' - [\nabla_{\mathbf{x}_0}\xi_z(t_k)]' \end{bmatrix} \quad (74)$$

Note that $\nabla_{\mathbf{x}_0}\xi_x(t_k)$, $\nabla_{\mathbf{x}_0}\xi_y(t_k)$ and $\nabla_{\mathbf{x}_0}\xi_z(t_k)$ are the three columns of the 6×3 matrix $\nabla_{\mathbf{x}_0}\boldsymbol{\xi}(t_k)'$. Once again differentiating the range $\rho_k(\mathbf{x}_0)$ in Eq. (5) we obtain

$$\nabla_{\mathbf{x}_0}\rho_k(\mathbf{x}_0) = \frac{[\nabla_{\mathbf{x}_0}\boldsymbol{\xi}(t_k)']\left[\boldsymbol{\xi}(t_k) - \boldsymbol{\eta}_{s(k)}(t_k)\right]}{\rho_k(\mathbf{x}_0)} \quad (75)$$

It can be seen from the above Eqs. (74) and (75) that in order to compute $\Gamma_k(\mathbf{x}_0)$ it only remains for the matrix $\nabla_{\mathbf{x}_0}\xi(t_k)'$ to be calculated. The computation of this matrix involves the evaluation of steps 8 to 12 (in addition to steps 1 to 7) of the algorithm given in Appendix A.

C. Computation of the Fisher information

The prior pdf $p(\mathbf{z}_0|\mathbf{x}_0)$ on the altitude and speed is modeled using an Nth order Butterworth function as:

$$p(\mathbf{z}_0|\mathbf{x}_0) = p(z_{0r}|\mathbf{x}_0)\,p(z_{0v}|\mathbf{x}_0) = \frac{\kappa(\Delta_r, N)}{1-\left[\frac{z_{0r}-r_0}{\Delta_r}\right]^{2N}} \cdot \frac{\kappa(\Delta_v, N)}{1-\left[\frac{z_{0v}-v_0}{\Delta_v}\right]^{2N}} \quad (76)$$

where $\Delta_r = 0.5\,[r_{\max} - r_{\min}]$, $\Delta_v = 0.5\,[v_{\max} - v_{\min}]$; $\kappa(\Delta, N) = \frac{N}{\pi\Delta} sin\left(\frac{\pi}{2N}\right)$ is the normalization constant for the Nth order Butterworth function and $\mathbf{z}_0 = [z_{0r}\ \ z_{0v}]'$ is the 2×1 pseudo measurement on the parameters

$$\mathbf{h}_0(\mathbf{x}_0) = \begin{bmatrix} r_0 \\ v_0 \end{bmatrix} = \begin{bmatrix} \|\xi_0\| \\ \|\dot{\xi}_0\| \end{bmatrix} \quad (77)$$

Since $p(\mathbf{z}_0|\mathbf{x}_0)$ depends on r_0 and v_0 and not directly on $\mathbf{x}_0$, we can expand the integral in Eq. (36) using the chain rule as

$$F_0(\mathbf{x}_0) = [\nabla_{\mathbf{x}_0}\mathbf{h}_0(\mathbf{x}_0)'] \cdot \int_{-\infty}^{\infty} \frac{\left[\nabla_{\mathbf{h}_0(\mathbf{x}_0)} p(\mathbf{z}_0|\mathbf{x}_0)\right] \cdot \left[\nabla_{\mathbf{h}_0(\mathbf{x}_0)} p(\mathbf{z}_0|\mathbf{x}_0)\right]'}{p(\mathbf{z}_0|\mathbf{x}_0)} \mathrm{d}\mathbf{z}_0 \cdot [\nabla_{\mathbf{x}_0}\mathbf{h}_0(\mathbf{x}_0)']' \quad (78)$$

Using the definition of $\mathbf{h}_0(\mathbf{x}_0)$ we obtain

$$\nabla_{\mathbf{x}_0}\mathbf{h}_0(\mathbf{x}_0)' = \begin{bmatrix} \frac{\xi_0}{\|\xi_0\|} & 0_{3\times1} \\ 0_{3\times1} & \frac{\dot{\xi}_0}{\|\dot{\xi}_0\|} \end{bmatrix} \quad (79)$$

The integration in the above equation can be separated in two separate integrals over z_{0r} and z_{0v} because the pdf $p(\mathbf{z}_0|\mathbf{x}_0)$ is separable in z_{0r} and z_{0v}. In addition using Eq. (79) in Eq. (78) yields the following simplification

$$F_0(\mathbf{x}_0) = \begin{bmatrix} \frac{\xi_0}{\|\xi_0\|}\frac{1}{\sigma_r^2}\frac{\xi_0'}{\|\xi_0\|} & 0_{3\times3} \\ 0_{3\times3} & \frac{\dot{\xi}_0}{\|\dot{\xi}_0\|}\frac{1}{\sigma_v^2}\frac{\dot{\xi}_0'}{\|\dot{\xi}_0\|} \end{bmatrix} \quad (80)$$

where

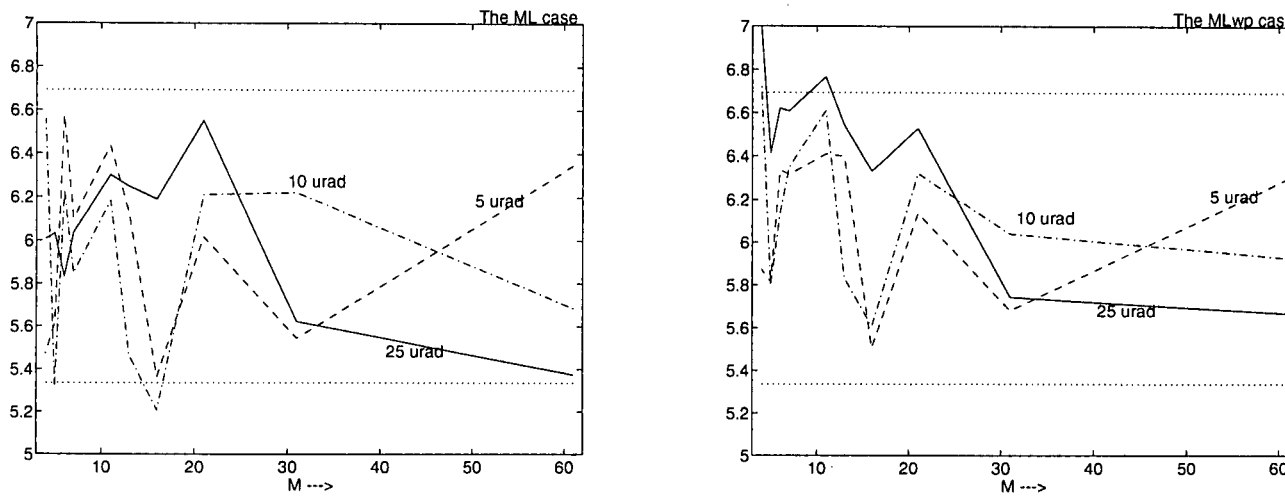
$$\sigma_r^{-2} = \int_{-\infty}^{\infty} \left(\frac{\partial p(z_{0r}|\mathbf{x}_0)}{\partial r_0}\right)^2 \frac{\mathrm{d}z_{0r}}{p(z_{0r}|\mathbf{x}_0)} \quad (81)$$

$$\sigma_v^{-2} = \int_{-\infty}^{\infty} \left(\frac{\partial p(z_{0v}|\mathbf{x}_0)}{\partial v_0}\right)^2 \frac{\mathrm{d}z_{0v}}{p(z_{0v}|\mathbf{x}_0)} \quad (82)$$

Using the definition of $p(z_{0r}|\mathbf{x}_0)$ and $p(z_{0v}|\mathbf{x}_0)$ given in Eq. (76) and simplifying, both the above definite integrals can be reduced to the following form, which has a closed form solution [16]

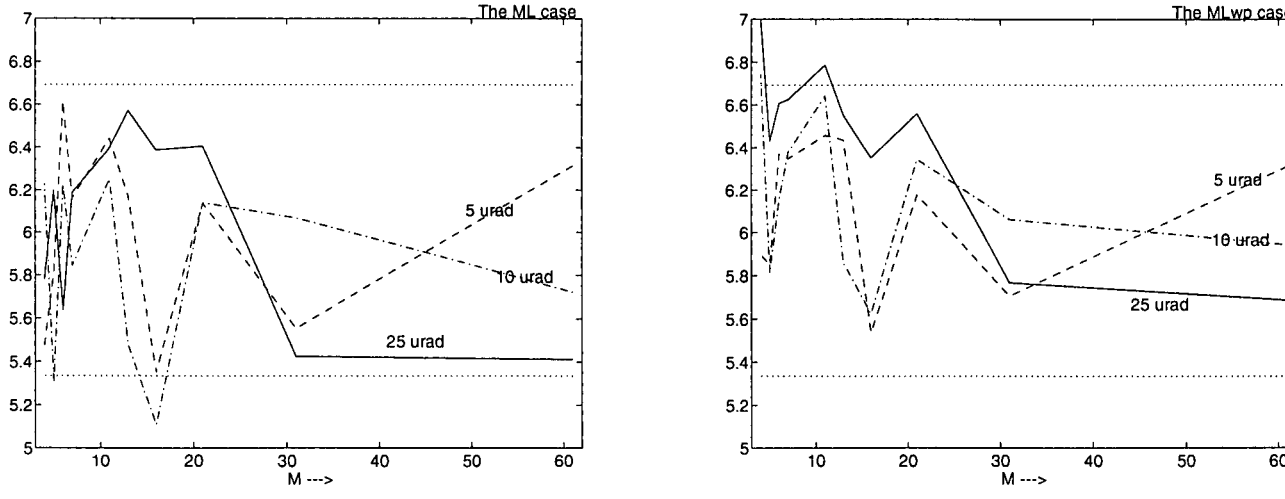
$$\sigma^{-2} = \frac{4N(2N-1)\,\kappa(\Delta, N)}{\Delta} \int_0^{\infty} \frac{z^{(2N-2)}}{[1+z^{2N}]^2}\,\mathrm{d}z = \frac{2-\frac{1}{N}}{\Delta} \quad (83)$$

Substituting the above result back in Eq. (80), the result given in Eqs. (44) and (46) follows.



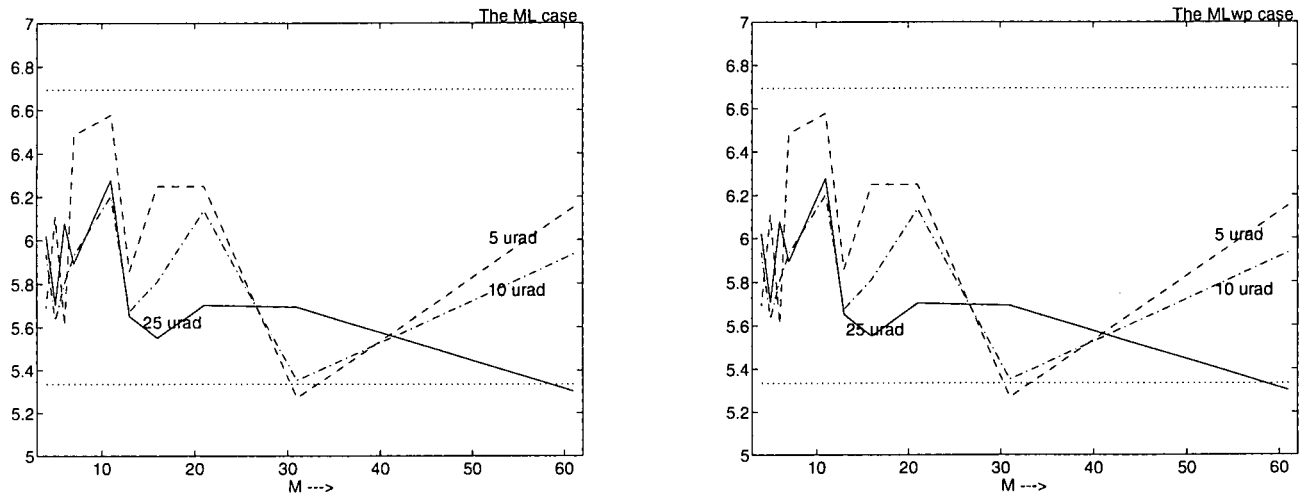
····· 95% Probability Region.

Figure 5: Estimator efficiency for the single sensor scenario with 60 s observation window.



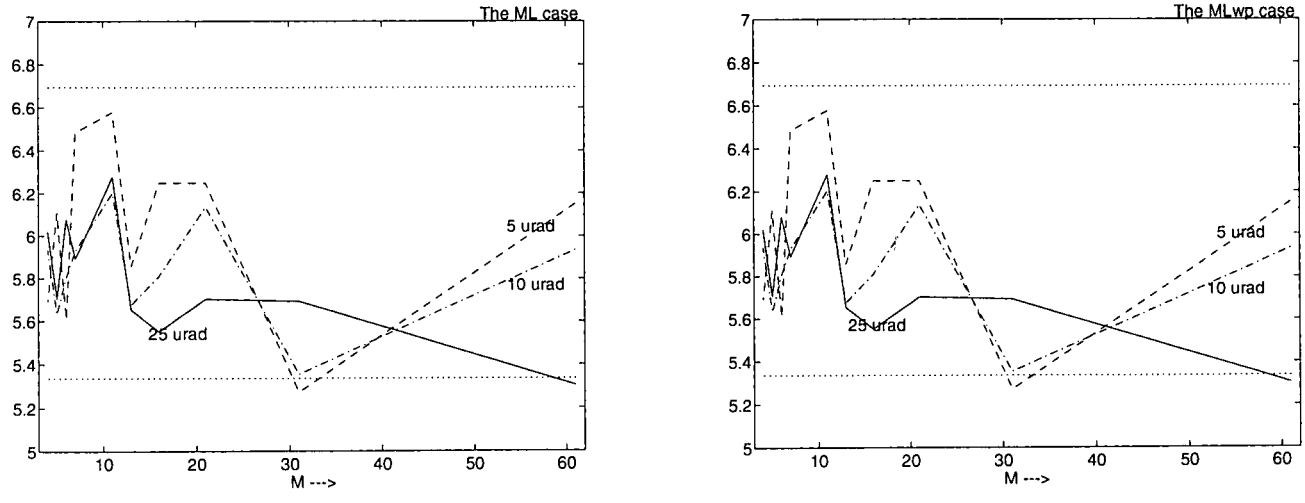
····· 95% Probability Region.

Figure 6: Estimator consistency for the single sensor scenario with 60 s observation window.



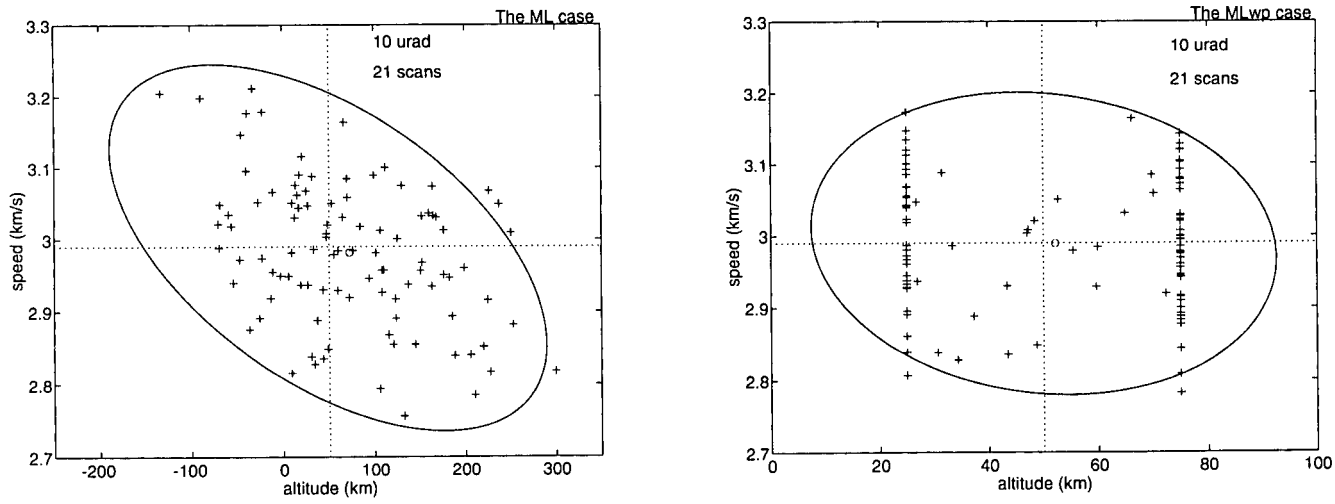
····· 95% Probability Region.

Figure 7: Estimator efficiency for the two sensor scenario with 60 s observation window.



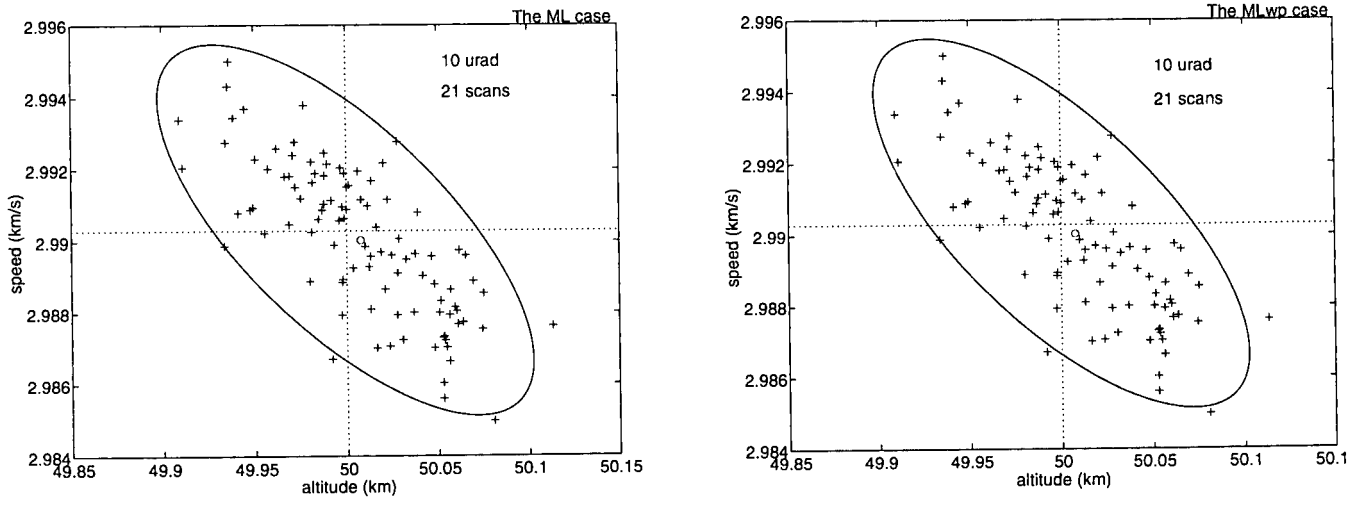
····· 95% Probability Region.

Figure 8: Estimator consistency for the two sensor scenario with 60 s observation window.



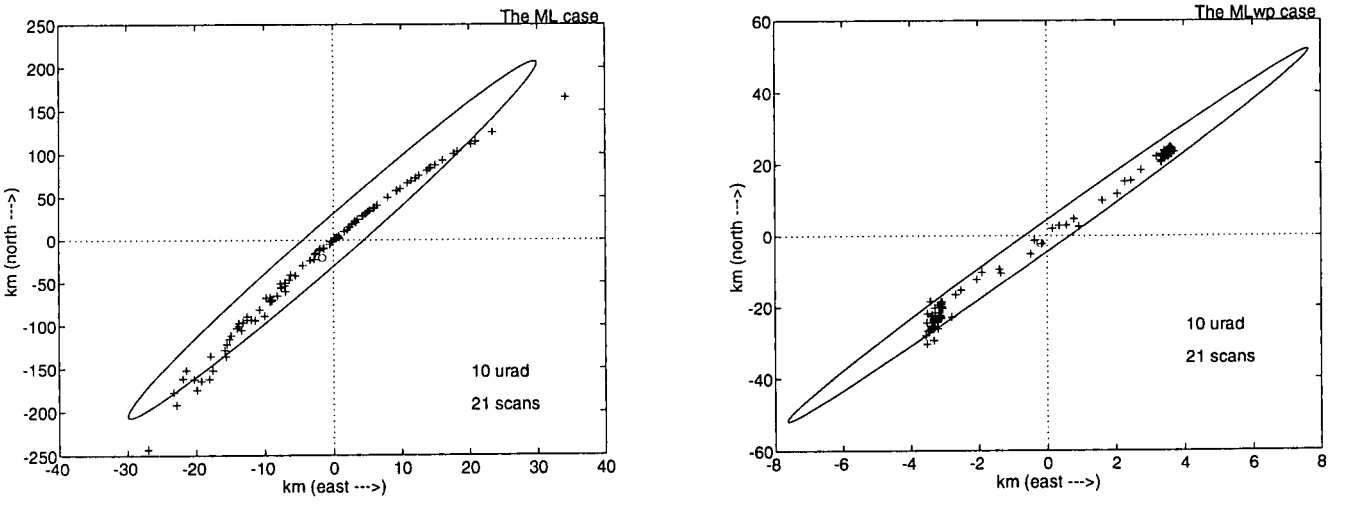
..... True values; —— 2σ Error ellipse (using CRLB); + Sample points; o Sample mean.

Figure 9: Altitude and Speed estimates for the single sensor scenario with 60 s observation window.



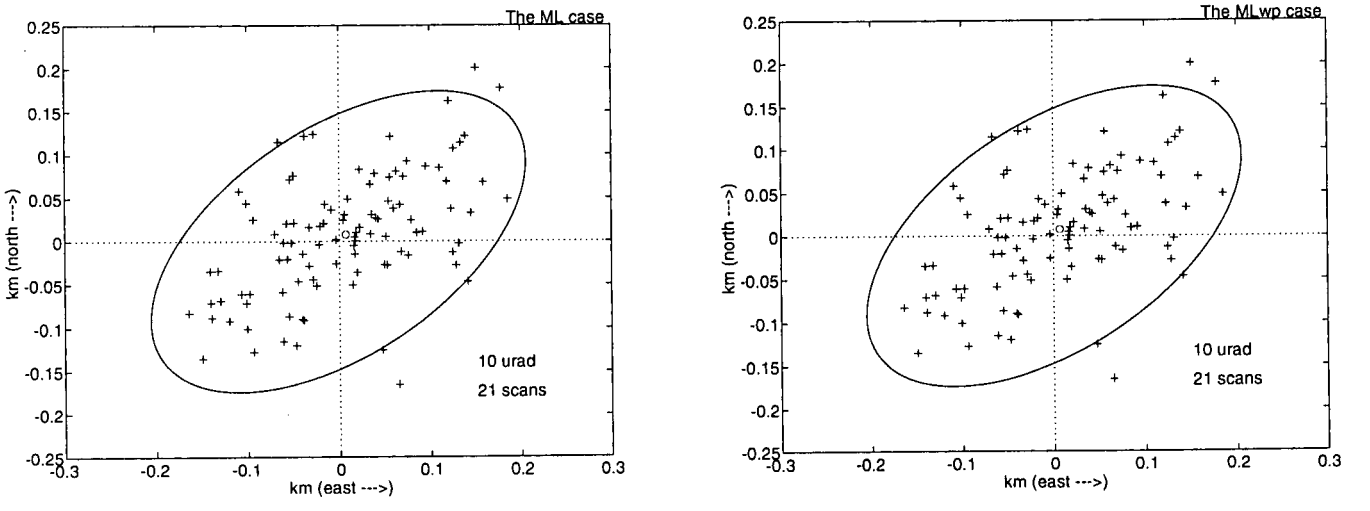
..... True values; —— 2σ Error ellipse (using CRLB); + Sample points; o Sample mean.

Figure 10: Altitude and Speed estimates for the two sensor scenario with 60 s observation window.



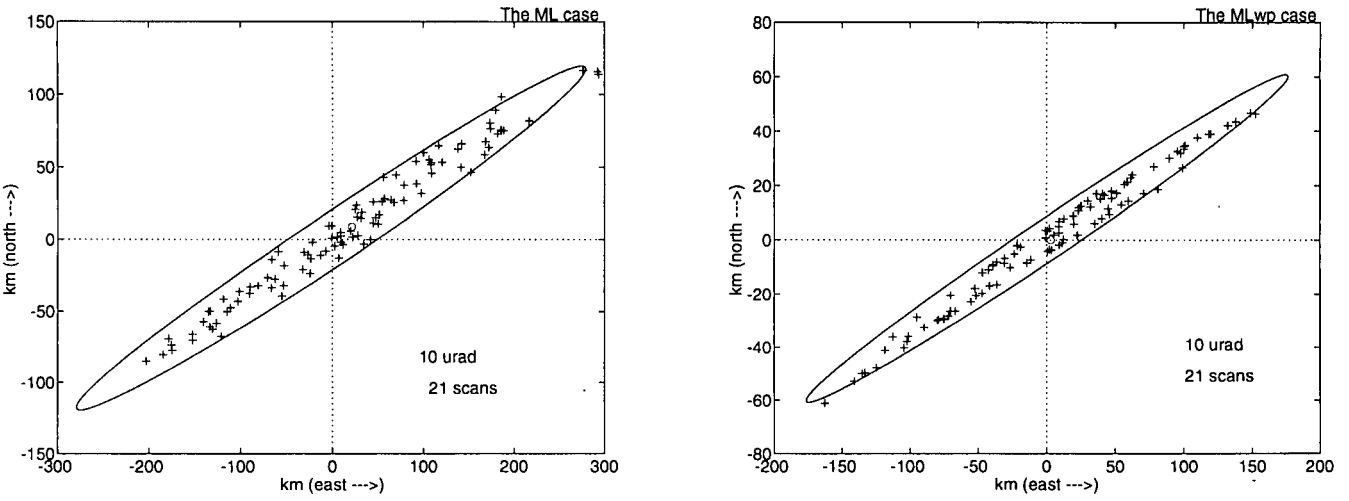
—— : 2σ Error ellipse (using sample covariance); + Sample points; o Sample mean.

Figure 11: Launch Point Error for the single sensor scenario with 60 s observation window.



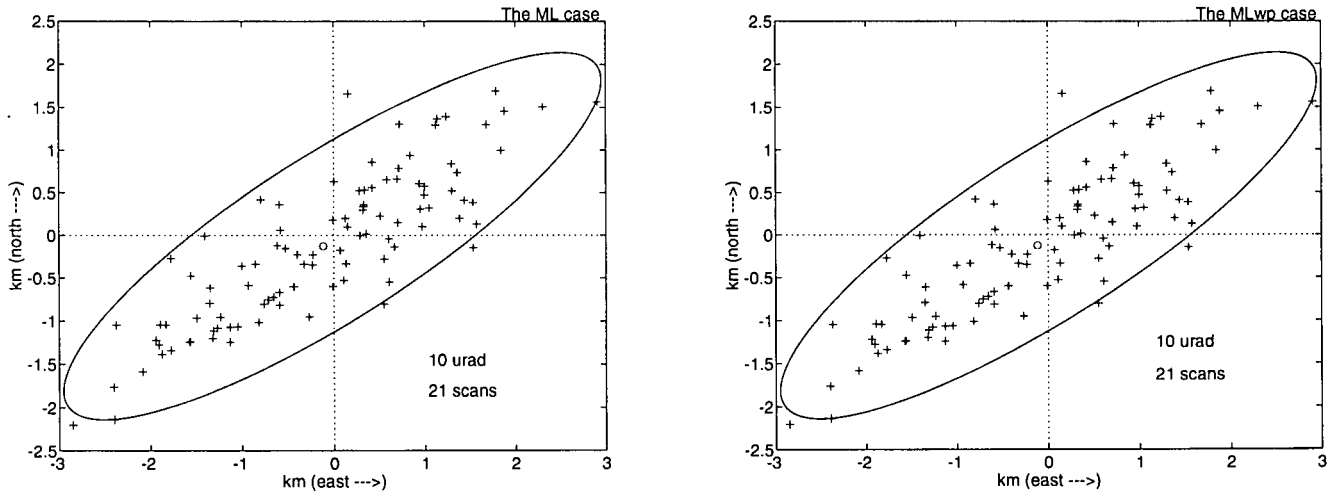
—— : 2σ Error ellipse (using sample covariance); + Sample points; o Sample mean.

Figure 12: Launch Point Error for the two sensor scenario with 60 s observation window.



—— : 2σ Error ellipse (using sample covariance); + Sample points; o Sample mean.

Figure 13: Impact Point Error for the single sensor scenario with 60 s observation window.



—— : 2σ Error ellipse (using sample covariance); + Sample points; o Sample mean.

Figure 14: Impact Point Error for the two sensor scenario with 60 s observation window.

APPLICATION OF MULTIPLE HYPOTHESIS TRACKING TO MULTI-RADAR AIR DEFENSE SYSTEMS

S. S. Blackman, R. J. Dempster and T. S. Nichols
Hughes Aircraft Company
P. O. Box 902
El Segundo, California 90245
USA
(310) 616-1918

SUMMARY

This chapter discusses a multi-radar air defense system implementation that utilizes central-level observation (plot) processing and Multiple Hypothesis Tracking (MHT) data association. The chapter begins with an overview of the air defense system application and a discussion of the reasons MHT data association has been chosen. Next, the transformation of plot data from multiple distributed radars to a common stereographic coordinate system is described and the approach to registration is outlined.

The mathematical basis and the implementation logic for the track-oriented MHT algorithm are described. This includes a discussion of the clustering, pruning and merging methods that have been developed to make implementation feasible. The manner in which the output of the MHT tracker is presented to the user is also described.

A number of features make the multi-radar air defense application more complex than single sensor systems. These include the variety and asynchronous nature of the input data, the potential for variable and heavy clutter densities and stringent false track confirmation requirements. This chapter will present the specific methods that have been developed to adapt the general MHT algorithms to these requirements. It will show how false track requirements can be related to the MHT track confirmation threshold through the use of the Sequential Probability Ratio Test (SPRT).

Finally, analytic and simulation methods that have been developed and applied to predict performance of an air defense tracking system are discussed and representative results are presented.

LIST OF SYMBOLS

a	semi-major axis of ellipsoid
A_c	common radar false plot area
A_i	area containing one false plot in one scan of radar i
b	semi-minor axis of ellipsoid
CV	coverage volume of the sensor
d_{max}	maximum extent of the surveillance region from the origin of the master coordinates
d^2	normalized statistical distance function
DBS	detection bin size
e	eccentricity of ellipsoid
E_m	radius of conformal sphere used in transformation between coordinate planes
E_s	geocentric earth radius at radar site
g	acceleration of earth's gravity
h_e	altitude extent of clutter returns
h_s	height of the radar site above sea-level
h_0	target height above sea-level measured by radar
h'_s	distance of radar site from center of earth
h'_0	distance of target from center of earth
H_0	hypothesis that a set of plots consists entirely of false detections

H_1 hypothesis that a set of plots arises from a single true target

J number of hypotheses

L_{H_j} likelihood score of hypothesis j

L_i likelihood score of track i

M sensor measurement dimensionally

N_{FA} average number of false alarms per second

N_R number of radars in system

$\hat{P}_D$ estimated probability of detection

P_{FA} false alarm probability

$P(T_i)$ probability of track validity

R transformed measurement covariance of target position error

R_0 target range measured by radar

R_g Ground range from radar to target

R_{max} Maximum range of radar

R_{min} Minimum range of radar

S residual covariance matrix

T_b basic scan interval for application of miss penalty

$\overline{T}_C$ expected true track confirmation time

$\overline{T}_{FC}$ expected time between false track confirmations

T_p processing interval length

T_s scan time required by the sensor to cover volume cv

T_{s_i} scan time of radar i

T_{SR} equivalent radar sampling interval

T_1 track deletion threshold

T_2 track confirmation threshold

T_{90} the expected time at which 90 percent of target tracks will be confirmed

w target location with respect to master stereographic plane in complex coordinates

w_s radar site location with respect to master stereographic plane in complex coordinates

$\overline{w}_s$ complex conjugate of w_s

x x-coordinate of transformed target position

x_0 x-coordinate of target position with respect to the local stereographic plane

x_g x-component of target ground range

y y-coordinate of transformed target position

y_0 y-coordinate of target position with respect to the local stereographic plane

y_g y-component of target ground range

$\underline{\tilde{y}}$ measurement residual vector

z target location with respect to local stereographic plane in complex coordinates

α false track acceptance probability

β true track rejection probability

β_{ms} angle of rotation between local and master stereographic planes

β_{FT} false target density

β_{FT_i} false target density for radar i

β_{FTN} nominal (average) false target (noise, clutter) density

β_{NT} new target density

β_{NTN} nominal (average) new target density

ε_0 target elevation measured by radar

ε_{lim} radar elevation limit

η_0 target azimuth measured by radar

η_{lim} radar azimuth limit

$\Delta L(k)$ track score increment at time interval k

$\Delta\lambda$ difference between longitudes of radar site and master stereographic coordinate system

λ_m longitude of master stereographic coordinate system

λ_s longitude of radar site

φ geodetic latitude

φ_m geodetic latitude of master stereographic coordinate system origin

φ_s geodetic latitude of radar site

Φ conformal latitude

Φ_m conformal latitude of stereographic coordinate system origin

Φ_s conformal latitude of radar site

ρ correlation of transformed target x and y position coordinates

σ_{h_0} standard deviation of measured target height

$\sigma^2_{R_g}$ variance of stereographic ground range

σ_{R_0} standard deviation of measured target slant range

σ^2_x variance of transformed target x position

σ^2_{xy} covariance of transformed target x,y position

σ^2_y variance of transformed target y position

σ_{η_0} standard deviation of measured target azimuth

ψ central spherical angle between the radar site and stereographic coordinate system origin

1. TARGET TRACKING

This chapter addresses target tracking in the context of air defense. The air defense mission is to defend assets against hostile aircraft and missiles by intercepting them with friendly aircraft or missiles. Both hostile and friendly targets may employ sharp maneuvers, up to 9 g for aircraft and even higher for missiles.

The goal of the target tracking function is to process sensor observations to provide accurate target position and velocity estimates, to maintain continuity of target identification, and to discriminate targets of interest (aircraft and missiles) from noise and clutter. Target measurements for this purpose come from two types of sensors, secondary surveillance radars (SSRs) and primary surveillance radars (PSRs).

1.1 Air Defense Radars

SSRs transmit interrogation signals. They detect and measure the responses of aircraft equipped with compatible transponders. As a result, SSRs detect only aircraft with active transponders, and they are not useful in tracking unequipped civilian aircraft or hostile aircraft. Based on the measured target parameters, an SSR outputs target observations, commonly known as plots, comprising the measured target azimuth and range, transponder codes (used for target identification) and, in some cases, target height as reported by the transponders

PSRs transmit radio or microwave pulses. They detect and measure the reflections of these pulses from objects in the environment. Objects that reflect sufficient energy back to the PSR are detected. As a result, primary radars detect birds, automobiles, swarms of insects, ocean waves, mountains and other extraneous objects. These extraneous observations are known as clutter. PSRs report plots comprising azimuth, range and, in some cases, height. However, many older PSRs do not measure target elevation and are thus unable to report height.

Capabilities of PSRs vary widely. While some high data rate PSRs are being employed, most are not very accurate (e.g. the MPDR-90, with a 90 kilometer range, 4 second data rate, and accuracies of 8 milliradians in azimuth and 250 meters in range). More typical are the long range, low-data rate PSRs, typified by the

AN/FPS-117, with a 12 second data rate, maximum range of 350-450 km, and accuracy of 250 meters in range and 2.5 milliradians in azimuth. PSRs are often positioned to provide overlapping coverage for the air defense mission, resulting in measurement update rates of 1 per second or more on close-in targets.

1.2 Plot-to-Track Association

Given a set of radar measurements, the target tracker must associate the new measurements of the targets with the tracks that represent them. The efficacy of this function determines whether the target is consistently identified with a single track. Errors or changes in track identity can result in track swaps or track loss. In the air defense mission, such ambiguities can result in significant errors such as firing on friendly aircraft.

The degree of difficulty of associating observations with the correct tracks depends on the type of sensor. An aircraft transponder provides an identification code to an interrogating SSR. While this code can be garbled in a number of ways, it provides a reasonably reliable means of associating observations with tracks. Some radar sites support both a PSR and an SSR aligned on the same mounting. The measurements of the two radars are often combined to form reinforced plots, constructed using the better source for each reported parameter. Observations made by PSRs alone have no identification codes. Thus, association of a PSR observation must be made on the basis of proximity of the measured target position to a track position estimate, an association that can have a high degree of uncertainty. The use of Multiple Hypothesis Tracking (MHT) for making reliable association decisions under uncertainty is the principle subject of this chapter.

1.3 Environmental Considerations

In air defense applications, the tracking environment can significantly affect the performance of the tracker. Noise jamming can be used by hostile forces to degrade or deny target detection. However, the newer generations of radars use adaptive nulling techniques to reduce the effects of noise jamming by suppressing the received signal at the jammer azimuths. Other electronic countermeasures (ECM) use deception to avoid detection by radars.

Clutter observations compound the difficulty of associating observations with tracks. If one or more clutter and target observations can be associated with the same track, then the likelihood of making an incorrect association increases.

1.4 Target Tracking Filters

When a measurement is associated with a track, it is used to update the track estimate of target position, velocity, and (in some cases) acceleration. This is typically accomplished with one of several types of tracking filters. The accuracy of the target position and motion estimates depend in part on how well the kinematic model (or models) used in the filter matches the target motion. If the range of maneuvers that the target can execute is small and well-defined, it is often possible to construct a filter that can provide very accurate position and velocity estimates, given accurate sensor measurements and a high data rate. If the target maneuvers do not match the filter's kinematic model(s), if the data rate is too low, or if the sensor measurement accuracy is poor, then track position and velocity estimate will be less accurate. In air defense applications, a trade-off often must be made between accuracy of the filter in tracking constant-velocity targets and accuracy and responsiveness of the filter during target maneuvers.

In the 1960s, low computer throughput rates necessitated the use of alpha-beta filters for target tracking. A tracking filter updates the track position and velocity by forming a weighted average of the predicted track position and the plot position. The alpha-beta filter uses a fixed gain (weighting) or a gain schedule. The gain schedule decreases the gains over time to an assumed steady-state level. A constant velocity target motion model is used and a maneuver detection scheme is employed to determine when the target dynamics no longer match the constant velocity model. At this point, the filter reverts to higher gains, allowing the filter to give more weight to target measurements. When the maneuver is completed, the gain schedule is again employed to force the filter to steady state convergence.

As computer throughput increased in the 1970s, Kalman filtering was employed in target tracking. Kalman filtering has the advantage of dynamically adapting gains to changing target trajectories at the expense of increased computational complexity. It allows faster convergence to a steady state. However, it assumes a single target motion model and a maneuver detector is still required to determine when the target's motion does not match the model. The methods used to respond to these occurrences are similar to those used with alpha-beta filters. For a discussion of tracking filters in general and the alpha-beta and Kalman filtering methods in particular, see [1].

In the 1980s, multiple model filters were developed. These tracking filters employ several Kalman filters running in parallel, each with a different target motion model. The goodness of the fit between the target measurements and the motion models is monitored. At any given time the filter state of the best model, or an appropriately weighted combination of filter states, is selected for further processing and for display. A successful multiple model filter for air defense applications is the Interactive Multiple Model (IMM) filter [2]. Several variations of IMM filters are being employed in the current generation of air defense trackers.

Simulation results for the system described in this chapter were obtained using Kalman filtering with maneuver detection.

1.5 Sensor Level Tracking vs. Centralized Tracking

Two philosophies of target tracking are used in air defense, sensor level tracking and centralized tracking. The sensor level tracking approach forms tracks using the output of each radar and then combines the tracks formed on each target to form a composite track. The centralized approach processes plots from multiple radars directly to form composite tracks.

The advantage of the sensor level method is that each sensor's data can be processed independently to form tracks using a simplified tracker and a computer with relatively low processing speed. Processing is distributed across multiple computers. The communication bandwidth required for sensor level tracking is small since only the processed tracks need to be transmitted to a central facility. However, in most applications, the plot data is desired for display and must be transmitted to the central level anyway.

The centralized tracking approach has the advantage of using all the information contained in the plots, as opposed to the limited information contained in single sensor tracks, to form the composite tracks. As a result, it produces more accurate tracks and has the potential to correctly resolve ambiguities more frequently than the sensor level approach. This improved tracking performance comes at the expense of a more complex tracker and higher performance computing resources.

Radars in an air defense system may be separated by hundreds of kilometers. Thus, both tracking philosophies require special methods to transform the multiple radar data to a common coordinate system such as the stereographic coordinate system. Also, registration methods are required to reduce multi-radar bias errors. The stereographic coordinate transformations and registration are discussed in Section 2.

1.6 MHT vs. Other Target Tracking Approaches

The MHT approach to target tracking differs from other approaches in the way in which plots are associated with tracks. This is a crucial step in any tracking approach. Association errors can cause inaccurate track state estimates, track loss and track misidentification. Where other methods make an immediate and irrevocable decision about which plots to assign to each track, MHT forms several branches on each track representing different possible associations. As time passes and additional plots are received, MHT eliminates branches which prove to be unlikely, based on a likelihood scoring method. In this way, it uses the evolution of plot histories over time to make better association decisions. Maintaining several branches for each track and deciding which branches to eliminate consumes substantially more processing than conventional tracking methods. However, with the exponential rate of increase in computer processing speed, the additional processing required has ceased to be an impediment to

implementing MHT trackers with high track capacity.

The MHT method, its application to the air defense system tracking problem and representative performance results are presented in Sections 3 through 5.

2. COORDINATE CONVERSION AND REGISTRATION

This section describes a method for converting radar measurements to a common stereographic coordinate system. It also deals briefly with the problem of radar registration.

2.1 The Stereographic Coordinate System

The problem of combining measured range, azimuth angle and either elevation angle or height from multiple distributed radar sites arises in the design of air defense systems. The aircraft tracked by these systems, as well as potential missile targets, typically maintain a nearly constant altitude above sea level. Thus, it is desirable to define a common coordinate system using x, y and altitude coordinates. A preferred system using these coordinates is the stereographic system defined below and described in more detail in [3-10].

The basic principles of stereographic coordinates are illustrated in Figures 1 and 2.

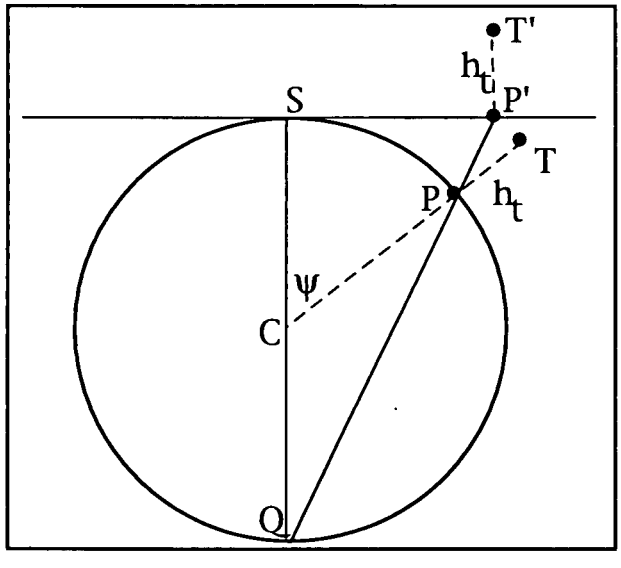


Figure 1 . Stereographic Projection. The projection of a target T with respect to a plane tangent to the earth at S is represented by T'.

The stereographic plane is drawn tangent to the surface of the earth at the center of the coordinate system. The target position (T') is represented with respect to this coordinate system by first projecting its true position (T) onto a point (P) on the surface of the earth. Then, the intersection point P' of the line drawn from the perspective point Q through P with the stereographic plane defines the target (x, y) position. Finally, the z (or altitude) position is set to the target height (altitude) above the earth's surface.

Several steps in the process of converting a radar measurement to stereographic coordinates use a spherical earth approximation. In actuality, the earth is better approximated as an ellipsoid which can be obtained by revolving an ellipse about its semi-minor axis. Defining a to be the semi-major axis (or equatorial radius) and b to be the semi-minor axis (polar radius), the degree of departure from circularity of the ellipse, as used in the required transformations, is defined by the eccentricity,

$$e^2 = 1-(b/a)^2 \tag{1}$$

A widely accepted standard is the WGS-84

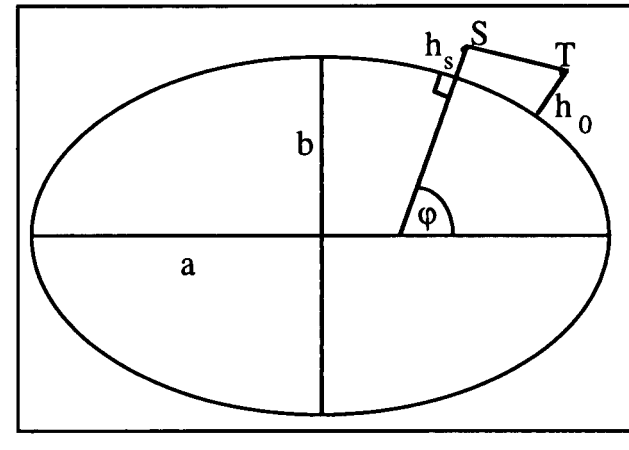


Figure 2. The Radar-Target Geometry with Respect to an Ellipsoidal Earth Model. The point S represents a radar site positioned at geodetic latitude φ and at an altitude of h_s. The target T is at an altitude h_0 and at a slant range given by the line segment ST. It is interesting to note that there is no closed form expression for the geodetic latitude of T given the radar slant range, azimuth and altitude.

ellipsoid for which a is 6,378,137 m and e^2 is 0.006694381.

The conversion of a measurement from a radar at a location displaced from the center of the stereographic coordinates to a position with respect to these coordinates requires several steps. Before processing any observations, the position of the radar site with respect to the system (or master) stereographic coordinate system is computed for later use in the conversion of the observations. Then, three steps are required to convert an observation for use in a tracking system centered at the master stereographic coordinate system.

First, the measurement is converted to a local stereographic system, centered at the radar. Second, a transformation is made from the radar site stereographic coordinate system to the master coordinate system. Finally, the radar measurement errors are converted to a measurement error covariance with respect to the master Cartesian coordinate system.

2.1.1 Computation of Radar Site Position Relative to Master Coordinates

The center of the master stereographic coordinate system is defined to be at the geodetic latitude φ_m and longitude λ_m. Similarly, the radar site is defined to be at geodetic latitude φ_s and longitude λ_s. (Latitudes are defined to be positive to the north of the equator and negative to the south. Longitudes are defined to be positive to the east of the prime meridian and negative to the west.) Since latitude and longitude are defined relative to the ellipsoidal earth and conformal mapping is relative to a sphere, a transformation of the geodetic latitude to a conformal latitude that is defined relative to a sphere is required. The longitude is unchanged. The conformal latitude Φ is computed from geodetic latitude φ according to the relationship [10]

$$\tan\left(\frac{\Phi}{2}+\frac{\pi}{4}\right)=\tan\left(\frac{\varphi}{2}+\frac{\pi}{4}\right)\left[\frac{1-e\sin\varphi}{1+e\sin\varphi}\right]^{e/2} \tag{2}$$

Once the conformal latitudes of the master system origin Φ_m and of the site origin Φ_s are computed from Eq. (2), the position components (u_s, v_s) of the site on the master stereographic plane are given by [10]

$$u_s=\frac{2E_m\sin\Delta\lambda\cos\Phi_s}{1+\cos\psi} \tag{3}$$

$$v_s=\frac{2E_m\left(\sin\Phi_s\cos\Phi_m-\cos\Phi_s\sin\Phi_m\cos\Delta\lambda\right)}{1+\cos\psi} \tag{4}$$

where

$$\Delta\lambda=\lambda_s-\lambda_m$$

$$\cos\psi=\sin\Phi_s\sin\Phi_m+\cos\Phi_s\cos\Phi_m\cos\Delta\lambda$$

The value of E_m in (4) is chosen to balance magnification effects between the center and the edges of the surveillance region. A good empirical value for E_m is

$$E_m = E\left[\frac{3}{4}+\frac{1}{4}\cos\left(\frac{d_{max}}{E}\right)\right] \tag{5}$$

where:

E = geocentric earth radius

$$= a\left[\frac{\cos^2\varphi_m+\left(1-e^2\right)^2\sin^2\varphi_m}{1-e^2\sin^2\varphi_m}\right]^{1/2}$$

d_{max} = maximum extent of the surveillance region from the origin of the master coordinates

Finally, the angle β_{ms} that will be required later for the transformation from the radar site to the master coordinates is given by [10]

$$\beta_{ms}=\tan^{-1}\left[\frac{-\left(\sin\Phi_s+\sin\Phi_m\right)\sin\Delta\lambda}{\cos\Phi_s\cos\Phi_m+\left(1+\sin\Phi_s\sin\Phi_m\right)\cos\Delta\lambda}\right] \tag{6}$$

The site position coordinates (u_s, v_s) in the master stereographic plane and the height of the site above sea-level h_s specify the position coordinates of the site with respect to the master coordinate system. These coordinates and the angle β_{ms}, given in Eq. (6), are used in the conversion, discussed next, of the radar data measured at the site into measurements with

respect to the master stereographic coordinate system.

2.1.2 Transformation of Radar Site Measurement to Master Coordinates

Each radar measurement consisting of slant range R_0, azimuth angle η_0 and either height above sea level h_0, or elevation angle ε_0 is first converted to a local stereographic plane that is tangent to the Earth at the radar site. The stereographic coordinates (x_0, y_0) of the radar report with respect to the local plane are given by

$$x_0 = x_g - 2Ax_g y_g \tag{7}$$

$$y_0 = y_g + A\left(x_g^2 - y_g^2\right) \tag{8}$$

where

$$A = \frac{b-a}{2a^2}\sin\left(2\varphi_s\right)$$

$$x_g = R_g \sin\eta_0$$

$$y_g = R_g \cos\eta_0$$

The stereographic ground range R_g is given by [7,8]:

$$R_g = 2E_m\left[\frac{F^2}{4\left(E_s+h_s\right)\left(E_s+h_0\right)-F^2}\right]^{1/2} \tag{9}$$

where:

$$F^2 = R_0^2 - \left(h_0 - h_s\right)^2$$

$$E_s = a\left[\frac{\cos^2\varphi_s + \left(1-e^2\right)^2\sin^2\varphi_s}{1-e^2\sin^2\varphi_s}\right]^{1/2}$$

The measured azimuth angle η_0 is the measurement after correction for registration error as discussed in Section 2.2

The measured elevation angle ε_0 (corrected at the radar for atmospheric refraction) can be used, along with measured range and radar height, to compute measured target height above sea level. The conversion is given by

$$h_0 = \sqrt{\left(E_s+h_s\right)^2 + 2R_0\sin\varepsilon_0\left(E_s+h_s\right) + R_0^2} - E_s \tag{10}$$

Next, the position (x_0, y_0) with respect to the site stereographic coordinates is converted to position (x, y) in the master coordinates. Note that the third coordinate (h = height above sea level) does not require further conversion. The transformation relating (x_0, y_0) to position components (x, y) with respect to the master coordinate system can be expressed in complex notation by [10]

$$w = \frac{w_s + z\exp\left(-i\beta_{ms}\right)}{1 - \dfrac{\overline{w}_s z\exp\left(-i\beta_{ms}\right)}{4E_m^2}} \tag{11}$$

where

$$w = x + iy$$

$$w_s = u_s + iv_s$$

$$\overline{w}_s = u_s - iv_s$$

$$z = x_0 + iy_0$$

A solution to Eq. (11) can be obtained through the second order approximation [10].

$$x = u_s + kx_1 + C(x_1^2 - y_1^2) + 2Dx_1y_1 \tag{12}$$

$$y = v_s + ky_1 - D(x_1^2 - y_1^2) + 2Cx_1y_1 \tag{13}$$

where

$$C = \frac{ku_s}{4E_m^2}$$

$$D = \frac{kv_s}{4E_m^2}$$

$$x_1 = x_0\cos\beta_{ms} + y_0\sin\beta_{ms}$$

$$y_1 = y_0\cos\beta_{ms} - x_0\sin\beta_{ms}$$

$$k = 1 + \frac{u_s^2 + v_s^2}{4E_m^2}$$

The solution given by Eqs. (12) and (13) is very accurate as long as the coordinate centers are

within about 2000 km of each other and the measurement displacements are about 300 km or less. For this case an approximate error bound is given by [10]

$$|\text{Error}| < \frac{|z|^3 |w_s|^2}{16E_m^4} \approx \frac{\left(3x10^5 m\right)^3 \left(2x10^6 m\right)^2}{16\left(6.4x10^6 m\right)^4} \quad (14)$$

$$< 5m$$

If larger distances are involved or greater accuracy is desired, exact values for x and y can be obtained from Eq. (11) by multiplying the numerator and denominator by the complex conjugate of the denominator.

2.1.3 Master Site Measurement Covariance Matrix

Central-level tracking in master site coordinates requires a measurement covariance matrix associated with the converted position measurement. The measurement error associated with the z (height) component and the target dynamics in z are typically considered to be very loosely coupled with those in the (x, y) components. Thus, the error in measured target height h_0 can be computed directly and input to an uncoupled z direction tracking filter. Then, a measurement error covariance that accounts for the correlation in the x and y measurement components is computed and input to a coupled x-y tracker.

The transformation converting the range, height and azimuth angle error variances $\sigma^2_{R_0}$, $\sigma^2_{h_0}$ and $\sigma^2_{\eta_0}$ associated with the radar site measurements to a covariance matrix R defining x and y components measurement errors with respect to the master coordinate system is given by

$$R = \begin{bmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{bmatrix} = B \begin{bmatrix} \sigma_{R_g}^2 & 0 \\ 0 & \sigma_{\eta_0}^2 \end{bmatrix} B^T \quad (15)$$

where

$$B = \begin{bmatrix} c & -d \\ d & c \end{bmatrix} \begin{bmatrix} f & g \\ -g & f \end{bmatrix} \begin{bmatrix} s & -t \\ t & s \end{bmatrix} \begin{bmatrix} m & R_g n \\ n & -R_g m \end{bmatrix} \quad (16)$$

$$c = k + 2Cx_1 + 2Dy_1$$

$$d = 2Cy_1 - 2Dx_1$$

$$f = \cos\beta_{ms}$$

$$g = \sin\beta_{ms}$$

$$s = 1 - 2Ay_g$$

$$t = 2Ax_g$$

$$m = \sin\eta_0$$

$$n = \cos\eta_0$$

$$\sigma_{R_g}^2 = G\left\{\left(2h'_0 R_0\right)^2 \sigma_{R_0}^2 + \left[R_0^2 + h'^2_0 - h'^2_s\right]^2 \sigma_{h_0}^2\right\} \quad (17)$$

$$G = \frac{R_g^6 h'^2_s}{4(E_m F^2)^4}$$

$$h'_s = E_s + h_s$$

$$h'_0 = E_s + h_0$$

The matrices containing c, d, s and t in Eq. (16) differ only slightly from identity matrices and can be dispensed with in most applications. Also, because $\sigma^2_{R_g}$ changes only slightly with changes in h'_0 and R_0, the right hand side of Eq. (17) can be replaced with an application-specific upper bound on $\sigma^2_{R_g}$.

2.1.4 Example

Application of the transformation defined above is illustrated using simulated data from three proposed radar sites for an air defense system currently under development. Table 1 gives the longitude, the geodetic and conformal latitudes and the height of the three sites and the origin of the proposed master stereographic coordinate system. Table 1 also gives the computed positions (u_s, v_s) of the three sites with respect to the origin of the master stereographic plane.

Next, Table 2 gives the result of transforming 3 simulated measurements from the same target which was taken to be moving along the X axis

Table 1. Master and Site Locations for Example Stereographic Coordinate System

Site	Longitude (rad)	Latitude (rad) Geodetic	Conformal	Height (m)	Location Relative to Master Coordinates u_s (km)	v_s (km)
Master	0.146660	0.828730	0.825385	0	0.00	0.00
1	0.144927	0.821914	0.818576	425	-7.54	-43.40
2	0.150777	0.827603	0.824268	497	17.79	-7.15
3	0.120371	0.817859	0.814509	511	-114.82	-68.12

Table 2. Example Results for 3 Sites Measurements Transformed to Master Stereographic Coordinates

Time	Site	Measured Az. Angle (rad)	Range (m)	Transformed Position (m) x	y	Error Statistics (m) σ_x	σ_y	ρ
1.01	3	1.0550	101,217	-25,856	-19,918	296	535	-0.96
3.57	1	5.6039	30,276	-26,520	-19,829	148	126	0.73
4.23	2	4.4344	46,274	-26,648	-19,979	102	267	-0.70

at a constant velocity of about -200 m/s. Table 2 gives the time, site and measured range (m) and azimuth angle (rad) of the three measurements taken at the three sites. The measured height h_0 was taken to be 400m in all cases and d_{max} was taken to be 300 km. Finally, Table 2 also presents the measurement covariance values computed using Eq. (15) assuming the following measurement accuracies for all three sites.

$$\sigma_{\eta_0} = 6.0 \text{ mrad}, \quad \sigma_{R_0} = 70 \text{ m}, \quad \sigma_{h_0} = 1000\text{m}$$

The final entry in Table 2 is defined to be

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

2.2 Radar Registration

The coordinate transformations given in the preceding sections assume that the radars are precisely calibrated to produce measurements free of any systematic biases. This is rarely the case. Systematic radar measurement biases are estimated in a process called registration. The estimated biases are removed from the radar measurements before they are converted to system coordinates. Reference [11] gives a detailed discussion of the registration methods that will be outlined below.

The most important parameters estimated in the registration process are range and azimuth biases. Large biases in these measurements can result in forming multiple tracks on a single target. Smaller biases can result in unstable track state estimation. Removing them allows proper functioning of plot-to-track correlation in central-level tracking and track-to-track correlation in sensor-level tracking.

Biases in time of detection and target elevation can also contribute to correlation errors. However, they are frequently not registered in air defense systems. Instead, time biases are often estimated based on known time delays associated with radar signal processors and communication links.

Elevation biases depend on a variety of factors such as atmospheric refraction of radar beams, bending of the antenna pedestal by winds, and other factors, resulting in biases that change rapidly over time and target position. The non-homogeneity of elevation biases makes their estimation both difficult and of little value. In practice, less reliance is placed on PSR height measurements than on azimuth and range in the plot-to-track correlation process.

Registration is usually performed on pairs of radars to eliminate relative biases between them. A prerequisite to the registration process is the identification of plots generated by the

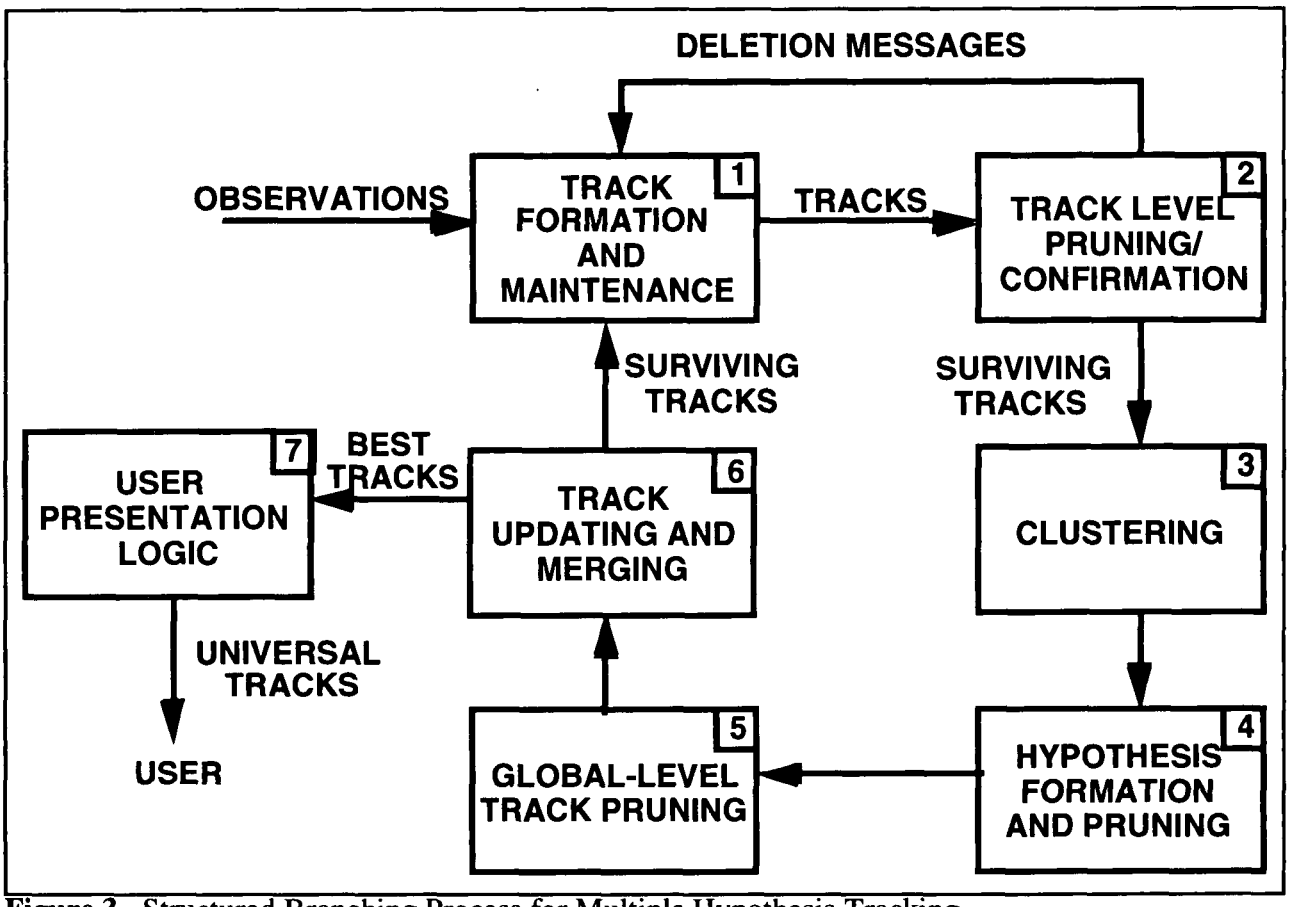


Figure 3. Structured Branching Process for Multiple Hypothesis Tracking.

two radars that represent the same target. This is accomplished either by associating SSR plots with matching identification codes, or by matching PSR plots according to their proximity. In the central-level tracking approach, the results of plot-to-track correlation can provide these correlated plots directly. Selected correlated plots are used in registration algorithms such as those described in [11] to reduce bias errors by an order of magnitude.

3. MHT ALGORITHM DESCRIPTION

Hughes Aircraft Company has developed a track-oriented MHT algorithm that is being applied to a variety of single and multiple radar and IR tracking applications. This section will give an overview of the general MHT logic flow shown in Figure 3. It will describe the clustering, pruning and merging methods that have been developed to make implementation possible. The score function used to evaluate alternative data association hypotheses will be defined and a unique method that has been developed to present MHT data to the user will be described. The next section will specifically address the manner in which the MHT algorithm has been tailored towards the multi-radar air defense application.

3.1 Functional Description of Track-Oriented MHT

The track-oriented approach to MHT starts by independently forming tracks. Our approach [12-15] has been denoted Structured Branching (SB) and is similar to other track-oriented MHT methods described in the literature [16-19]. Using this approach, observations (denoted plots) are formed into tracks without imposing the usual constraints that a plot not be used to update more than one track and that a track not be updated by more than one plot. The tracks that are formed may not be consistent with each other—two tracks may both use the same plot, for example. These inconsistencies are resolved later through the formation and evaluation of

hypotheses composed of sets of consistent tracks.

In order to satisfy computational constraints and produce information that can readily be interpreted by a user, it is necessary to limit the number of track hypotheses. The main method for doing this is to delete (or prune) unlikely tracks. As detailed below, pruning is performed at two stages. First, individual track hypotheses are compared with the hypothesis that all included plots are false alarms. Then, tracks that survive this first test (versus the false alarm option) are compared at the global level by forming, evaluating and pruning hypotheses. Next, the blocks in the flow chart of Figure 3 will be described.

3.1.1 Track Formation and Maintenance

The first element of Figure 3, Track Formation and Maintenance, represents the central track file where all tracks are maintained and the operations performed there. As plots are received, standard gating methods [1] are used to determine viable plot to track pairings. Then, existing tracks are updated with all plots within the gates and extrapolated, no-update-observation, tracks are formed. Also, essentially every plot is used to form the first point of a new track. Thus, a great many tracks are potentially formed and many of the tracks are inconsistent in the sense that the same plots are used for more than one track.

A list is compiled for each track specifying the plots that compose the track. This list is used in the clustering, compatibility testing, and merging logics described below.

3.1.2 Track Level Pruning and Confirmation

Each track has a track level probability that can be computed from its score. The score is the log likelihood ratio, or the log of the probability that the track is valid divided by the probability that all observations are false alarms. The track-level pruning process compares the track-level probability to a suitably chosen deletion threshold. The tracks that fail this test are deleted. The surviving tracks are tested for confirmation and passed to the next stage, which is clustering. Track confirmation status is used later to determine the eligibility of tracks for presentation to the user.

3.1.3 Clustering

The process of clustering is the collection of all tracks that are linked by common observations. Tracks that share observations are defined to be incompatible and a record of incompatible tracks is maintained from scan to scan. This record is updated as tracks are deleted and as new tracks are formed from the current scan's observations.

A cluster can include tracks that do not share plots directly but that both share plots with a third track. Thus, if track 1 shares a plot with track 2 and track 2 shares another plot with track 3, all three tracks are in the same cluster. A standard algorithm from [20] is used for clustering.

The result of clustering is a list of tracks that are interacting (or linked through common plots). These tracks are ranked in order of log likelihood ratio (the score function discussed in the next section). The next step is to form hypotheses of compatible tracks.

3.1.4 Hypothesis Formation and Pruning

Multiple track hypotheses are formed to represent the multiple targets in the scene. Hypotheses are defined to be sets of consistent (compatible) tracks in the sense that no two tracks within a given hypothesis share plots. There can theoretically be any number of tracks within a hypothesis. Thus, the process starts with the definition of one track hypotheses (one and only one track is valid) and expands by adding new tracks to existing hypotheses.

The new tracks that are added to any hypothesis as the hypothesis is expanded cannot share plots with any tracks in the existing hypothesis. This can be accomplished directly because each track has a compatibility list and so a compatibility list can be defined for the hypothesis as a whole. Then, when the hypothesis is expanded, only those tracks in the hypothesis compatibility list can be used.

The hypothesis generation process forms a set of N-track hypotheses (starting with $N=1$) and expands a subset of these hypotheses into $(N+1)$-track hypotheses. This process is continued until the potential scores that are associated with further expansion are no longer deemed adequate to justify expansion. This is

basically a breadth-first type of search that can be terminated at any time (if, for example, time or computer storage allocations are exceeded) with a broad cross-section of the tracks surviving.

To summarize, hypotheses are formed by a breadth-first expansion process that forms (N+1)-track hypotheses from N-track hypotheses. A continual process of pruning is performed so that the number of hypotheses is kept under control. At the end of this process, there will be a list of hypotheses, each comprising a different set of tracks. Finally, hypothesis scores are converted to probabilities, using the transformation presented in the next section, and low probability hypotheses are pruned (deleted).

3.1.5 Global-Level Track Pruning

The (a posteriori) probability of a given track can be computed as the sum of the probabilities of all the hypotheses containing that track. Some tracks, for example, may have only been contained in hypotheses that were deleted. Thus, these tracks will be computed (as an approximation) to have probability zero and can be immediately deleted. Also, each track whose probability is below a deletion threshold is removed from the track file. Finally, following [16,17], an N-scan (extended to N-observation) pruning approach is used to delete selected confirmed tracks. This implementation of N-observation pruning for a multiple sensor system is discussed in detail below.

3.1.6 Track Update and Merging

Filtered state estimates are formed for those tracks that survive pruning. This computationally demanding Kalman filtering step should not be performed until poor tracks are deleted by pruning. However, the next processing step, merging, performs a comparison of state estimates that requires accurate filtered estimates.

Tracks that potentially share observations will have been identified during clustering. Merging logic is performed to determine which tracks are redundant representations of the same target. Merging rules have been defined to use both common observation history and similar state vectors to identify those tracks that should be merged.

Once two tracks are determined to be similar, the track with the higher a posteriori probability is retained and the other track is deleted. Thus, in effect, a single track now represents the two tracks that previously represented essentially the same potential target. An increment to the score of the retained track is also made in order to account for the probability of the track that is deleted. This is described in Section 14.3 of [1].

Merging is the last logical operation performed in order to reduce the number of tracks that are to be maintained. Tracks that survive the pruning and merging steps are predicted ahead to the time of the next observation data and the process continues. Finally, as described next, the best tracks are output to the user.

3.1.7 User Presentation Logic

Any MHT system will be required to maintain at least several and possibly up to about 10 tracks for each true target. These tracks could all be displayed with some indication of their probability. This approach can be useful for the algorithm designer but is too difficult to interpret for use in an air defense system. Thus, a special logic is required to operate upon the current most likely tracks to provide a consistent, continuous output to the user. Using this method, although track numbers change internal to the MHT logic, a single track with a consistent track number is displayed for each target. The track file containing these tracks is denoted the universal track file. The logic, described in detail in [21], required to form the universal track file is outlined next.

After each scan the confirmed tracks in the most likely hypothesis of each cluster are defined to be primary tracks. These primary tracks are the tracks that best represent the expected numbers of targets in the clusters. A primary to universal track-to-track assignment process is performed to link the current best (primary) tracks with the extrapolated (universal) tracks that best represented the targets on the previous scan. Unassigned primary tracks are used to establish new universal tracks. Assigned primary tracks replace the old universal tracks to which they were assigned. Thus, this approach does not require any track fusion. Finally a logic is defined to inhibit the presentation of tracks that are identified to be stationary clutter points.

The state estimates can be obtained at any time by extrapolation of the universal tracks. A universal track is allowed to be extrapolated for several (typically 3 or 4) scans without a primary track assignment before that universal track is deleted.

3.2 Track and Hypothesis Evaluation

This section presents the basic track and hypothesis scoring relationships and shows how track confirmation and deletion thresholds can be derived as applications of the classical Sequential Probability Ratio Test (SPRT).

3.2.1 Track and Hypothesis Scoring

A track-oriented MHT approach starts by forming candidate tracks and evaluating the likelihood that each track represents a true target, as opposed to the alternative hypothesis that all plots are false alarms or random clutter. It is most convenient to convert likelihoods to the log likelihood ratio (or score function) as defined in [1,13,15]. Then, the score function for track i, L_i, can be computed at time interval k using the recursive relationship

$$L_i(k) = L_i(k-1) + \Delta L(k) \tag{18}$$

where

$$\Delta L(k) = \begin{cases} \ln\left(1-\hat{P}_D\right), & \text{no track update} \\ \ln\left[\dfrac{\hat{P}_D}{\beta_{FT}(2\pi)^{M/2}\sqrt{|S|}} - \dfrac{d^2}{2}\right], & \text{track updated with observation} \end{cases}$$

$\hat{P}_D$ = estimated probability of detection

β_{FT} = false target density

M = sensor measurement dimensionally

S = residual covariance matrix

d^2 = normalized statistical distance function

$= \underline{\tilde{y}}^T S^{-1} \underline{\tilde{y}}$

$\underline{\tilde{y}}$ = measurement residual vector

In order to use prior knowledge of new target density β_{NT} and false target density β_{FT}, the initial score can be defined to be

$$L_i(1) = \ln\left[\frac{\beta_{NT}/\beta_{NTN}}{\beta_{FT}/\beta_{FTN}}\right] \tag{19}$$

β_{NTN} = nominal (average) new target density

β_{FTN} = nominal (average) false target (noise, clutter) density

Thus, under the nominal (average) conditions the initial score is zero. If, for example, the new target density in a given region is expected to be larger than the nominal, then track confirmation is favored in that region by starting new tracks with a higher initial score.

The track score L_i can be converted to a probability of track validity $P(T_i)$ through the relationship

$$P(T_i) = \frac{\exp(L_i)}{1+\exp(L_i)}$$

This is the probability that the track is valid when compared with the single alternative that the plots are all false alarms. These probabilities can be used to prune highly unlikely tracks. Next, for tracks that survive this first pruning step, hypothesis and track probabilities are computed considering all data association hypotheses. The score of any given hypothesis H_j is just the sum of the scores of all component tracks

$$L_{H_j} = \sum_{T_i \in H_j} L_i \tag{20}$$

Implicit in the hypothesis score of Eq. (20) is the score of zero that is associated with plots that are declared to be false alarms by that hypothesis. Given hypothesis scores L_{H_j}, the probability $P(H_j)$ of hypothesis j can be computed using all J hypotheses [13],

$$P(H_j) = \frac{\exp\left(L_{H_j}\right)}{1+\sum_{i=1}^{J}\exp\left(L_{H_i}\right)}$$

Note that a given track can be contained in more than one hypothesis. Thus, the probability of a track is the sum of the probabilities of all hypotheses that contain the track. This global track probability is used for track deletion. That is, if a track's global probability does not exceed a threshold (typically 0.001 to 0.005), the track is deleted.

3.2.2 Track Confirmation and Deletion

The number of tracks that the system will be required to maintain is limited according to the computational capabilities. Also, there will typically be a false track confirmation rate requirement to be satisfied. For example, a typical criterion used in air defense system applications is that false track confirmation be limited to no more than ten per hour. Thus, track confirmation and deletion criteria must be defined in accordance with these system requirements. This is done by noting that the track confirmation process can be stated as an application of the sequential probability ratio test (SPRT).

The SPRT is used here as a decision rule for the classification problem: does a specified set of plots consist entirely of false alarms (H_0) or do they arise from a single true target (H_1)? Thus, the track confirmation (accept H_1) and deletion (accept H_0) criteria can be expressed as upper and lower thresholds (T_2 and T_1, respectively) on the score function. Using SPRT theory [22,23] these thresholds are defined in terms of the score L_i for track i such that

$$\begin{aligned} &L_i \le T_1 = \ln\left[\frac{\beta}{(1-\alpha)}\right], && \text{accept } H_0 \text{ (delete track i)} \\ &L_i \ge T_2 = \ln\left[\frac{(1-\beta)}{\alpha}\right], && \text{accept } H_1 \text{ (declare track i a confirmed target)} \\ &T_1 < L_i < T_2, && \text{continue test} \end{aligned} \tag{21}$$

Predetermined allowable false decision probabilities (α, β) are defined:

α = false track acceptance probability

β = true track rejection probability

In particular, α is defined to be the probability of accepting H_1 when H_0 is correct and can be related to the allowable false track confirmation rate. The choice of β is made in order to best minimize expected true target track confirmation time $\overline{T}_C$ while satisfying the constraints on the number of tracks that can be maintained. Results indicate a good overall choice for β to be 0.1.

In order to relate α to the false track confirmation requirements, it is first necessary to define the additional system parameters

CV = coverage volume of the sensor

T_s = scan time required by the sensor to cover volume CV

$\overline{T}_{FC}$ = expected time between false track confirmations (taken to be 3600 divided by the number of false track confirmations allowed per hour)

DBS = detection bin size

DBS is the volume of an elemental detection element in which the false alarm event is independent from other elements and has probability P_{FA}. For example, a radar detection bin might be defined by the beam width in azimuth and elevation angles, and by range and Doppler (range rate) bins.

Then, the average number of false alarms produced per second is given by

$$N_{FA} = \frac{CV}{T_s DBS} P_{FA} \tag{22}$$

For small α the probability that any one second interval will produce a false alarm that ultimately will produce a false track confirmation is αN_{FA}. Thus, in order to limit the expected number of false track confirmations to one per $\overline{T}_{FC}$ seconds, we have

$$\alpha = \frac{1}{N_{FA}\overline{T}_{FC}} = \frac{T_s DBS}{\overline{T}_{FC} CV P_{FA}} \tag{23}$$

Choice of threshold relationships as defined in Eq. (21) and with α as defined in Eq. (23) is an approximation that ensures that the false track rate will be nearly the required value. The nature of the SPRT test, with thresholds as defined by Eq. (21), is such that the actual false decision probabilities may be less than specified

[23]. As discussed further below, simulation results indicate that the false track confirmation rate resulting from this procedure may be somewhat less than that allowed by the requirement. But, the simulation assumes spatially uncorrelated clutter and, in practice, clutter is spatially correlated. Correlated clutter may lead to more false track confirmations than predicted by the simulation. However, since the specified error probabilities are small, it is expected that the differences between specified and true error probabilities will be insignificant.

4. MHT FOR MULTI-RADAR AIR DEFENSE SYSTEM APPLICATION

As previously discussed, the characteristics, such as scan rate, measurement accuracy and even measurement dimension, can differ significantly for the individual radars of an air defense system. Thus, a centralized architecture in which all plots are sent to the central-level MHT tracking system is conceptually preferable and will optimize data association and state estimation performance. Applying MHT to a centralized architecture is much less complex than the use of MHT with a distributed network such as described in [24]. However, there are a number of modifications to a general MHT algorithm that must be made in order to accommodate the unique characteristics of the air defense radar system input data and requirements.

4.1 Handling Multi-Radar Asynchronous Data

The differing radar scan rates and coverages mean that observations are received from different spatial sectors in what is effectively a random pattern. Thus, the definition of a scan (or sampling) interval that is required for track state prediction and update and for track score processing must be modified from what would typically be used for a single sensor system.

For the purposes of track prediction and update, a pseudo scan interval T_p with length about 1.0 s is defined. Tracks are predicted to the center of the interval and during the initial gating, the plots are assumed to be valid for the center of the interval. Since the plots are not actually detected at the center of the interval, there can be a timing error. After selection of candidates using a gate large enough to include time differences, time correction is performed for gate tests based on finer measures of correlation between plots and tracks.

In order to allow a track to be updated by more than one radar during an interval T_p, each radar's input plots are handled separately. Consequently, data from several radars may be processed by the MHT tracker during a single time interval T_p.

Figures 4a and 4b illustrate a potential problem that can occur when processing radar data in scan intervals of T_p. In this example, three plots (O11, O12, O13) are available from radar 1, but only O11 and O12 are received on the first scan interval (k). Similarly, radar 2 produces O21 at interval k and O22 at interval $k+1$. This pattern could likely occur for the condition of three closely spaced targets which, due to limited sensor resolution, may produce

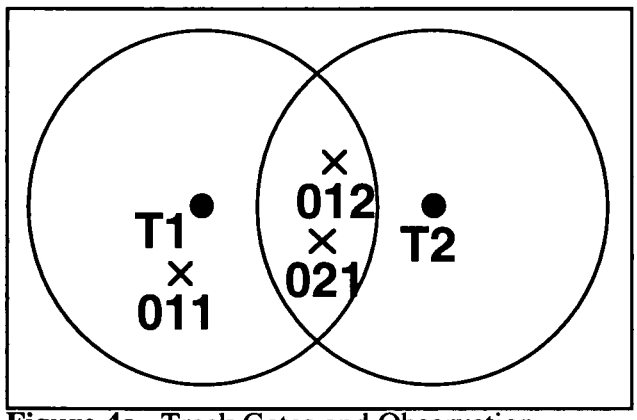


Figure 4a. Track Gates and Observation Received on Scan Interval k.

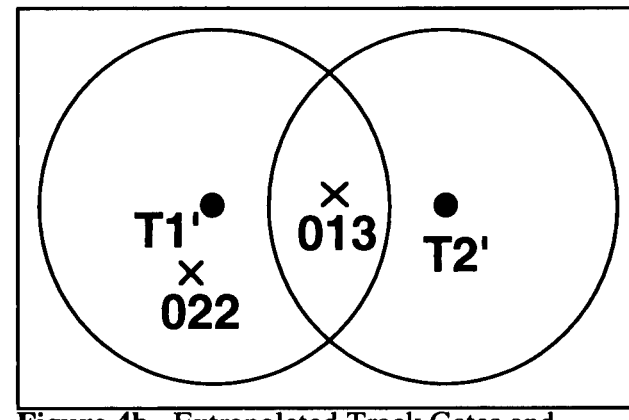


Figure 4b. Extrapolated Track Gates and Observations Received on Scan Interval k+1.

only two plots.

The processing in interval k will start with tracks T1 and T2 and the plots (O11, O12) from radar 1. Then, O21 from radar 2 will be processed. Since plots O13 and O22 will not be processed until the next interval ($k+1$), likely hypotheses on scan k will update T1 with plot O11, T2 with O12 and then either T1 or T2 with O21, as shown in Figure 4a. However, MHT will also maintain extrapolated versions (T1', T2') of T1 and T2 that are not updated with observations O12 and O21. Then, as shown in Figure 4b, the processing on interval k+1 will produce a likely hypothesis that O22 and O13 update T1' and T2', respectively and that O12 and O21 start a new track.

To address situations such as illustrated in Figures 4a and 4b, it is necessary to avoid updating a track with more than one plot from the same radar on the same scan. This can be accomplished by prohibiting a track from being updated on adjacent processing intervals by the same radar. For example, the branch of T1 that is updated on interval k with O11 is not allowed to be updated on interval $k+1$ with O13.

Finally, the manner in which the miss penalty $\ln\left(1-\hat{P}_D\right)$ is added to the track score defined in Eq. (16) must be modified from that used for a single sensor system. One approach is to define a basic scan interval T_b such that any target within the system field of view should be illuminated at least once (by one of the radars) during this time period. This time period is typically about 6-12 seconds. Thus, a track score miss penalty is given to those tracks that are not updated at least once in this time interval. A second, more accurate, but more complex approach is to apply the miss penalty separately to each track according to the expected radar illumination pattern of that track and its previous update history.

4.2 N-Scan (Observation) Pruning

Following [16,17], the basic idea of N-observation pruning is to use the tracks in the current most likely hypothesis to prune other tracks based upon conflict over observations received N observations back in time.

The N-observation pruning logic is facilitated by use of the family tree structure illustrated in Figure 5. Using this structure, root nodes are

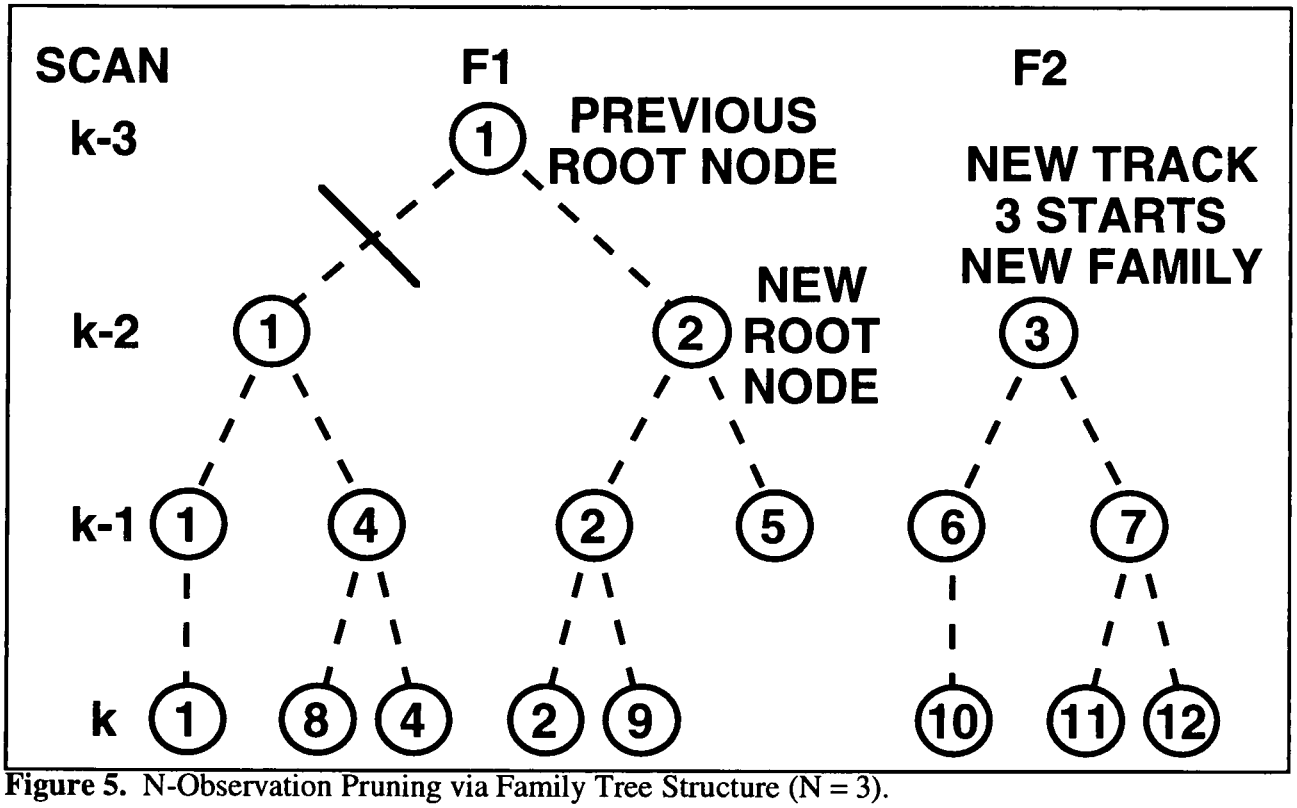


Figure 5. N-Observation Pruning via Family Tree Structure (N = 3).

identified and all confirmed tracks that are not descendants of one of these root nodes are deleted. The root nodes are identified using the confirmed tracks in the most likely hypotheses of each cluster. The history of each of these tracks is traced back for N observations and the ancestor track that was updated with the Nth observation in the past becomes the family tree node. Then, all confirmed tracks that are not descendants of one of the root node tracks are deleted.

The Nth-observation pruning logic is illustrated in Figure 5 with N=3. Track T2 of family F1 is contained in the most likely hypothesis at scan k. Assume that it was updated on scans k, k-1 and k-2. Using N=3, the history of track T2 is traced back for 3 observations to the root node represented by T2 on scan k-2. Then, within F1, only those tracks that can be traced back to T2 in scan k-2 (in this case only T2 and T9) are allowed to survive. All other tracks are deleted. This logic is, in effect, using current data to make irrevocable decisions regarding the correct association hypothesis N observations ago. Finally, each family that does not have a track in the most likely hypothesis and contains at least one track with N observations is identified (such as F2). Each confirmed track within such a family is deleted.

As discussed in [17], the single sensor N-scan pruning logic has been extended for a multi sensor system where, potentially, all targets may not be seen by all sensors. The logic is restricted to all tracks that contain at least N observations and the number of scans back is the number required to cover N observations. Thus, the logic is actually an N-observation pruning rule. This logic reduces the number of tracks required for maintenance in typical scenarios by about half when compared with a system that only uses hypothesis and track probabilities for pruning.

4.3 False Target Density and Track Confirmation

Computation of the track score function, as discussed above, requires the definition of a false target (or false alarm) density. Also, determination of the track confirmation threshold requires an estimate of the false target input rate. This section defines these quantities for a typical multi-radar air defense system application.

The majority of false detections produced by radar air defense systems are clutter returns. Many clutter returns will persist over extended time periods and thus will have a similarity with true targets. Tracks formed from persistent stationary clutter can be identified as false targets by their small velocity. These tracks should be maintained so that future clutter returns can be correctly identified and not assigned to true target tracks. Then, only those false observations that are essentially uncorrelated from scan-to-scan will remain to present a problem to the tracking system. Characterization of these false returns is discussed next.

Typical clutter density is defined to be one false observation (or plot) per unit area A_i (on the ground) per scan of a given radar i. A representative value for A_i is 25 nmi^2. Assume that the returns are uniformly distributed over an altitude extent h_e, that radar i has scan time T_{s_i} and that data over interval T_p are processed. Then, the false target density for use in processing this data and updating the track score is

$$\beta_{FT_i} = \frac{T_p}{T_{s_i} A_i h_e} \tag{24}$$

false plots per time interval T_p per unit volume

Using typical parameters

$A_i = 25$ nmi^2, $h_e = 15{,}000$ ft, $T_{s_i} = 6$s, $T_p = 1$s

gives

$$\beta_{FT_i} = 1.2 \times 10^{-14}\ \text{ft}^{-3} = 4.2 \times 10^{-13}\ \text{m}^{-3}$$

Determination of the track confirmation threshold for an air defense system basically follows from Eqs. (21) and (23). However, for this application, as derived next, the input number of false alarms N_{FA} is computed from the clutter model rather than from Eq. (22). In order to determine N_{FA}, assume that the air defense system will cover a volume defined by maximum R_{max} and minimum R_{min} ranges.

Also, the angular coverage is taken to be $\pm\eta_{lim}$ in azimuth and 0 to ε_{lim} in elevation. Then, the total coverage volume becomes

$$CV = \frac{2\left(R_{max}^3 - R_{max}^3\right)}{3}\eta_{lim}\sin\varepsilon_{lim}$$

Given N_R radars the total or (system) false target density will be a sum over β_{FT_i} defined in Eq. (24). Then, the number of false alarms per second N_{FA} becomes

$$N_{FA} = \frac{1}{T_p}\sum_{i=1}^{N_R} CV_i\beta_{FT_i} = \frac{1}{h_e}\sum_{i=1}^{N_R}\frac{CV_i}{T_{s_i}A_i}$$

As a typical example, consider a system with 4 radars each with a scan period of 6 s and with area A_i equal to 50 nmi^2. Further take

h_e = 2.5 nmi, R_{max} = 65 nmi, R_{min} = 2 nmi

η_{lim} = 180 deg, ε_{lim} = 10 deg.

Then, the expected number of false alarms N_{FA} per second is about 500.

The above development has assumed uniformly distributed false alarms while in practice the density will vary. In particular, the false alarm density will typically be larger at closer range. In order to extend the development to non-uniform clutter densities, a set of zones with different clutter densities is defined for each radar. Similarly, the expected new true target density can be chosen to vary with zone. Then, the initial track score can be defined according to the zone using Eq. (19).

5. SYSTEM EVALUATION

Three levels of analysis and simulation are used to predict true and false track confirmation performance and true target track maintenance performance. First, as discussed in the previous section, the track confirmation process is a direct application of the classical SPRT. Thus, analytic prediction of the expected decision times for true track confirmation and false track deletion can be obtained directly from the Wald approximation [13,15,22,23].

A second performance prediction approach uses a simplified Monte Carlo simulation to follow the score history of simulated tracks. False track confirmation performance is evaluated by examining the histories of tracks that are initiated by false alarms. The simulation is simplified by only examining the highest score false track hypothesis. Each Monte Carlo simulation run is continued until the track score satisfies either the confirmation or deletion criterion, as defined in Eq. (21).

The simplified Monte Carlo simulation approach for predicting true target track confirmation and maintenance history follows the true target path and generates observations according to the assumed system detection and measurement statistics. This simulation just propagates the correct hypothesis (no false alarms are generated) and compares the score of this hypothesis with the confirmation and deletion thresholds. This approach uses the optimistic assumption that the correct hypothesis will ultimately prevail over incorrect hypotheses that contain false alarms. Thus, its results represent an upper bound on performance. However, experience indicates that performance predictions from more detailed (and much more time consuming) simulations closely follow those from the simplified MHT simulation [15].

Finally, a detailed Monte Carlo simulation of target detection and MHT processing is required in order to consider scenarios with multiple interacting targets and to obtain final system performance predictions. This simulation is also used for basic system design. Thus, along with detailed Monte Carlo statistics, a detailed print out capability has been developed so that track and hypothesis time histories can be examined for individual Monte Carlo runs. The next sections will present representative results from system design studies.

5.1 False Track Confirmation

Due to the extremely large numbers of Monte Carlo samples required in order to obtain accurate statistics, false track confirmation performance is predicted using the simplified Monte Carlo simulation approach. For example, consider a hypothetical system with 500 false alarms/s input and with a required time $\overline{T}_{FC}$ between false track confirmations.

Table 3. Simulated False Track Confirmation Results

Parameters			Simulation Results	
$\overline{T}_{FC}$	False Tracks per Hour	α	Number of False Tracks	Estimated α
36	100	5.6×10^{-5}	63	3.50×10^{-5}
120	30	1.7×10^{-5}	15	0.83×10^{-5}
360	10	5.6×10^{-6}	10	5.60×10^{-6}
720	5	2.8×10^{-6}	4	2.20×10^{-6}

Then, if 10 false track confirmations/hr are allowed

$\overline{T}_{FC} = 360$ s, $\alpha = 5.6\text{x}10^{-6}$

Table 3 presents false track confirmation results as a function of the various $\overline{T}_{FC}$ (and resulting α) that were simulated. An hour of tracking was simulated for each case. This table gives the expected number of false tracks and the number of false tracks actually formed by the simulation (with equivalent α). These simulation results generally agree with the Wald approximation but show, as in [15], that the numbers of tracks formed by simulation tend to be somewhat less than predicted.

5.2 Single Target Track Confirmation and Maintenance

True target track confirmation performance has been examined as a function of false alarm density and system coverage. First, the analytic Wald approximation is used to predict expected track confirmation time $\overline{T}_C$. Second, the simplified Monte Carlo simulation approach is used to compute $\overline{T}_C$ and the expected time T_{90} at which 90 percent of the target tracks will be confirmed. Finally, the full Monte Carlo simulation, with a multi-radar detection model was used to compute $\overline{T}_C$ and T_{90}.

The multi-radar system simulated used 3 radars with coverage volume out to 120 nmi and azimuth and elevation limits of ± 60 deg and 15 deg, respectively. Two radars had sampling interval 3.75 s while the third had sampling interval 10 s. The measurement error standard deviations were taken to be 0.5 deg in angle and 0.07 nmi in range. System evaluation was performed for different values of the clutter density parameters A_i, but for a given system the same value A_c was used for all radars.

In order to use analytic and simplified Monte Carlo simulation methods for evaluation, the multi-radar system was approximated by a single radar system with an equivalent single radar sampling interval T_{SR} given by

$$\frac{1}{T_{SR}} = \frac{1}{3.75} + \frac{1}{3.75} + \frac{1}{10}, \quad T_{SR} = 1.58 \text{ s}$$

Also, in the case where the clutter area parameters A_i are the same for all radars, the effective clutter area for the equivalent single radar system is also equal to A_c.

Table 4 summarizes true target track confirmation performance using the analytic prediction and the results of simplified and full Monte Carlo simulation. Results are presented as a function of A_c, which is defined to be the area over which one false alarm per radar per scan is expected. The simplified Monte Carlo results were derived using 500 samples while the full Monte Carlo used 20 samples for the $A_c = 100$ case, 100 sample for the $A_c = 50$ and 25 cases and 10 samples for the case of $A_c = 12.5$.

Table 4 shows a consistency between the results obtained using the three performance prediction approaches. The track confirmation times predicted by the analytic method are somewhat longer than those found from the Monte Carlo simulation. This can be explained by the fact

Table 4. Track Confirmation as Function of Clutter Density

Clutter Density Parameter	Track Confirmation Times				
	Analytic	Simplified Monte Carlo		Full Monte Carlo	
A_c	$\overline{T}_C$	$\overline{T}_C$	T_{90}	$\overline{T}_C$	T_{90}
100	12	10	14	13	16
50	15	12	17	14	16
25	20	16	21	18	24
12	35	22	32	27	30

that the analytic prediction is, in effect, averaging over an ensemble of target maneuvers while the Monte Carlo simulations used nonmaneuvering targets.

Track maintenance in the presence of severe target maneuver as well as heavy clutter has also been studied. Typical results are summarized in Figures 6 through 9. First, Figures 6 through 8 show the probability that a track is confirmed and maintained as a function of time for three clutter densities (A_c = 50, 25, 12.5) examined using a full Monte Carlo study. These results include a 7 g target maneuver that began at 37 s and led to a 180 deg turn. To satisfy the maintenance requirement, a track must be continuously maintained on the target. The

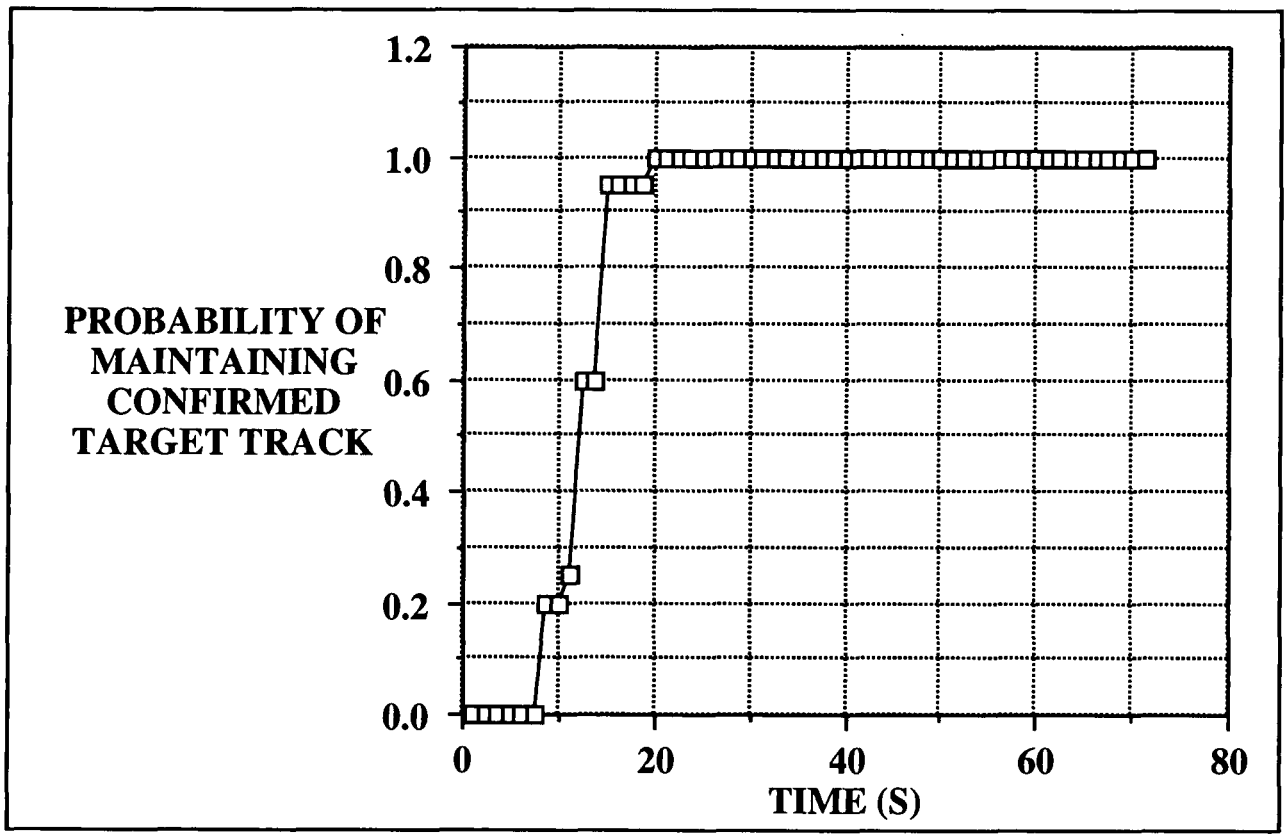


Figure 6. Track Confirmation and Maintenance for A_c = 50 nmi^2.

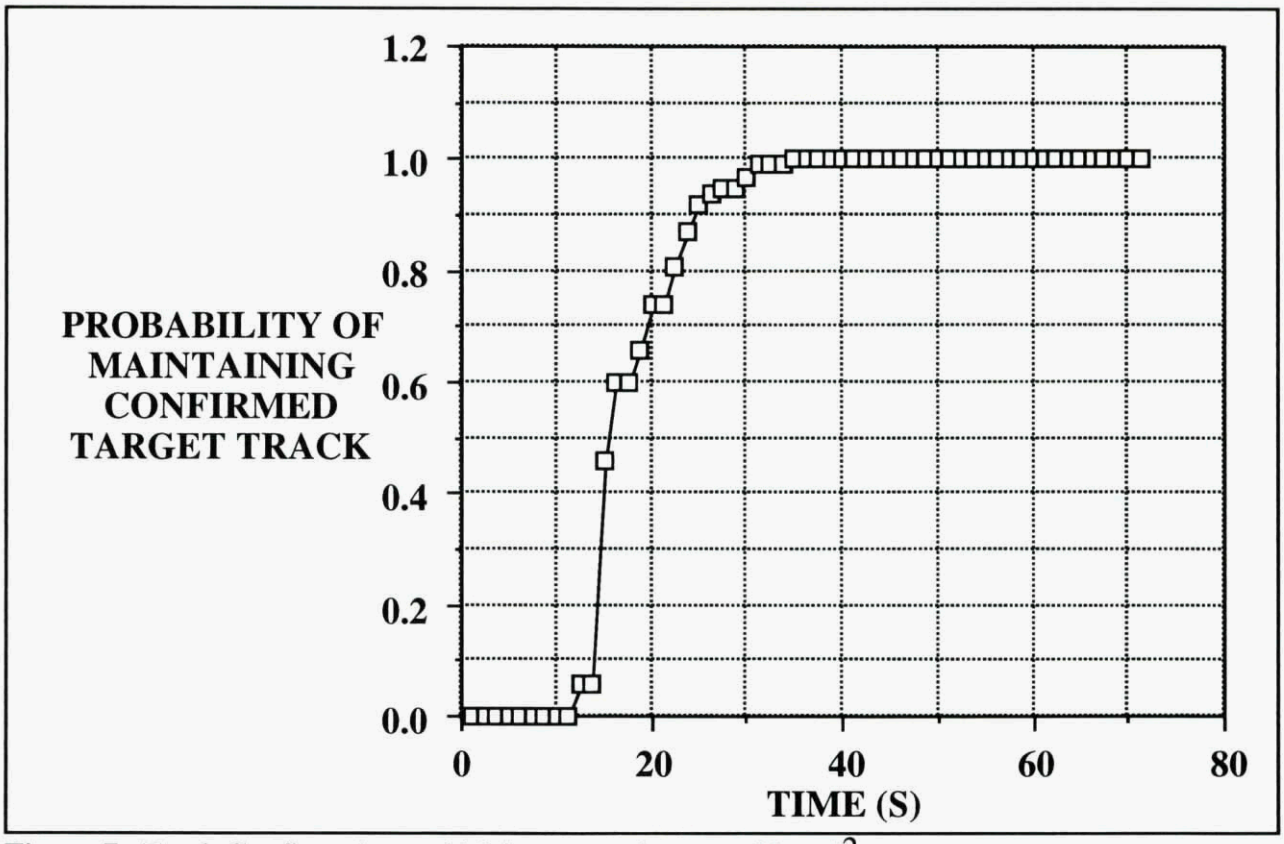


Figure 7. Track Confirmation and Maintenance for $A_c = 25$ nmi^2.

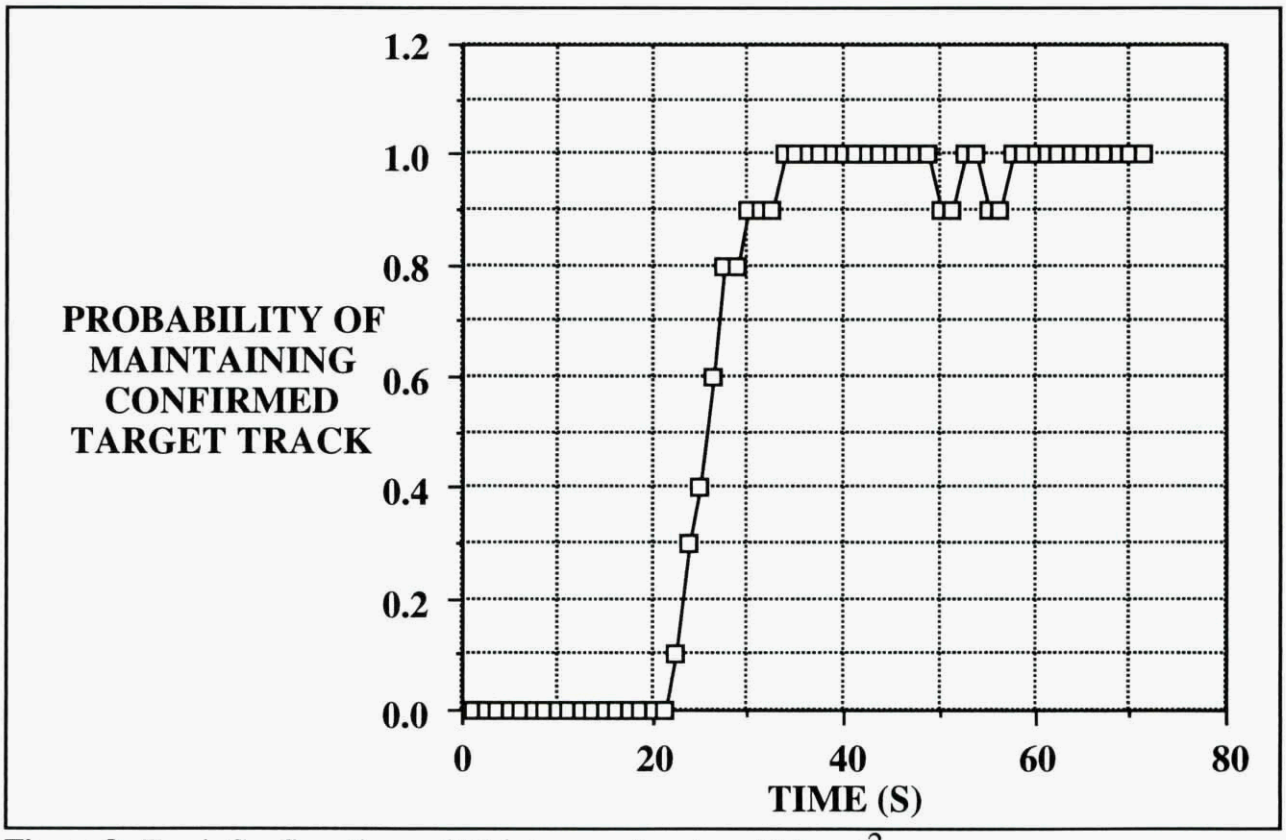


Figure 8. Track Confirmation and Maintenance for $A_c = 12.5$ nmi^2.

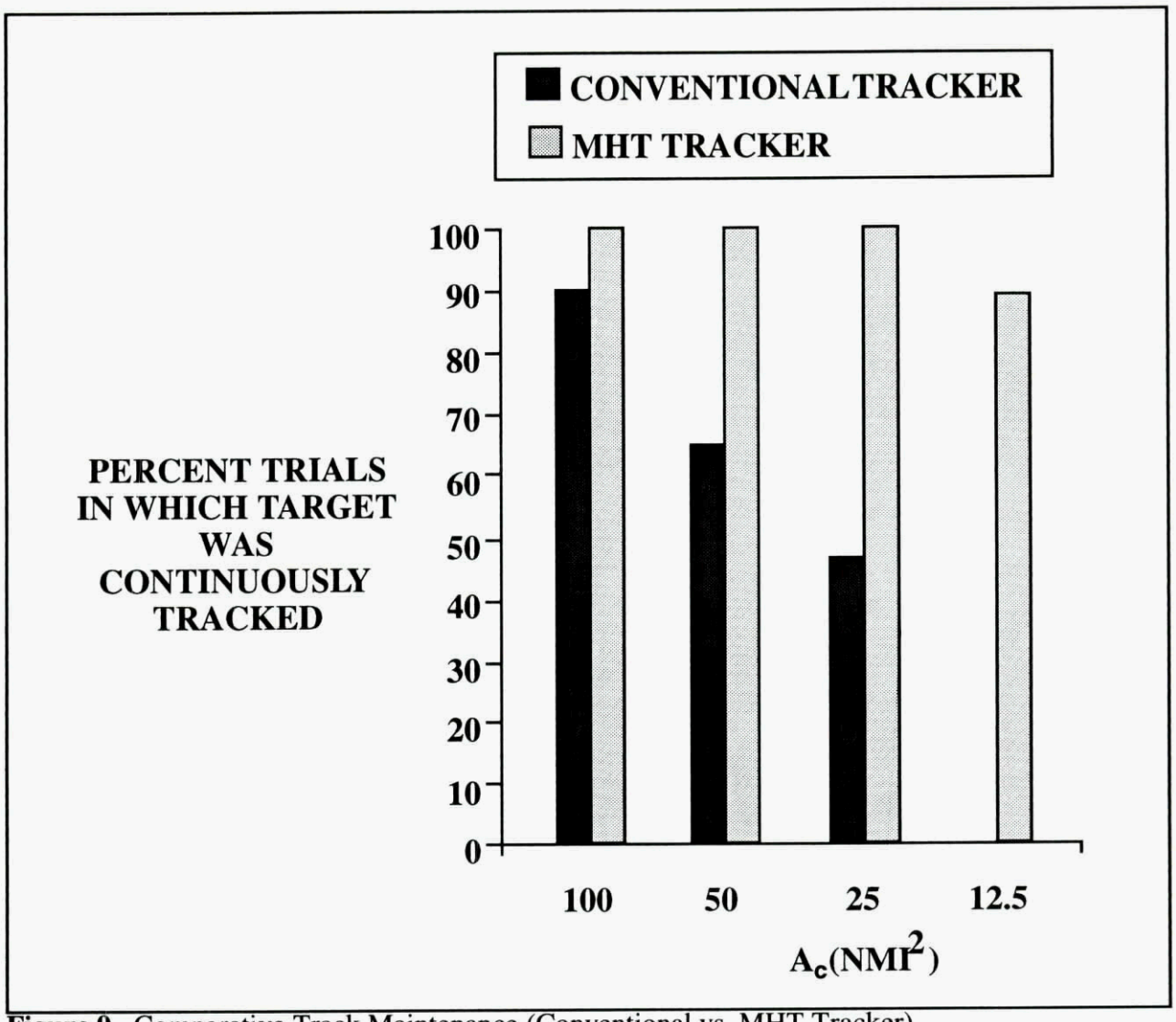


Figure 9. Comparative Track Maintenance (Conventional vs. MHT Tracker).

maintenance requirement also includes a difference criterion between the true and estimated position and velocity. This distance was briefly exceeded for two of the ten Monte Carlo runs under the most dense clutter condition (A_c = 12.5). However, a continuous track was maintained in both cases and, as shown in Figure 8, the distance criterion was only exceeded for about 2.5 s in both cases.

Figure 9 shows comparative track maintenance performance results for MHT versus a conventional (single hypothesis tracking) system in current operational use. The probability of track maintenance through a 7g 180 deg turn is given as a function of the clutter density, as defined by A_c. These results, as well as similar results obtained from IR system analysis, indicate that the MHT approach has the capability to operate in a clutter density that is at least 10 times the operational limit for a conventional (single hypothesis) tracking system.

6. SUMMARY

MHT is readily applicable to a multi-radar air defense system. Plots from all radars are transmitted to the central-level MHT processor. Although not discussed in detail in this chapter, modern computational capabilities and recently developed MHT algorithm efficiencies assure that real-time processing can be achieved for systems currently under development.

Simulation results indicate that the MHT approach to data association will lead to a significant performance improvement over that currently achievable by conventional (single hypothesis) tracking systems.

7. ACKNOWLEDGMENTS

The authors would like to thank Robert Wenski of Hughes Aircraft Company and Dr. Joseph Mingrone of Hughes Aircraft of Canada, Ltd. for their consultation on coordinate conversion and their expert review of this work.

8. REFERENCES

1. Blackman, S.S., *Multiple Target Tracking with Radar Applications*, Artech House, Norwood, MA 1986, Chapters 9,10 and 14.
2. Busch, M. T. and S. S. Blackman, "Evaluation of IMM Filtering for an Air Defense System Application", *Proc. of Signal and Data Processing of Small Targets 1995*, San Diego, CA, July 1995.
3. F. Pearson II, *Map Projections: Theory and Applications,* CRC Press Inc., Boca Raton, Florida, 1990
4. J. P. Snyder, *Map Projections - A Working Manual,* U.S. Geological Survey Professional Paper 1395, United States Government Printing Office, Washington, DC 1987.
5. P. Thomas , "Conformal Projections in Geodesy and Cartography", Special Publication No. 251, U.S. Department of Commerce, Coast and Geodetic Survey, U.S. Government Printing Office, Washington DC 1952.
6. O. Adams, "General Theory of Polyconic Projections", Special Publication No. 57, U.S. Department of Commerce, Coast and Geodetic Survey, U.S. Government Printing Office, Washington DC 1934.
7. E. Wolf, "A Stereographic Coordinate System for the Utilization of Data from Several Radars", The MITRE Corporation, SR-2, March 1959.
8. C.H. Nordstrom, "Stereographic Projection in the Joint Surveillance System", The MITRE Corporation, ESD-TR-76-169, September 1976.
9. D. Goldenburg and E. Wolf "A Common Coordinate System for the Utilization of Data from Several Radars", Lincoln Laboratory Technical Report TR 67, 1954.
10. J.J. Burke, "Stereographic Projection of Radar Data in a Netted Radar System", The MITRE Corporation, ESD-TR-73-210, AD 771544, November 1973.
11. Dana, M. P., "Registration: A Prerequisite for Multiple Sensor Tracking", *Multitarget-Multisensor Tracking: Advanced Applications,* in Y. Bar-Shalom (ed), Artech House, Norwood, MA, 1990, Chapt. 5.
12. Broida, T.J., et al, "Performance Prediction and Structured Branching for Multiple Sensor Tracking Systems", *Proc. of 1989 Tri-Service Data Fusion Symposium*, May 1989.
13. Demos, G.C., et al, "Applications of MHT to Dim Moving Targets", *Proc. of Signal and Data Processing of Small Targets 1990,* SPIE Vol. 1305, April 1990, pp. 297-309.
14. Werthman, J.R., "Step-by-Step Description of a Computationally Efficient Version of Multiple Hypothesis Tracking", *Proc. of Signal and Data Processing of Small Targets 1992*, SPIE Vol. 1698, April 1992, pp. 288-300.
15. Blackman, S.S., R.J. Dempster and T. J. Broida, "Multiple Hypothesis Track Confirmation for Infrared Surveillance Systems," *IEEE Trans. on Aerospace and Electronic Systems,* Vol. 29, No. 3, July 1993, pp. 810-823
16. Kurien, T., "Issues in the Design of Practical Multitarget Tracking Algorithms," in Y. Bar-Shalom (ed), *Multitarget-Multisensor Tracking: Advanced Applications,* Artech House, Norwood, MA, 1990, Chapter 3.

17. Kurien, T. and M. E. Liggins, "Report-to-Target Assignment in Multisensor-Multitarget Tracking", *Proc. of 27th Conference on Decision and Control,* Austin, Texas, Dec 1988, pp. 2484-2488.

18. Kovacich, M. et al, "An Application of MHT to Group to Object Tracking", *Proc. of Signal and Data Processing of Small Targets 1991,* SPIE Vol. 1481, Orlando, FLA, April 1991, pp. 357-370.

19. Blostein, S.D. and H.S. Richardson, "A Sequential Detection Approach to Target Tracking", *IEEE Trans. on Aerospace and Electronic Systems,* Vol. 30, No. 1, January 1994, pp. 197-211.

20. Baase, S., Computer Algorithms: *Introduction to Design and Analysis,* Addison-Wesley, Reading, MA 1978.

21. Blackman, S.S., R.J. Dempster and J.T. Fagarason, "Continuous Time Representation of Multiple Hypothesis Track Data," *Proc. of Signal and Data Processing of Small Targets 1993*, SPIE Vol. 1954, Orlando, FLA, April 1993, pp. 352-360.

22. Wald, A., *Sequential Analysis*, J. Wiley and Sons, New York, 1947, Chapter 3.

23. Ferguson, T.S., *Mathematical Statistics*, Academic Press, New York, 1967, Section 7.6.

24. Chong, C.Y., S. Mori and K.C. Chang, "Distributed Multitarget Multisensor Tracking", in Y. Bar-Shalom (ed), *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, MA., 1990, Chapt. 8.

25. Daum, F. E. and R. J. Fitzgerald, "The Importance of Resolution in Multiple Target Tracking:, *Proc. of Signal and Data Processing of Small Targets 1994*, SPIE Vol. 2235, Orlando, FL, April 1994, pp. 329-338.

26. Trunk, G. V and J. D. Wilson, "Track Initiation of Occasionally Unresolved Radar Targets", *IEEE Trans. on Aerospace and Electronic Systems*, Vol. AES-17, No. 1, January 1981, pp. 122-130.

27. Chang, K. C. and Y. Bar-Shalom, "Joint Probabilistic Data Association for Multitarget Tracking with Possibly Unresolved Measurements and Maneuvers", *IEEE Trans. on Automatic Control*, Vol AC-29, No. 7, July 1984, pp. 585-594.

Multisensor Tracking and Fusion with MTI Radars

Kuo-Chu Chang
Systems Engineering Department
George Mason University
Fairfax, VA 22030, USA
E-mail:kchang@gmu.edu

Yaakov Bar-Shalom*
Electrical and Systems Engineering Department
University of Connecticut
Storrs, CT 06269, USA
E-mail:ybs@eng2.uconn.edu

1. SUMMARY

Multisensor tracking and data fusion deals with combining data from various sources to arrive at an accurate assessment of the situation. Technical difficulties in performing multisensor tracking and fusion include not only ambiguous data, but also disparate data sources. The tracking and fusion problem is further complicated by the facts that targets may not be detected by some sensors, dense false alarms and clutter detections may be present, and the target model may not be known exactly. In this chapter, a multitarget tracking problem which involves data obtained from multiple MTI radars is considered. A tracking and fusion algorithm which takes into account the uncertainties in both data origin and target dynamics under a dense clutter environment is presented.

2. INTRODUCTION

Tracking and fusion with multiple sensors has attracted a great deal of attention recently [1-6]. It deals with integration and correlation of data from various sources to arrive at an overall assessment of the situation. The difficulties in performing multisensor tracking and fusion include not only ambiguous data, but also disparate data sources. First of all, the identity of the objects responsible for each individual data set is unknown so there is uncertainty as how to associate data from one sensor which are obtained at one time and location to those of another sensor at another point in time and location. Secondly, the data sources may include various active and passive sensors such as radar, infrared, acoustic, as well as other sources such as COMINT and ELINT.

The tracking and fusion problem is further complicated by the facts that (1) targets may not be detected by some sensors due to the variation of signals and the sensor characteristics, (2) dense false alarms and clutter detections which are not easily distinguishable from the true target measurements may be present; and (3) target models may not be known exactly and they may vary in time.

The *multiple-hypothesis* approaches for multitarget tracking which are near-optimal, have gained popularity [1-6] since the pioneer work of Reid [7]. In these approaches, all feasible data association hypotheses between measurements and targets are formed, evaluated, and maintained. Although they can handle complex target and sensor models and include track initiation and continuation in one framework, they require computing resources, both time and memory, that increase exponentially with time especially under dense target and clutter environments. Thus only suboptimal implementations are feasible, which are still very expensive.

Other approaches for multitarget tracking include *target-oriented* and *track-oriented* approaches. In the *target-oriented* approach [6, Sec. 6.2], the number of targets is assumed to be known and all data association hypotheses are combined into one for each target. Even though the computational requirements are fixed for this type of approach, it can only handle track continuation, and requires a separate module for track initiation. In the *track-oriented* approach, tracks are initiated, updated, and possibly terminated based on the associated measurement history. Each track is treated individually and a data association and evaluation scheme is needed to form and evaluate the track. This approach represents a trade-off between performance and computation and is most suitable for dense target or clutter applications.

In this chapter, a multitarget tracking problem under a dense clutter environment is considered. Detections from ground vehicles are assumed to be available from multiple airborne MTI (moving target indicator, i.e., with Doppler) radars located on different platforms. The goal of this chapter is to present a simple, yet comprehensive multiple-target tracking and fusion algorithm suitable for dense target and clutter environment. It is shown in [5, Ch. 7] that the optimal data association algorithm is only marginally better when the target or clutter density is either very high or very low compared to the simple nearest-neighbor algorithm.

We therefore propose here a simple *track-oriented* approach based on a form of "greedy" nearest-neighbor and multiple model algorithms. In this approach, a centralized fusion architecture is assumed, i.e., data collected from multiple sensors are pooled together in a central site where they are combined. The goal here is to be able to pick up a potential track as quickly as possible and eliminate false tracks as effectively as possible. To do so, tracks are initiated based on a single measurement and a score is obtained for each track to determine its "strength". The track score is calculated based on the associated measurement history as well as the target and sensor models. To eliminate false tracks effectively, tracks with scores below a specified threshold will be pruned. The pruning threshold is one of the system parameters and can be adjusted adaptively based on the scenario and performance requirements.

The algorithm has been implemented in a MATLAB environment [8]. Simulation results with multiple MTI sensors have also been obtained. Extensive analysis with Monte Carlo simulations shows good performance. The chapter is organized

---

* Sponsored by AFOSR Grant F49620-95-1-0229 and ONR/BMDO N00014-91-J-1950

as follows. Section 3 presents the tracking algorithm, Section 4 describes sensor and target models, and Section 5 shows the simulation results.

3. TRACKING AND FUSION ALGORITHM

The centralized architecture is chosen for tracking and fusion with multiple sensors. In other words, all detections from different sensors are pooled together in the fusion center where they are processed. To handle a dense target and clutter environment, a track-oriented approach is proposed together with the multiple-model approach for tracking maneuvering targets.

The tracking and fusion algorithm is summarized as follows (see Figure 1).

1. Track clustering: At each scan when a new data set (measurements) arrives, first decompose the existing tracks and new measurements into independent groups (clusters) based on gating so that data association can be performed within each cluster. This step is to eliminate unnecessary associations between far apart (in the measurement space) tracks and measurements.

2. Data association: Within each cluster, form and select the most likely data association event to be processed. To speed up the process in dense environments, a "greedy" algorithm is used to associate measurements to tracks.

3. Track scoring/updating: A score is assigned to each track and is obtained based on the association history. The score indicates the "true track probability" (TTP) and is used in the decision of eliminating or confirming tracks.

4. Track initiation: Measurements not associated with any existing track will be used to initiate a new track. A new track is initiated with a single measurement where a Gaussian distribution for the target state is created based on that measurement.

5. Track management: To avoid redundant tracks, similar tracks are combined and tracks with scores below a certain threshold are pruned out.

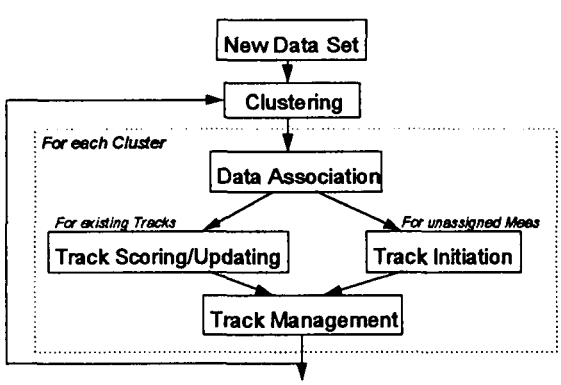


Figure 1. Algorithm Summary

Data Association

The simplest data association algorithm is the so called "nearest neighbor" (NN) algorithm [3]. There are several variations of the algorithm including the standard NN and the "greedy" NN algorithms. In the standard NN algorithm, one simply assign tracks to measurements according to the order of the association likelihoods. Namely, the "nearest neighbor" assignment is

$$(i^*, \tau^*) = \arg\max_{i,\tau} L(y_i|\tau) \tag{1}$$

The selected track and measurement pair is then removed from the list for future consideration and the next most likely association will be chosen from the remaining list. This procedure is repeated until all tracks are considered.

In the "greedy" nearest neighbor (GNN) algorithm, the tracks are first prioritized. The idea is to start with the track with the highest "priority" and assign it with the most likely measurement. Namely, the first pair is

$$\begin{aligned} \tau^* &= \arg\min_{\tau} H(\tau) \\ i^* &= \arg\max_{i} L(y_i|\tau^*) \end{aligned} \tag{2}$$

where $H(\tau)$ is the track priority, which could be defined as a function of the track score (probability), relative track estimated position, current track association uncertainty, or a combination of part or all of the above. For example, the association uncertainty of a track is defined as the entropy of the association probabilities between the track τ and all the current "feasible" measurements y_i, i.e.,

$$H(\tau) = -\sum_i p_i \log(p_i) \tag{3}$$

where

$$p_i = \frac{L(y_i|\tau)}{\sum_i L(y_i|\tau)} \tag{4}$$

and $L(y_i|\tau)$ is the association likelihood between measurement y_i and track τ (for specific example, see Eq. (18)).

Once a certain track-to-measurement association pair is chosen, they will be removed from the list and the next track with the highest priority will be processed. This process continues until all tracks are considered. Note that measurements which have been chosen earlier will not be included in the current consideration. It is possible that a track may not have any validated[1] measurement, in that case, no measurement will be chosen and the track will not be updated. After processing all tracks, measurements that have not been assigned to any track will then be used to initiate a new track.

[1]measurements inside the validation gate [3].

Track Initiation

The track initiation module initiates a track based on a single measurement. To initiate a track, we need a Gaussian representation (approximation) of the target state distribution of an undetected target. It is reasonable to assume that the position information contained in this distribution is negligible compared with that contained in the measurements because variances of measurements are typically much smaller than that of the prior distribution. Therefore, a track will be initiated from a single measurement as follows:

1. First create a gaussian distribution from the three positional measurements, i.e., the range, the azimuth and the elevation.

2. Create a gaussian distribution of the velocity vector with a priori information.

3. Modify the Gaussian target state distribution by the Doppler (range rate) measurement.

Track Scoring/Updating

A score is assigned to each track and is updated based on the association history. A multiple model algorithm [6, Sec. 4.4] is used to define and evaluate track scores. In this algorithm, two models are used: one for "observable target", designated as Model A ("alive"), one for "unobservable target", designated as Model D ("dead"). The "unobservable target" can represent either a true target outside the sensor coverage or an erroneously hypothesized target, that is, it is equivalent to "no target". In both models, measurements can originate from the target (with detection probability P_D) or clutter. However, in the "no target" model we have $P_D = 0$.

A Markov chain will model the observable and unobservable situation as follows. Denoting by M_X the event that model X is in effect during the current sampling interval[2] and $\overline{M}_X$ for the previous interval, the following transition probabilities are assumed:

$$\begin{aligned} P(M_A|\overline{M}_A) &= 1-\varepsilon_A, \quad P(M_D|\overline{M}_A) = \varepsilon_A \\ P(M_D|\overline{M}_D) &= 1-\varepsilon_D, \quad P(M_A|\overline{M}_D) = \varepsilon_D \end{aligned} \tag{5}$$

That is, transitions between the models (sudden death probability ε_A and resurrection probability ε_D) are assumed with low probabilities. Practical values for these "designed parameters" are discussed in [6, Sec. 4.4].

The initial score for the new track is calculated based on the ratio of new target density β_{NT} and clutter density β_λ, specifically, $P_0(M_A) = \dfrac{\beta_{NT}}{\beta_{NT}+\beta_\lambda}$, where β_{NT} and β_λ are defined as the expected numbers of true targets and false alarm per unit surveillance volume per scan. Denoting by $P(M_X)$ and $P(\overline{M}_X)$ the probabilities that track τ and its ancestor $\overline{\tau}$ are in model X respectively, then at each scan, depending on the specific association for the track, the score is updated based on the following formula.

$$P_y(M_A) = \frac{1}{C_y}\left[P(M_A|\overline{M}_A)P(\overline{M}_A) + P(M_A|\overline{M}_D)P(\overline{M}_D)\right] \cdot P_D(\overline{\tau})L(y|\overline{\tau}) \tag{6}$$

$$P_y(M_D) = \frac{1}{C_y}\left[P(M_D|\overline{M}_A)P(\overline{M}_A) + P(M_D|\overline{M}_D)P(\overline{M}_D)\right]L_{FA}(y) \tag{7}$$

for track $\overline{\tau}$ with associated measurement y where $C_y = P_y(M_A) + P_y(M_D)$, and

$$P_\varnothing(M_A) = \frac{1}{C_\varnothing}\left[P(M_A|\overline{M}_A)P(\overline{M}_A) + P(M_A|\overline{M}_D)P(\overline{M}_D)\right] \cdot \left(1-P_D(\overline{\tau})\right) \tag{8}$$

$$P_\varnothing(M_D) = \frac{1}{C_\varnothing}\left[P(M_D|\overline{M}_A)P(\overline{M}_A) + P(M_D|\overline{M}_D)P(\overline{M}_D)\right] \tag{9}$$

for track $\overline{\tau}$ with no associated measurement where $C_\varnothing = P_\varnothing(M_A) + P_\varnothing(M_D)$. In (6)-(9), $P_D(\overline{\tau})$ is the detection probability of the predicted track $\overline{\tau}$, $L(y|\overline{\tau})$ is the association likelihood between the measurement y and $\overline{\tau}$, and $L_{FA}(y)$ is the false alarm likelihood.

For the tracks with associated measurement, they will be updated based on the standard filtering equations such as Kalman filter or Extended Kalman Filter (EKF) (see Section 3 for a specific example). For those track with no associated measurement, they will only be extrapolated without updating.

Track Management

Since many tracks can be initiated under a dense clutter environment, to avoid redundant tracks, a test of similarity is carried out to determine whether two tracks represent the same target. This test first checks if the two tracks share the same current associated measurement, if so, a track-to-track association technique [6] is then used to determine their similarity. Assuming the tracks are independent, this test will accept the hypothesis that the two tracks represent the same target if $(\hat{x}_i - \hat{x}_j)'(\hat{V}_i + \hat{V}_j)^{-1}(\hat{x}_i - \hat{x}_j) \le \psi$ where $\hat{x}_i, \hat{x}_j$ and $\hat{V}_i, \hat{V}_j$ are the state means and covariances of the two tracks and ψ is the decision threshold.

To judge if a track is real or not at any moment of time, the scoring threshold is used. A track is judged to be false and will be pruned if the score of the track is below the given threshold. This is to effectively eliminate unwanted tracks under dense

[2]The k-th sampling interval is $(t_{k-1}, t_k]$.

clutter environment. The pruning threshold is one of the system parameters and should be adjusted adaptively based on the scenario and performance requirements.

4. SIMULATION MODELS

We first describe the target and sensor models used for the simulation. Targets of interest are generally assumed to follow nearly constant velocity trajectories. To model the effects of randomness in target dynamics (target maneuvering), white accelerations are assumed. The target motion is modeled in two-dimensional Cartesian coordinates.

The MTI radar observes targets in a low signal-to-noise ratio environment. We assume the MTI radar is modeled by the two functions, measurement model $p_M(y|x)$ and detection model $p_D(x)$. The detection model can be described as

$$p_D(x) = \prod_{i \in \{R,A,E,D\}} p_D^i(x_i) \tag{10}$$

where p_D^i is the detection function for each component i in $\{R,A,E,D\}$ which represent *range, azimuth, elevation*, and *Doppler (range rate)*[3]. Each x_i represents the *true* value for component i. For i in $\{R, A, E, D\}$, we have

$$p_D^i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\int_{FOV_i} \exp\left[-\frac{1}{2}\left(\frac{y_i - x_i}{\sigma_i}\right)^2\right] dy_i \tag{11}$$

where FOV_i is the field of view[4] for the component i, and σ_i is the standard deviation for measurement component i, which will be discussed later. The measurement value model $p_M(y|x)$ is likewise decomposed as

$$p_M(y|x) = \prod_{i \in \{R,A,E,D\}} p_M^i(y_i|x_i) \tag{12}$$

For i in $\{R, A, E, D\}$, we have

$$p_M^i(y_i|x_i) = \frac{\exp\left(-\frac{1}{2}\left(\frac{y_i - x_i}{\sigma_i}\right)^2\right)}{\frac{1}{\sqrt{2\pi}\sigma_i}\int_{FOV_i} \exp\left(-\frac{1}{2}\left(\frac{y_i - x_i}{\sigma_i}\right)^2\right) dy_i} \tag{13}$$

[3] signal strength can also be explicitly modeled, see [9] for details.

[4] It is generally an interval or a union of intervals so that the integral in Eq. (9) can be expressed by error functions.

where σ_i is the measurement error standard deviation given by [2]

$$\sigma_i = \max\left\{\frac{\delta_i}{\sqrt{2SNR}}, \sigma_i^{\min}\right\} \tag{14}$$

with δ_i being the *sensor resolution* for the component i and SNR is the signal-to-noise ratio.
The false alarm probability is determined by [2]

$$P_{FA} = \exp(-SNR_{TH}) \tag{15}$$

from which the expected number of false alarms can be calculated as

$$n_{FA} = P_{FA} \prod_{i \in \{R,A,E,D\}} \frac{\mu(\mathrm{FOV}_i)}{\delta_i} \tag{16}$$

where SNR_{TH} is the SNR threshold for detection and $\mu(FOV_i)$ is the volume of the field of view for measurement component i.

We model the target state as a 4-dimensional vector, i.e., 2-D position and 2-D velocity as described earlier. We also assume that the target state distributions of all the *old* tracks $\bar{\tau}$ have gaussian representations, as

$$p(x|\bar{\tau}) = \left(\det(2\pi\bar{V})\right)^{-1/2} \exp\left(-\frac{1}{2}\|x - \bar{x}\|^2_{\bar{V}^{-1}}\right) \tag{17}$$

Then it follows from the linearization of nonlinear measurement equations that the measurement to track association likelihood is

$$L(y|\bar{\tau}) = \left(\det(2\pi S)\right)^{-1/2} \exp\left(-\frac{1}{2}\|y_{RAED} - \bar{y}_{RAED}\|^2_{S^{-1}}\right) \tag{18}$$

where $y_{RAED} = (R, A, E, D)$ is the measurement vector, $\bar{y}_{RAED} = (\bar{R}, \bar{A}, \bar{E}, \bar{D})$ is the nonlinear projection of target state estimate $\bar{x}$ onto the sensor's measurement space, S is the innovations variance matrix defined by

$$S = H\bar{V}H^T + \Sigma \tag{19}$$

with

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \bar{R}\cos\bar{E} & 0 & 0 \\ 0 & 0 & \bar{R} & 0 \\ 0 & 0 & 0 & \bar{R} \end{bmatrix}^{-1} \begin{bmatrix} g_R^T & 0 \\ g_A^T & 0 \\ g_B^T & 0 \\ (\bar{v} - v_S)^T & g_R^T \end{bmatrix} \begin{bmatrix} I_3 & 0 \\ -\bar{D}I_3 & \bar{R}I_3 \end{bmatrix} \tag{20}$$

$$\begin{bmatrix} g_R^T \\ g_A^T \\ g_E^T \end{bmatrix} = \begin{bmatrix} \cos\bar{A}\cos\bar{E} & \sin\bar{A}\cos\bar{E} & \sin\bar{E} \\ -\sin\bar{A} & \cos\bar{A} & 0 \\ -\cos\bar{A}\sin\bar{E} & -\sin\bar{A}\sin\bar{E} & \cos\bar{E} \end{bmatrix} \begin{bmatrix} \ell_R^T \\ \ell_A^T \\ \ell_E^T \end{bmatrix} \tag{21}$$

and

$$\Sigma = diag(\sigma_R^2, \sigma_A^2, \sigma_E^2, \sigma_D^2) \tag{22}$$

where $\bar{v}$ is the velocity component of $\bar{x}$, v_s is the sensor velocity vector, I_3 is a 3x3 identity matrix, and $(\ell_i)_{i\in\{R,A,E\}}$ is the sensor pointing system defining azimuth and elevation directions.

The initial state estimate is obtained as described in Section 2. After a track is initiated, it is a rather straightforward application of the *Extended Kalman filter* to update a track by a measurement. Namely

$$\hat{x} = \bar{x} + K(y_{RAED} - \bar{y}_{RAED}) \tag{23}$$

with

$$K = \bar{V}H^T S^{-1} \tag{24}$$

and $\hat{V} = (I - KH)\bar{V} = \bar{V} - KSK^T$ (25)

that update the mean vector to $\hat{x}$ and the estimation error variance matrix to $\hat{V}$. When no measurement is assigned, a track is simply extrapolated based on the target dynamic model.

5. SIMULATION RESULTS

This section describes the results of simulations. The simulation is implemented in a MATLAB environment. A two sensors, four targets scenario shown in Figure 2 was created and simulated measurements based on the models described in Section 3 were generated. The two MTI sensors are located at approximately 2000 feet above the locations (0,0) and (50 km,0). The nominal set of key parameters used are given in Table 1.

Figure 3 shows the receiver operating characteristic (ROC) of the system. With $\overline{SNR} = 14\text{dB}, SNR_{th} = 10\text{dB}$ corresponds to a false alarm rate of, approximately, $P_{FA} = 4.5 \cdot 10^{-5}$, which translates into an average number of 6 false alarms in the sensor FOV (field of view). Figure 4 shows the typical sensor detections from two MTI sensors for the entire 80 minutes of simulation. It can be seen that the clutter density is relatively high and targets are often missed especially when their relative velocities (range rate) are low. Figure 5 shows one sample run results with a single MTI sensor only and Figure 6 shows the results with both sensors. Figures 7 and 8 show the performance curve with 50 Monte Carlo simulations. It is clear that with two sensors, not only all four targets are reliably tracked (versus two or three targets tracked with a single sensor) but the track probabilities are also higher while the average false track probabilities remain low. Since we initiated a track with a single detection, there will be more tracks created in a multisensor situation than in the single sensor cases as shown in Figure 7. However, with the algorithm presented here, the life of most of the false tracks is very short, as shown in Figure 9. In other words, the tracking/fusion algorithm is able to prune the false tracks rather effectively.

Azimuth FOV	sensor 1 (0 - 90°), sensor 2 (-90 - 0°)
Range FOV	15 - 55 km
Sensor Resolution	Range:300m, Doppler:1m/s, Azimuth: 0.7 degree
Doppler threshold	2.5 m/s
Sampling Interval	2 minutes[5]
Mean SNR	14 dB
SNR threshold	10 dB
Track pruning threshold	$P_{TH} = 0.01$

Table 1. Nominal Parameter Values

A more stressing example with the same $\overline{SNR}$ but with lower threshold $SNR_{th} = 8\text{dB}$ was also simulated which corresponds to the false alarm rate of $P_{FA} = 1.2 \cdot 10^{-4}$, that generates about 16 false detections per scan in the surveillance region. The simulation results given in Figures 10 and 11 show that even though the average number of false track increases significantly, the tracking performance does not degrade too much. Specifically, most of the four targets are tracked with reasonably high probabilities. We also varied the average SNR to study its impact on the tracking performance. Since the detection probability is a function of the SNR (see Figure 3), the tracking performance, particularly, the average good track probabilities, decreases from 0.9 to about 0.7 when the average SNR decreases from 14 dB to 10 dB (see Figures 8 and 12). The pruning threshold is the last factor we studied in the simulation. It turns out that the tracking performance (not shown here) is relatively insensitive to the pruning threshold, particularly for the threshold ranging from 0.01 to 0.1. However, the performance degrades notably when the threshold goes beyond that. This simulation can be used also for threshold optimization [10].

[5]This low sampling rate is typical for an air to ground surveillance system.

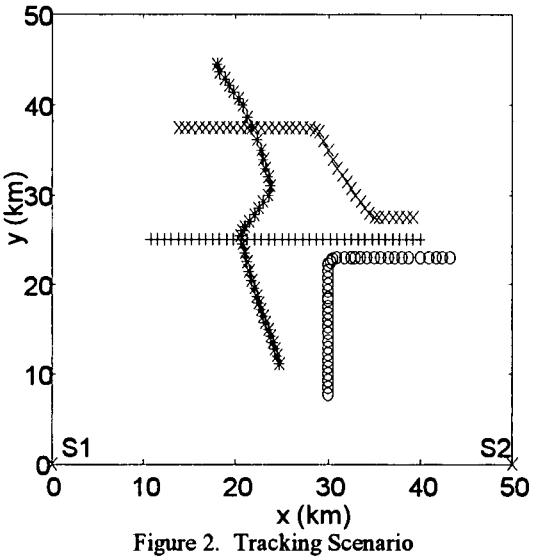


Figure 2. Tracking Scenario

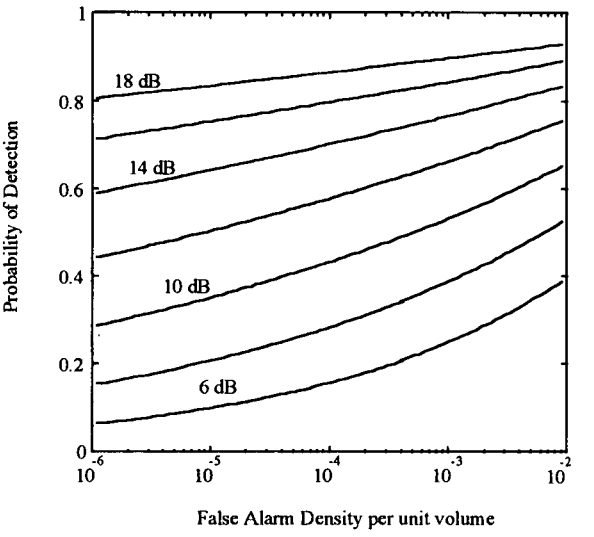


Figure 3. Receiver Operating Characteristic vs. SNR

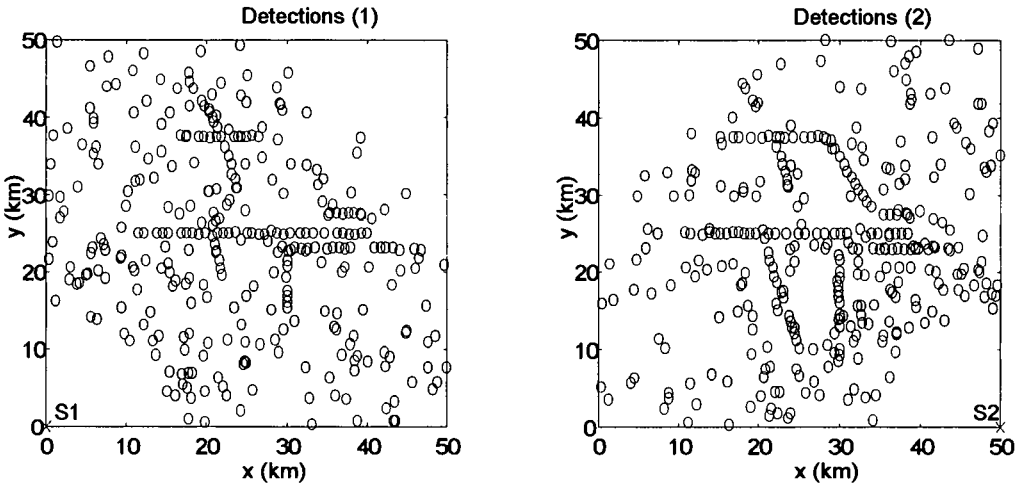


Figure 4. Sensor 1 and sensor 2 detections

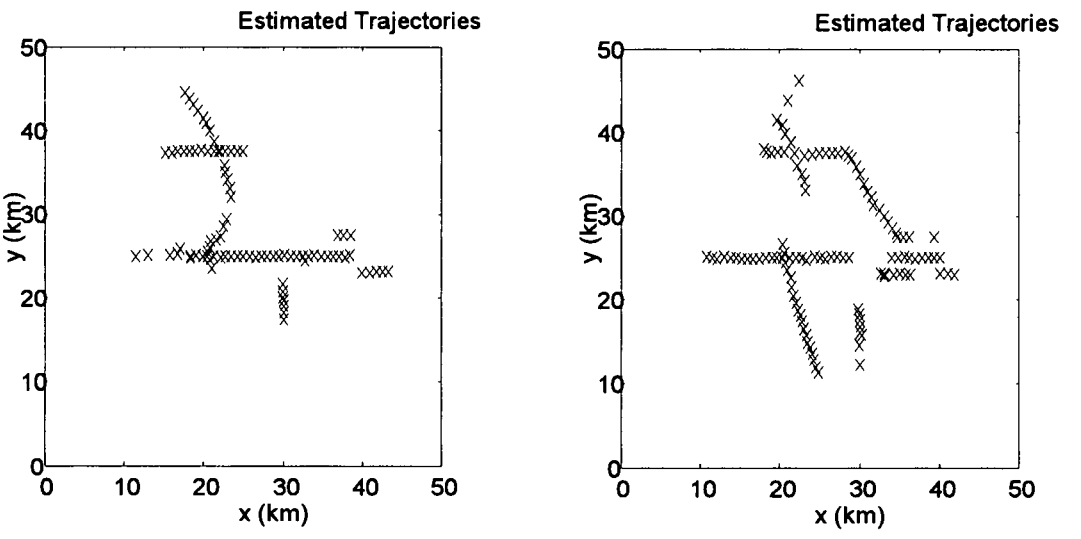


Figure 5. Tracking results with a single sensor (left: sensor 1; right: sensor 2)

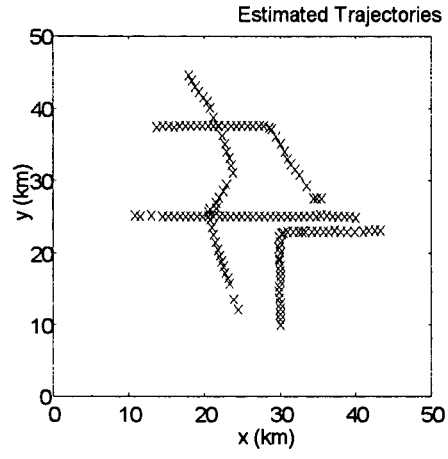


Figure 6. Tracking results with both sensors

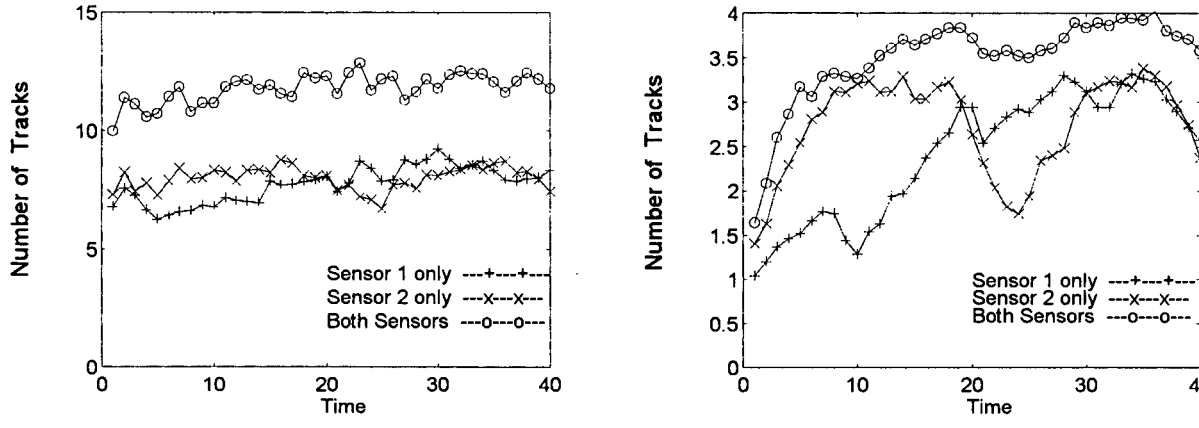


Figure 7. Monte Carlo performance curves (I): average number of total tracks and good tracks ($\overline{SNR}$ = 14 dB, SNR_{th}= 10 dB)

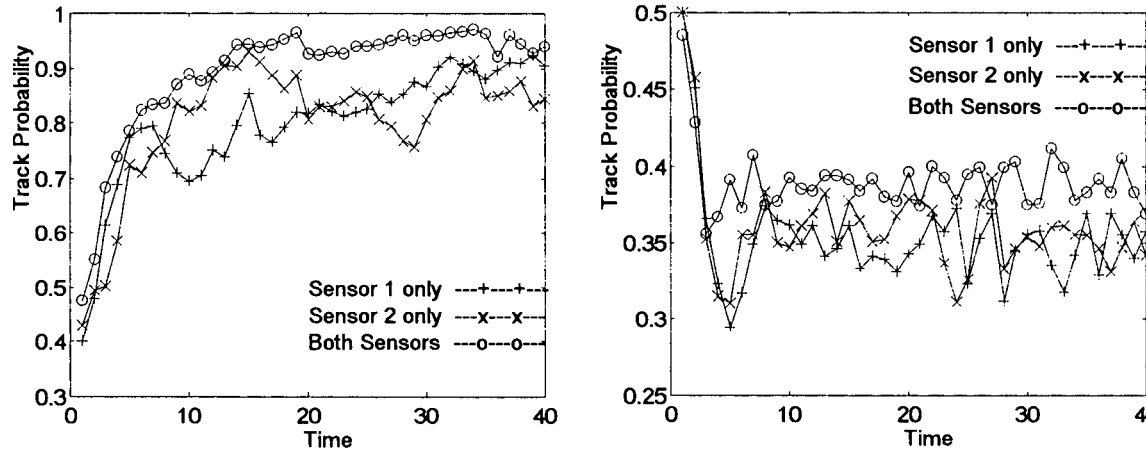


Figure 8. Monte Carlo performance curves (II): average good track and false track probabilities ($\overline{SNR}$ = 14 dB, SNR_{th}= 10 dB)

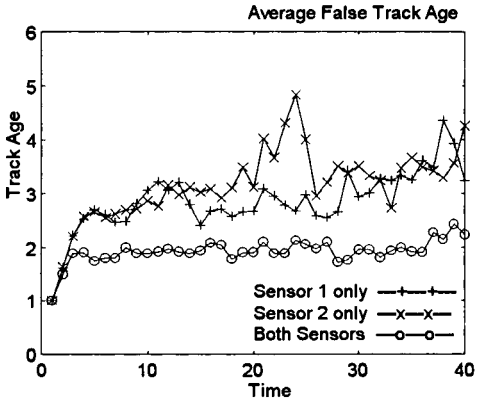


Figure 9. Average life of false tracks

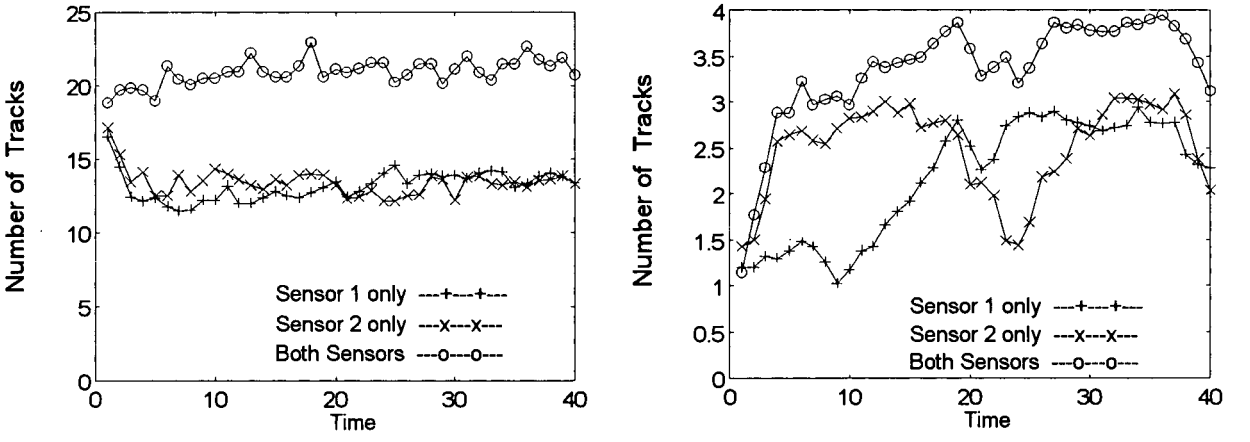


Fig. 10. Monte Carlo performance curves (I): average number of total tracks and good tracks ($\overline{SNR}$ = 14 dB, SNR_{th} = 8 dB)

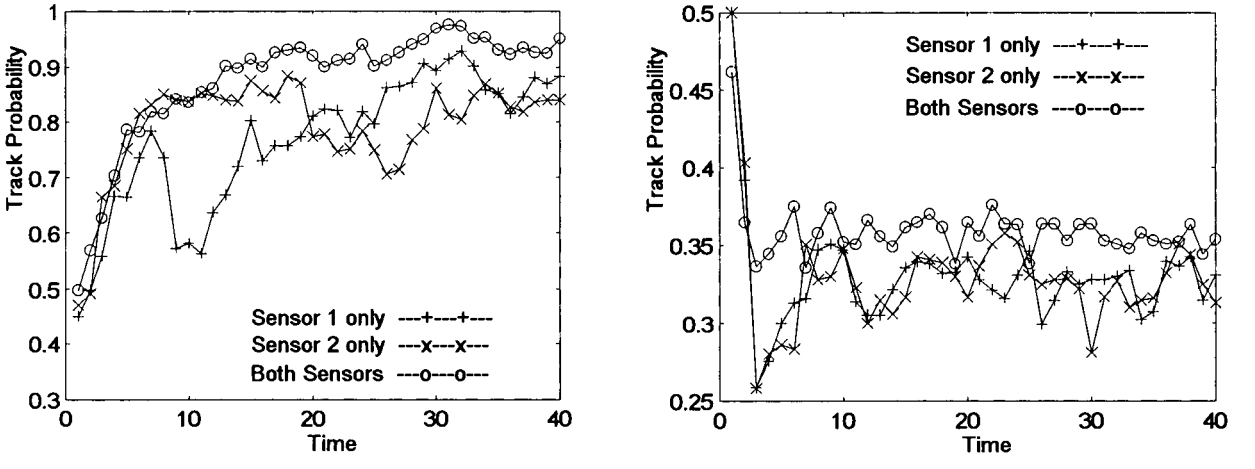


Fig. 11. Monte Carlo performance curves (II): average good track and false track probabilities ($\overline{SNR}$ = 14 dB, SNR_{th} = 8 dB)

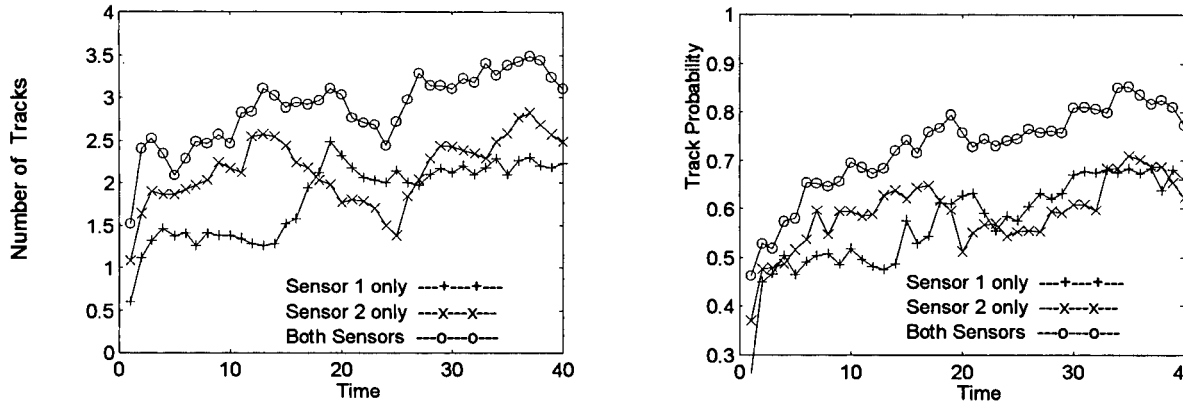


Fig. 12. Monte Carlo performance curves : average number of good tracks and average good track probabilities ($\overline{SNR}$ = 10 dB, SNR_{th} = 10 dB)

6. SUMMARY AND CONCLUSION

This chapter considers the MTI tracking and fusion problem for scenarios with high false alarm density. In order to develop a simple, yet comprehensive, algorithm suitable for a dense clutter environment, we adopted the track-oriented approach where a form of "greedy" nearest-neighbor association algorithm is used. In this approach, a centralized fusion architecture was assumed. To establish good tracks and eliminate false tracks effectively, tracks are initiated based on a single measurement and a probabilistic track score is assigned for each track to determine its strength. This track score is computed based on the association history using a multiple model algorithm with an underlying Markov chain. Tracks with scores below a certain threshold are then discarded and similar tracks are combined.

The algorithm has been implemented in a MATLAB environment. Tracking results with multiple MTI sensors using extensive Monte Carlo simulations have also been obtained. Several key parameters such as SNR and detection threshold were varied to demonstrate the feasibility of the algorithm. Future work includes the development of algorithms for incorporating terrain information and integrating data from other intelligent sources such as infrared and COMINT.

7. REFERENCES

[1] Mori, S., C. Y. Chong, E. Tse, and R. P. Wishner (1986). Tracking and Classifying Multiple Targets without A Priori Identification, *IEEE Trans. on Automat. Contr.*, Vol. AC-31, No. 5, pp. 401 - 409, May, 1986.

[2] Blackman, S. S. (1986). Multiple-Target Tracking with Radar Applications, Artech House, Inc.

[3] Bar-Shalom, Y. and T. Fortmann (1988). Tracking and Data Association, Artech House, Inc.

[4] Bar-Shalom, Y. (1990), editor. Multitarget-Multisensor Tracking: Advanced Applications, Vol. I, Artech House, Inc.

[5] Bar-Shalom, Y. (1992), editor. Multitarget-Multisensor Tracking: Applications and Advances, Vol. II, Artech House, Inc.

[6] Bar-Shalom, Y. and X. R. Li (1995). Multitarget-Multisensor Tracking: Principles and Techniques, YBS Publishing (Storrs, CT 06269-3157).

[7] Reid, D. B. (1979). An Algorithm for Tracking Multiple Targets, *IEEE Trans. on Automat. Contr.*, Vol. AC-24, pp. 843 - 854, Dec., 1979.

[8] Bar-Shalom Y. and K. C. Chang (1994). FUSEDAT - Fusion, Estimation and Data Association for Tracking, YBS Interactive Software, 1994.

[9] Chang, K. C., S. Mori, and C. Y. Chong (1994). Evaluating a Multiple-Hypothesis Multi-Target Tracking Algorithm, *IEEE Trans. on Aerospace and Electronic Systems.*, Vol. AES-30, No. 2, pp. 578-590, April, 1994.

[10] Chang, K. C. (1994). Multitarget Tracking with Adaptive Detection Thresholds, submitted to *IEEE Trans. on Aerospace and Electronic Systems* for publication, 1994.

System-Level Track Fusion for Command and Control

M. P. Dana and J. L. Dana

Hughes Aircraft Company
P. O. Box 3310
Fullerton, California 92634, USA

SUMMARY

The central elements of this paper are the development of a statistical approach to the problem of track fusion, specifically, track correlation, and system track maintenance in distributed command and control systems. The basic approach is a sequential decision process based on the concepts proposed by Bar-Shalom [1] and Waltz and Llinas [2]. This paper presents (1) the functional requirements for track management in conventional air defense systems and extended air defense systems (which include low-observable aircraft, cruise missile, and tactical ballistic missile defense), (2) the theoretical foundations for the sequential, multiple hypothesis decision processes used in the FAAD C^2I, IADS and NATO AEW Integration Programs; (3) track maintenance in distributed systems, and (4) a discussion of the practical aspects of the application of the theory to systems constrained by current data link message standards. The essential theory for sequential correlation decisions is based on the sequential probability ratio test for multiple alternative hypotheses [3-8]. The decision statistics and criteria are derived together with the theoretical performance trade-offs among type I and type II errors and decision times.

1. INTRODUCTION

Conventional air defense systems (such as the NATO AEW/Ground System Integration Programs, NATO Iceland Air Defense System, and the U.S. Army Forward Air Defense C^2I System) generally employ combinations of sensors, weapon subsystems, and external system data to compile a surveillance air picture. Data from external systems usually is received via tactical data links (for example, TADIL B, Link 11, JTIDS). In order to maintain the best possible recognized air picture and to make the most efficient use of the limited capacity of the tactical data links, the C^2I subsystems therefore must be able to integrate track data from a local sensor suite with track position data from other C^2I subsystems, Airborne Early Warning, and external track sources into a composite system track file. The system track file must be complete, non-redundant and of sufficient quality to support the requisite Command and Control decisions such as distribution of selected track information to weapon subsystems, target-to-weapon pairing, acquisition by the weapon subsystems, and coordination with other C^2 and Battle Management subsystems. The critical issue often is to conserve the communication resource by suppressing dissemination of redundant track information.

In extended air defense systems, which include low-observable, cruise missile and tactical ballistic defense in addition to conventional air defense, the problem is somewhat more complicated. First, it is compounded by the addition of high precision tracking radars with fixed face, phased-array antennas, which are necessary to detect and track low-observable objects and ballistic missiles. These radars may be employed both for surveillance and fire control. Second, the ground-based elements may not be as constrained by the limitations of tactical data links as conventional systems; future systems may be netted by wide area networks of fiber-optical cables. The critical issues are (1) to ensure the maximum surveillance coverage against low-altitude, low-observable objects, (2) to provide high quality track information to multiple, geographically distributed users, and (3) to maintain a common data base at multiple C^2 subsystems.

On a qualitative level, the track management requirement at the C^2 subsystems of extended air defense systems is to correlate multiple tracks of a common objects in order:

1. To maintain track number continuity as an object moves through the coverage envelopes of multiple sensors or subsystems;
2. To fuse the sensor or local identification, classification and discrimination data into a single system-level track description; and
3. To maintain a common data base of high quality track information at multiple, distributed subsystems.

The last item is particularly important in order to avoid saturation of the data links and to avoid ambiguous or conflicting decisions at distinct C^2 and weapon subsystems.

The central elements of a track management function for an extended air defense system with the characteristics described above are (1) the track correlation logic and (2) a method for updating multiple, distributed data bases. The paper also provides a top-down systems analysis approach to the design of a track correlation and maintenance logic for multiple radar surveillance systems, which is illustrated with the extended air defense design problem derived from NATO air defense systems. Moreover, all of the performance data is presented parametrically versus separation normalized by the subsystem tracking errors. Therefore, the same technique and performance analyses can be used for any system design problem once the subsystem errors are known; the performance data can be scaled to whatever subsystem tracking errors are anticipated. These techniques have been applied to the U.S. Army Forward Area Air Defense C^2I System, the AEW integration programs for NATO and to the Iceland Air Defense System now under development.

This paper will present in detail discussions of the following topics: (1) functional requirements for track management in extended air defense systems, particularly those with geographically distributed C^2 and weapon subsystems; (2) examples of quantitative performance requirements for track correlation; (3) a statistical approach to track correlation; (4) track maintenance in distributed systems based on Kalman filtering theory; and (5) implementation issues in communications limited systems.

2. AN EXAMPLE: FAAD C^2I

In order to motivate the general theory which will be addressed in Section 3, consider the US Army Forward Area Air Defense (FAAD) System. The primary operational mission of the FAAD System is to provide effective counter-air protection for the Division ground units. This is accomplished by the Command, Control and Intelligence (C^2I) element of the system through the detection and engagement of hostile air targets before those targets can attack the Division assets. Therefore, the basic requirement for the FAAD C^2I System is to collect, process and disseminate to the FAAD Weapon Systems the time-critical air track information necessary for engagement of hostile aircraft.

This processing must be accomplished in the context of the FAAD C^2I design, which consists of a collection of spatially distributed, semi-autonomous, redundant C^2 subsystems. In this concept, each subsystem (i.e., the Sensor C^2 units, the Air Battle Management Operation Center, and the Army Airspace Control Center) is self-contained. Each subsystem may have one or more sources of local track data, including a FAAD Ground Based Sensor (GBS), HIMAD, Air Force JTIDS PPLI, Army EPLRS Position Reports, and Adjacent FAAD Systems, as illustrated in Figure 2.1.

The requirement "to process" air track information specifically includes the requirement that each C^2 element be able to assemble a single system-level track file from the remote tracks received from other C^2 elements and one or more local track files. The essential functional requirement, therefore, is a Track Correlation function. This function must correlate the local track reports from the individual sources against a system track file; when two or more sources are determined to

report tracks for the same aircraft, one of the tracks is selected to represent the "system level" track for that aircraft. An individual sensor track is adequate to cue a weapon system; averaging tracks for additional accuracy is not required. Thus reporting responsibility rules based on track quality are used to ensure that each subsystem maintains an identical copy of the system track file. *The data fusion problem for the FAAD C^2I application, therefore, reduces to that of track association or correlation.*

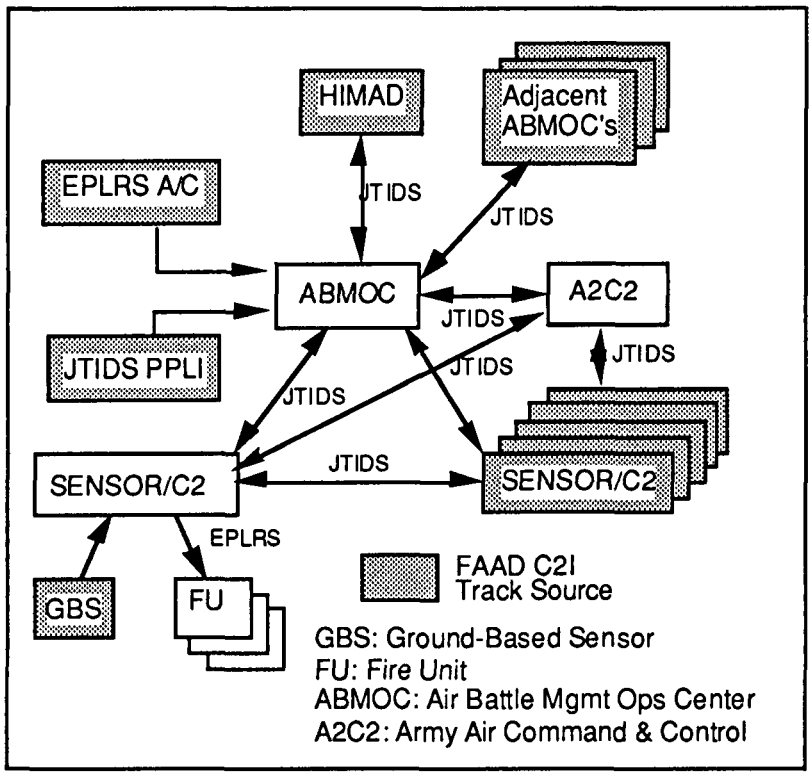


Figure 2.1: FAAD C^2I System Architecture. System tracks and ID information are exchanged among the C^2 subsystems over the FAAD JTIDS net.

The discussion throughout the remainder of this paper will consider the case of FAAD Ground Based Sensor tracks only; external tracks will not be considered. Thus, the correlation logic design problem is limited to the problem of correlation at a Sensor C^2 subsystem of new local GBS tracks with remote GBS tracks received from other Sensor C^2 subsystems via the FAAD JTIDS data link.

2.1 FAAD Error Sources

In order to design and evaluate a track correlation logic, it is necessary first to describe quantitatively the sources of track error which drive the correlation process. It is important for this application to consider both the random and systematic error components in the tracks. The random components include the random measurement errors at the sensors, and data link and time reference quantization errors. The systematic errors include sensor calibration, north alignment, and leveling as well as the error introduced by the coordinate transformation from local sensor coordinates to system coordinates.

Table 2.1: Systematic Errors

Coordinate Conversion Errors (σ)	
Sensor Position σ_{CC}	25 m
North Alignment σ_{NR}	60 m
Data Link Quantization Errors (σ)	
Position σ_{PQ} (LSB=100 m)	28.9 m
Velocity σ_{VQ} (LSB=20 m/s)	5.8 m/s

The quantitative performance requirements on the C^2I system assume that the sensor can deliver tracks (under some conditions) with position accuracy[1] (1σ) no worse than 100 meters in the horizontal plane and 167 meters in the vertical direction. The errors in the delivered track include coordinate conversion errors from radar to system coordinates and quantization errors in the data link in addition to the normal random measurement and track prediction errors; the values assumed are shown in Table 2.1. The north alignment error is based on a 1σ value of 2 milli-radians for the azimuth error against an aircraft at 30 km from the radar.

Tracks in the FAAD C^2I System are five dimensional, consisting of a horizontal (x,y) position, velocity in the horizontal plane, and height. However, if one considers one dimension of the track data, for example the x coordinate, then the track state vector is $X = [x \quad \dot{x}]^t$ and, the covariance matrix Σ can be represented [1, 9] by

$$\Sigma = E\left[X X^t\right] = \begin{bmatrix} P & C \\ C & V \end{bmatrix} = \begin{bmatrix} \sigma_P^2 & \sigma_{PV} \\ \sigma_{PV} & \sigma_V^2 \end{bmatrix} \tag{2.1}$$

Also, let $\overline{X}$ and $\hat{X}$ denote the filtered and predicted state vectors, respectively. The variance of the position error at the time t_1 of delivery to the C^2I system is, therefore,

$$\begin{aligned} \hat{P}(t_1) &= \overline{P}(t_0) + 2(\Delta t_1)\,\overline{C}(t_0) + (\Delta t_1)^2\,\overline{V}(t_0) \\ &\quad + \sigma_{CC}^2 + \sigma_{NR}^2 + \sigma_{PQ}^2 \\ \hat{C}(t_1) &= \overline{C}(t_0) + (\Delta t_1)\,\overline{V}(t_0) \\ \hat{V}(t_1) &= \overline{V}(t_0) + \sigma_{VQ}^2 \end{aligned} \tag{2.2}$$

where, $\Delta t_1 = (t_1 - t_0)$ is the elapsed time from the last track update to track delivery at the first C^2I subsystem.

An actual variance, σ_M^2, for the radar measurement error can be obtained if some assumptions are made about the radar tracking logic. For example, if it is assumed that a standard (α, β)-filter is used with the gains limited to provide approximately 3-point smoothing [1, 9], then

$$\overline{P}(t_0) = 0.833\,\sigma_M^2,\ \overline{C}(t_0) = \frac{\sigma_M^2}{2\tau},\ V(t_0) = \frac{2\,\sigma_M^2}{2\,\tau^2} \tag{2.3}$$

where τ is the radar scan rate (in this case, 2 seconds). If the delivery time Δt_1 is 0.25 seconds and if delivered accuracy is 100 meters, then the standard deviation σ_M of the measurement error must be approximately 70 meters.

The 100 meter track accuracy noted above refers to the accuracy of a local GBS track at a

[1] Position accuracy means explicitly the standard deviation ($1\,\sigma$) of the error. Track errors are assumed to be independent and normally (Gaussian) in each of the three coordinates. Moreover, the standard deviations of the errors for both horizontal components (that is, X and Y) are assumed to be equal.

Sensor C^2 unit. For the remote track, the transmitting Sensor C^2 unit must extrapolate the track forward, using the quantized velocity information, to the anticipated time of transmission. At the interface to the JTIDS radio, the position is quantized for a second time. Thus, the variance model for the remote track is

$$\begin{aligned}\hat{P}(t_2) &= \overline{P}(t_0) + 2(\Delta t_2)\overline{C}(t_0) + (\Delta t_2)^2 \overline{V}(t_0) \\ &\quad + (\Delta t_2)^2 \sigma_{VQ}^2 + \sigma_{CC}^2 + \sigma_{NR}^2 + 2\sigma_{PQ}^2 \\ \hat{C}(t_2) &= \overline{C}(t_0) + (\Delta t_2)\left[\overline{V}(t_0) + \sigma_{VQ}^2\right] \\ \hat{V}(t_2) &= \overline{V}(t_0) + \sigma_{VQ}^2 \end{aligned} \tag{2.4}$$

where, Δt_2 is the total time necessary to predict the remote track forward (from the time of the last update by the remote GBS) to obtain time coincidence of the local and remote tracks. If one considers, in addition to the normal 2 second update rate of the GBS, the I/O, buffer, and processing delays at both the remote and local Sensor C^2 units, the JTIDS terminal delays, and the transmission time necessary when multiple relays are required, a worst-case value for Δt_2 of 4 seconds is possible. Thus, the standard deviation of the remote GBS track can be as large as 175 meters.

2.2 Quantitative Performance Goals

The qualitative requirement for a **complete** and **non-redundant** system track file must be refined to quantitative upper bounds for the probabilities of **false correlation** of tracks of two distinct aircraft and **missed correlation** of two or more tracks of the same aircraft, respectively. The probability of false correlation must be further qualified by an operational requirement on the separation of aircraft. It is not reasonable, for example, to expect a high degree of discrimination between aircraft separated by less than resolution capabilities of the sensors.

In general, these quantitative requirements must be derived from operational requirements and the limitations of the physical hardware, such as the communications links, processing rate, and so forth. For the FAAD C^2I System, the system-level probability of false correlation, denoted by $P_{SYS}(FC)$, must not exceed 0.10; the probability of missed correlation denoted by $P_{SYS}(MC)$, must not exceed 0.004; that is,

$$P_{SYS}(FC) \le 0.10, \quad P_{SYS}(MC) \le 0.004 \tag{2.5}$$

It will be assumed that the missed correlation applies to aircraft within 100 meters of each other. The false correlation requirement is not complete; there is the issue of the aircraft separation for which the requirement applies. Operationally, distinct formations of aircraft may approach to within one-half mile or approximately 700 meters of each other in normal operations. Thus, the requirement will be restated as

$$P_{SYS}(FC \mid \text{Separation} \ge 700\text{ m}) \le 0.10 \tag{2.6}$$

3. A STATISTICAL THEORY OF TRACK CORRELATION

In order to formulate the track correlation problem, let X_1 and X_2 represent tracks from two sources (systems or sensors). In the following it will be assumed that each track X_i is a random vector in the n dimensional real, Euclidean space $\mathbf{R}^n$ and are distributed as a multivariate normal random variable, denoted

$$X_i \in \mathbf{R}^n, X_i \approx N(\mu_i, \Sigma_i) \tag{3.1}$$

The expected value and variance of X_i are denoted

$$E[X_i] = \mu_i, \quad E\left[(X_i - \mu_i)^2\right] = \Sigma_i \tag{3.2}$$

In the following it will be assumed that a track is an estimate of the state (at a minimum, position and velocity) of an aircraft based on a time-ordered set of measurements of the position of the aircraft; more specifically,

$$X_i = \sum_{k=1}^{m} A_k Z_{i,k} = A_1 Z_{i,1} + A_2 Z_{i,2} + \ldots + A_m Z_{i,m} \quad (3.3)$$

where the matrix weighting factors $\{A_k \mid k = 1, 2, \ldots m\}$ are given, for example, by the Kalman filter formulation of the optimal estimation problem, and the $\{Z_{i,k} \mid k = 1, 2, \ldots m\}$ are the measurements at source i [1, 9, 10].

The tracks represent the same aircraft if $\mu_1 = \mu_2$; otherwise, they represent two distinct aircraft. A correlation or association test now can be formulated as a statistical test of two alternative hypotheses about the means of the two underlying probability distributions. The space of possible mean vectors m is $\mathbf{R}^n$. Given a covariance matrix S (that is, a symmetric, positive definite matrix), let S_N be a neighborhood of the zero vector 0 defined by

$$S_N = \left\{\mu \in \mathbf{R}^n \mid \mu^t \Sigma^{-1} \mu \le \lambda_N\right\} \quad (3.4)$$

and let S_A be a subset of the complement of S_N defined by

$$S_A = \left\{\mu \in \mathbf{R}^n \mid \mu^t \Sigma^{-1} \mu \ge \lambda_A\right\} \quad (3.5)$$

where $\lambda_N < \lambda_A$. Note that S_A and S_N are disjoint subset of $\mathbf{R}^n$; that is,

$$S_N \cup S_A \subset \mathbf{R}^n, \quad S_N \cap S_A = \varnothing \quad (3.6)$$

A correlation test is now a test of the two alternative hypotheses H_N and H_A (called the null and the alternative hypotheses, respectively) defined by

$$H_N = \left\{(\mu_1 - \mu_2) \in S_N\right\}$$
$$H_A = \left\{(\mu_1 - \mu_2) \in S_A\right\} \quad (3.7)$$

An appropriate statistic for testing the null hypotheses H_N against the alternative hypotheses H_A is the chi-squared statistic defined in [3, 6, 7, 11].

$$\xi = \left(X_1 - X_2\right)^t \left(\Sigma_1 + \Sigma_2\right)^{-1} \left(X_1 - X_2\right) \quad (3.8)$$

In particular, a convenient test would be of the form: accept the null hypothesis H_N if $\xi \le C$; otherwise reject H_N and accept the alternative hypothesis H_A. The constant C can be defined by the requirement that

$$\Pr\left[\xi \le C \mid H_N\right] \ge 1 - \alpha \quad (3.9)$$

where the error rate α is the acceptable error for rejecting H_N when it is true. Alternatively, the constant C could be chosen in order to meet the acceptable false correlation rate β; that is,

$$\Pr\left[\xi \le C \mid H_A\right] \le \beta \quad (3.10)$$

In general, the statistic ξ has a non-central chi-squared probability distribution, denoted as $\chi^2(n,\lambda)$, since X_1 and X_2 are assumed to be multivariate normally distributed random vectors, where

$$\lambda = \left(\mu_1 - \mu_2\right)^t \left(\Sigma_1 + \Sigma_2\right)^{-1} \left(\mu_1 - \mu_2\right) \quad (3.11)$$

The probability density function for the non-central chi-squared probability distribution is defined as follows [2, 8, 9 and 10]:

$$f(\xi \mid \lambda) = h(\xi)\, e^{-\lambda/2}\, G\left(\frac{n}{2}, \frac{\lambda\xi}{4}\right)$$
$$h(\xi) = \frac{\xi^{n/2-1}\, e^{-\xi/2}}{2^{n/2}\, \Gamma\left(\frac{n}{2}\right)}, \quad G(n,\xi) = \sum \frac{\Gamma(n)\xi^k}{\Gamma(n+k)\, k!} \quad (3.12)$$

where Γ is the standard gamma function

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}\, e^{-y}\, dy \quad (3.13)$$

The parameter n -- then number of degrees of freedom -- in Equation 3.11 is the dimension of the track state vector. The probability that $\xi \leq C$ can be evaluated numerically with the procedures given in Press, et. al., *Numerical Recipes* [12]; the American Mathematical Society has published extensive tables [13].

Now consider the requirements of the FAAD C^2I outlined in Section 2. The error rate α for missed correlations is 0.004, and the error rate β for false correlations is 0.10 when the actual separation between aircraft is 700 meters. The 700 meter criterion represents a normalized separation of approximately 3.5; the 100 meter criterion represents a normalized separation of 0.5. Thus, the parameters λ_N and λ_A which define the two hypotheses (Equations 3.4 and 3.5) can be assumed to be 0.25 and 12.25, respectively. Tracks have five components; thus the number of degrees of freedom is 5 in Equation 3.11. The gate C for the missed correlation criterion is the solution of

$$\Pr\left[\chi^2(5,\lambda) \leq C\right] \geq 0.996 \qquad (3.14)$$

for $\lambda = 0.25$; that is, C = 18.2. Similarly, the gate C for false correlation criterion is determined by

$$\Pr\left[\chi^2(5,\lambda) \leq C\right] \leq 0.10 \qquad (3.15)$$

for $\lambda = 12.25$; in this case, C = 8.2.

A comparison of the two correlation decision criteria is shown in Figure 3.1. Based on a single correlation decision or trial, only one of the two criteria can be satisfied; both cannot be satisfied simultaneously.

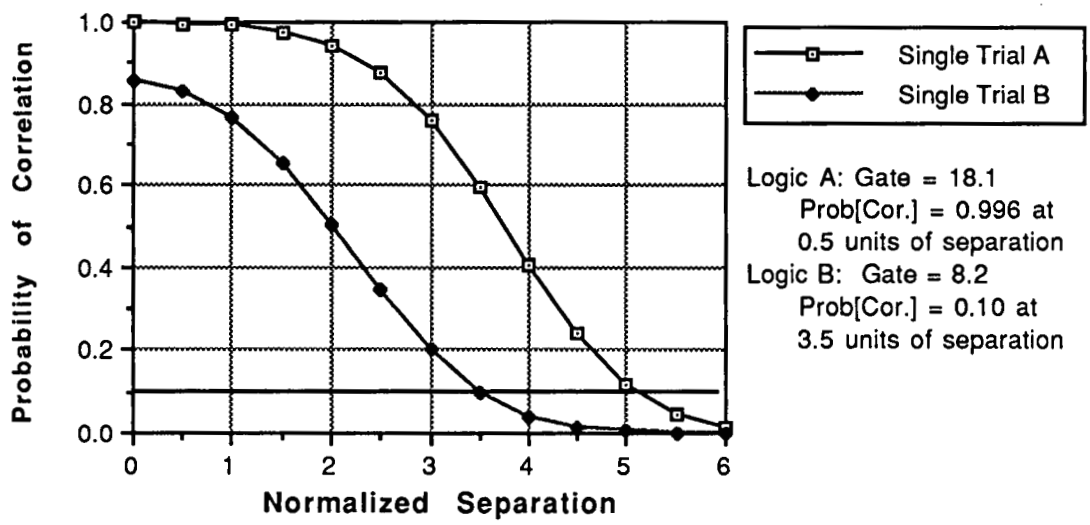


Figure 3.1: Single Trial Probabilities of False Correlation versus Aircraft Separation. Single Trial Logic A is optimized for the required missed correlation rate; Logic B is optimized for the required false correlation rate.

The solution for this apparent dilemma is to apply Wald's [4, 5, 8] concept for sequential analysis, often called the sequential probability ratio test. In this concept, two decision criteria are used, corresponding to an inner gate (e.g., $\xi \leq C_1$) for "correlation" decisions and an outer gate (e.g., $\xi \geq C_2$) for "no correlation" decisions. For the region of ambiguity (e.g., $C_1 < \xi < C_2$) the decision is deferred until another observation or sample is obtained. By selecting the two gates C_1 and C_2 appropriately, both error rates can be achieved simultaneously. This price, of course, is that an extended sample time may be required

before sufficient data is collected for the decision.

Formally, the sequential probability ratio test is defined by the statistic Z, where

$$Z = \frac{f\left(\xi \mid H_N\right)}{f\left(\xi \mid H_A\right)} \tag{3.16}$$

The correlation test is then:

(i) Accept the null hypothesis H_N if $Z \leq a$

(ii) Accept the alternative hypothesis H_A if $Z \geq b$

(iii) Continue sampling if $a < Z < b$.

"Continue sampling" is the context of track correlation means to obtain the next update of each of the two track under consideration. A track is, by Equation 3.3 a statistic, that is, a function of all of the measurements available to a system concerning the state of an aircraft. Wald gives a convenient approximation for the choice of the two decision criteria, a and b. Specifically,

$$a \cong \frac{\beta}{1-\alpha}, \quad b \cong \frac{1-\beta}{\alpha} \tag{3.17}$$

The problem now is to reduce the formal definition of the sequential probability ratio test to a usable test in terms of ξ rather than Z. To this end, set [4, p. 315]

$$z = \mathrm{Ln}\,(Z) = \frac{\lambda_A - \lambda_N}{2} + \mathrm{Ln}\left\{G\left(\frac{n}{2}, \frac{\lambda_N\,\xi}{4}\right)\right\} - \mathrm{Ln}\left\{G\left(\frac{n}{2}, \frac{\lambda_A\,\xi}{4}\right)\right\} \tag{3.18}$$

Two gates, C_1 and C_2, are obtained by solving Equation 3.17 for ξ; first with z = Ln(a), and then with z = Ln(b). Given the non-linear nature of Equation 3.17, Brent's method[2] was applied to this problem in order to find the roots.

[2] Brent's method is a straightforward, numerical method for finding roots of non-linear equations; see Ref. 12 for details. The convergence properties, when applied to this problem, are quite good.

In the FAAD C^2I example, the solutions for the two gates, C_1 and C_2, are 4.7 and 24.6. That is, the two tracks are associated if $\xi \leq 4.7$; they are not associated if $\xi \geq 24.6$; otherwise, there is no decision. The performance of the sequential decision logic is shown in Figure 3.2.

4. APPLICATIONS TO AIR DEFENSE

In air defense systems there are two limitations of the sequential approach to decisions. First, one does not have the luxury of an extended decision time; an actual decision is necessary at some point in time. Second, only rarely can a correlation problem be reduced to a one-versus-one test; there usually is more than one candidate for correlation when a new track is received from a track source.

Consequently a modified sequential decision logic, which takes the two constraints into consideration, was developed for FAAD based on the sequential probability ratio test. This logic uses Small, Large and Gross Gates, as illustrated in Figure 4.1 for the case of a plane of two degrees of freedom, to identify probable correlations, non-correlations, and the ambiguous cases. The small gate corresponds to the gate C_1 described previously; the gross gate corresponds to the C_2 gate.

The actual correlation decisions are determined as follows. If there are no remote tracks in the Gross Gate (GG), then the new local track is unique or uncorrelated (UC). If there is exactly one remote track in the Gross Gate, and it is also within the Small Gate (SG) (that is, R2 through R6 of Figure 4.1 do not exist), then the new local track is Firmly Paired or Correlated (FP) with the remote track (that is, R1 in Figure 4.1). Thus, the logic yields rapid (that is, single trial) decisions in low density environments. For any other situation, a conditional correlation decision is made which is confirmed or modified based on subsequent track updates from both the local and remote sensors.

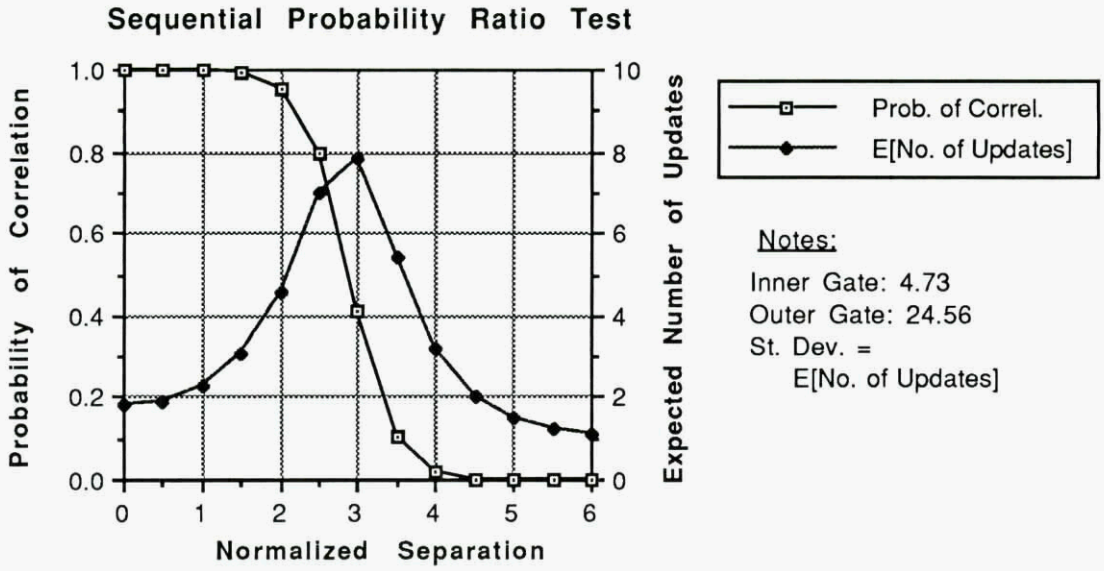


Figure 3.2: Probability of Correlation and the Expected Decision Time versus Aircraft Separation. The standard deviation of the decision time is approximately equal to the expected time; therefore, extended decision times can be anticipated for aircraft separated by 3.5 normalized units or, for FAAD C^2I, 700 meters.

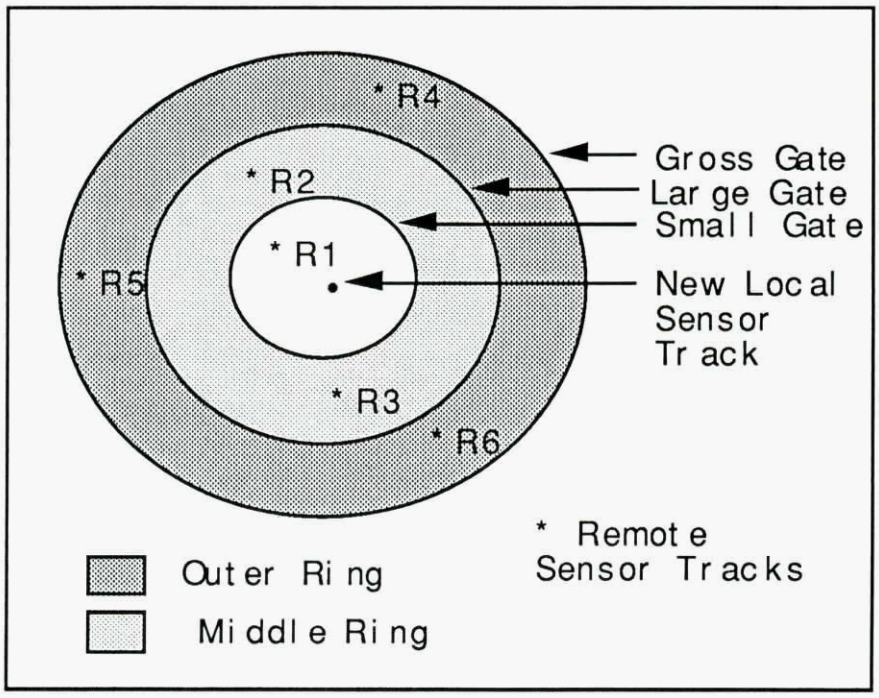


Figure 4.1: Sequential Correlation Logic Gate Structure (Horizontal Position). The Small Gate identifies the remote tracks which represent a common aircraft with a high probability; the Gross Gate identifies remote tracks which represent different aircraft with a high probability; remote tracks in the middle and outer rings are ambiguous correlation decisions which require further observations for resolution.

For example, if there is one or more remote tracks in the Outer Ring (OR) (that is, in the Gross Gate but not in the Large Gate; for example, R4, R5 and R6 in Figure 4.1), then the new local track is declared to be Conditionally Single (CS); Conditionally Single tracks are eligible for dissemination to other subsystems. If this condition persists for the next four updates of the local track, then the Conditional Single label is changed to Firm Single (or, equivalently, Uncorrelated); see Table 4.1. On the other hand, if there is one remote track in the Middle Ring (MR) or multiple remote tracks within the Large Gate (LG), then the local track is Conditionally Paired (CP) with each remote track in the Large Gate. A scoring logic based on the presence of subsequent updates in the Middle Ring or Small Gate is used to select one remote (generally, the closest of the remote tracks) for the Firmly Correlated decision; refer to Table 4.1. This process of conditional correlation decisions and confirmation produces high confidence decisions in high target densities by deferring decisions in ambiguous situations until target motion resolves the ambiguity.

As noted previously, the small gate and the gross or outer gate are defined by the sequential probability ratio test. However, the large or middle gate is not an element of the theory. This gate must be defined partly by intuition and partly by trial and error through performance modeling. Performance modeling is the subject of the next section.

Table 4.1: Scoring Logic and Decision Criteria for Tracks with Conditional Status

Status	Correlation Gates	Score
CS	No remote in GG	+5
	No remotes in LG	+2
	At least one remote in LG	-1
CP*	Remote in SG	+2
	Remote in MR	+1
	Remote in OR	-1
	Remote not in GG	Delete
Decision Criteria		
CS --> UC if Score $\geq$ 5 or no remotes in GG		
CS --> CP if Score = 0		
CP --> CS if Score = 0 for all remote CP's		
CP --> FP if Score $\geq$ 5 **		

Notes:
CP: conditional pair; CS: conditional single.
* A score is maintained for each conditionally paired remote track.
** If more than one remote has Score $\geq$ 5, then choose the closest in the horizontal plane.

5. PERFORMANCE ANALYSES

The performance analyses outlined in this section are based on the use of (1) central and non-central chi-squared statistics to evaluate the gate probabilities as functions of the error statistics as noted in Section 3 and (2) a Markov Chain model with which the sequential logic can be evaluated in closed-form. These models, in addition, can be applied to derive the system level requirements for sensor registration and coordinate conversion accuracy, which are the prerequisites for successful track management.

The sequential logic outlined in Section 4 can be represented by a state transition model. For the case of one remote track and one local track, the assignment of states is given in Table 5.1. Table 5.2 defines the possible state transitions together with the respective probabilities. A representation of the state transition structure is shown in Figure 5.1. P_G, P_L, and P_S are probabilities that the remote track will be in the Gross, Large and Small Gates, respectively; P_O and P_M are probabilities that the remote track will be in the outer and middle rings:

$$P_O = P_G - P_L \,,\; P_M = P_L - P_S \,. \qquad (5.1)$$

With this framework, a Markov chain model [14, Chapter 3] can be applied to evaluate the probability of each of the two possible decisions as well as the statistics for the number of samples necessary to reach the final decision. The set of absorbing states, in this case the Firmly Paired (or Correlated) and Firmly Uncorrelated states, is an ergodic set of states in a finite Markov chain. The fundamental theorem of finite absorbing Markov chains states that, independent of the starting state, the probability that the process is in an ergodic state after n transitions tends to unity as n tends to infinity [4]. The issues of (1) which of the two states represents the terminal point, and (2) the expected time to absorption are addressed in [14].

Table 5.1: State Definitions

State	Description	Score
0	Null state	0
1	Conditional Single	1
2	Conditional Single	2
3	Conditional Single	3
4	Conditional Single	4
5	Conditional Pair	1
6	Conditional Pair	2
7	Conditional Pair	3
8	Conditional Pair	4
9	Firmly Uncorrelated	5
10	Firmly Paired	5

Table 5.2: State Transitions and Probabilities

Current	Next States	Transition Prob.
0	1, 5, 9, 10	P_O , P_M , $1-P_G$, P_S
1	3, 5, 9, 10	P_O , P_M , $1-P_G$, P_S
2	1, 4, 9, 10	P_O , P_M , $1-P_G$, P_S
3	2, 9, 10	P_M , $1-P_L$, P_S
4	3, 9, 10	P_M , $1-P_L$, P_S
5	1, 6, 9, 10	P_O , P_M , $1-P_G$, P_S
6	3, 7, 9, 10	P_O , P_M , $1-P_G$, P_S
7	6, 8, 9, 10	P_O , P_M , $1-P_G$, P_S
8	7, 9, 10	P_O , $1-P_G$, P_L
9	9	1
10	10	1

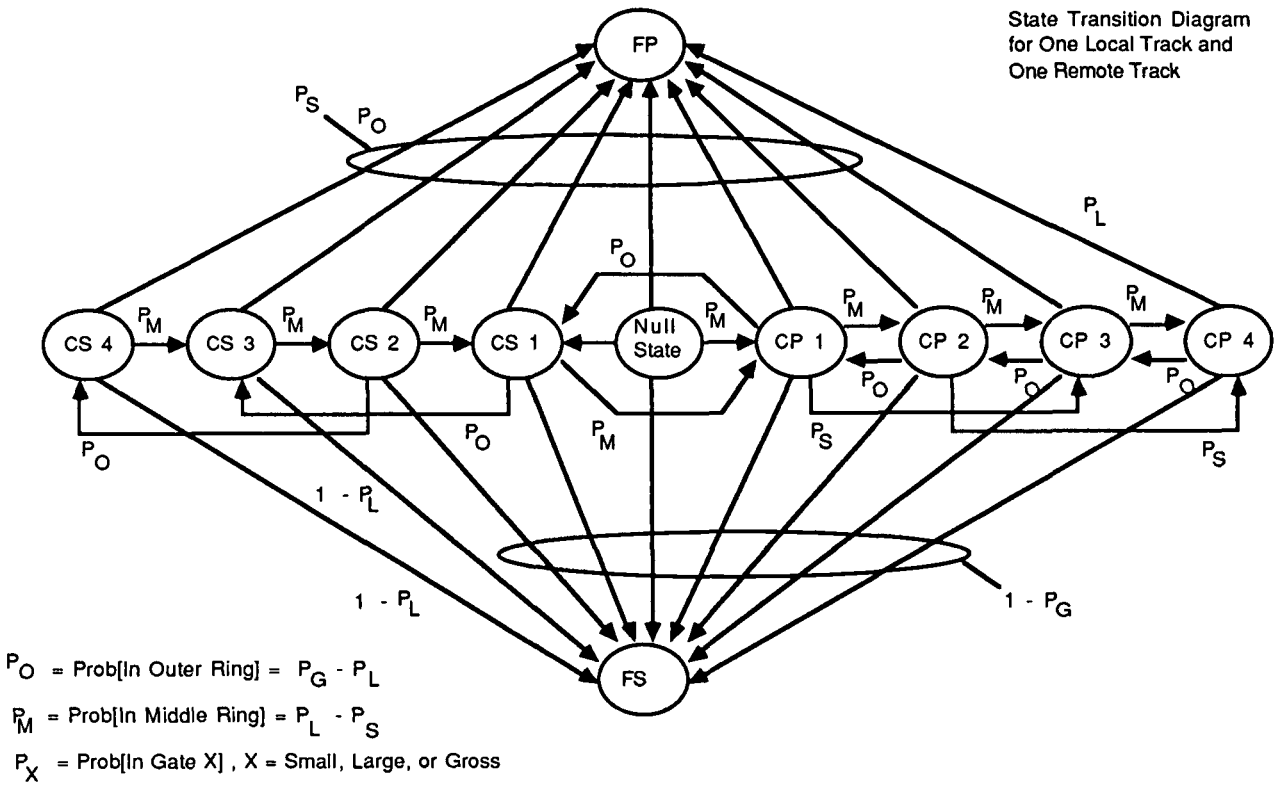


Figure 5.1. State Transition Diagram for Correlation of One Local Track against One Remote System Track. The transitions among the states are defined by the probabilities that the remote track will fall within the specified gates.

The Markov chain model was applied to obtain the probability of "correlation" and "no correlation" decisions as a function of aircraft separation. For the results shown in Figure 5.2, it was assumed that the correlation logic received only one track from each of two subsystems. The techniques discussed by Johnson and Kotz [11] were employed to compute the gate probabilities for the non-central chi-squared distributions which result from off-set normal distributions. The performance of three alternative logics is shown in Figure 5.2; the alternatives are defined by the approximate percentage points of the small and large gates, respectively. The 90/50 Gate Logic was chosen for the FAAD application since it provides the best match to the system requirements. The 50 percent gate corresponds approximately to the acceptance criterion of 4.4 for the null hypothesis in the sequential ratio test. All three of the logics used a Gross Gate size of 29.0, which corresponds to the criterion for rejection of the null hypothesis.

Since the sequential logic may require multiple observations in order to reach a firm decision, the expected or average number of observations required is a performance parameter of interest. The expected number and standard deviation of the number of transitions is provided in Figure 5.3 as a function again of the normalized aircraft separation. As will be noted from Figure 5.3, the standard deviation of the decision time is approximately equal to the average time which indicates the possibility of a

wide variation in actual decision times, particularly at a normalized separation approximately equal to the small gate size. On the other hand, for separations greater than the gross gate size, the logic reduces essentially to a single trial logic.

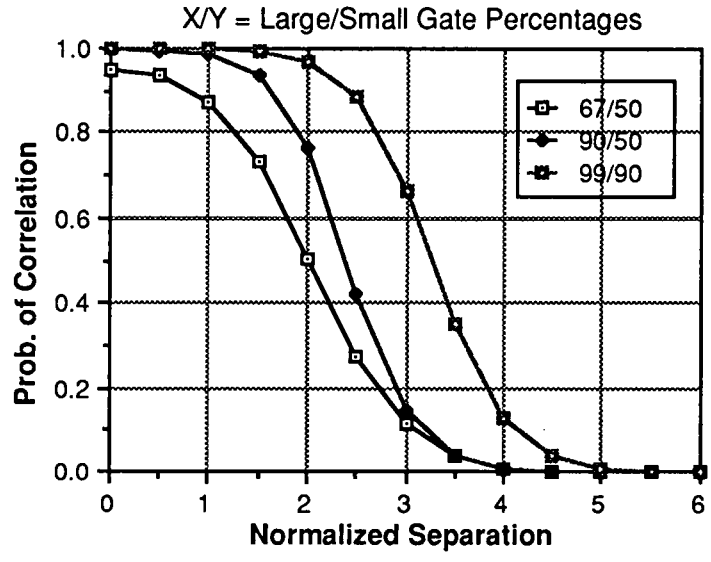


Figure 5.2: Probability of Correlation versus Aircraft Separation. In the Sequential Logic, the Large Gate can be used to control the missed correlation rate while the Small Gate can be used to control the false correlation rate.

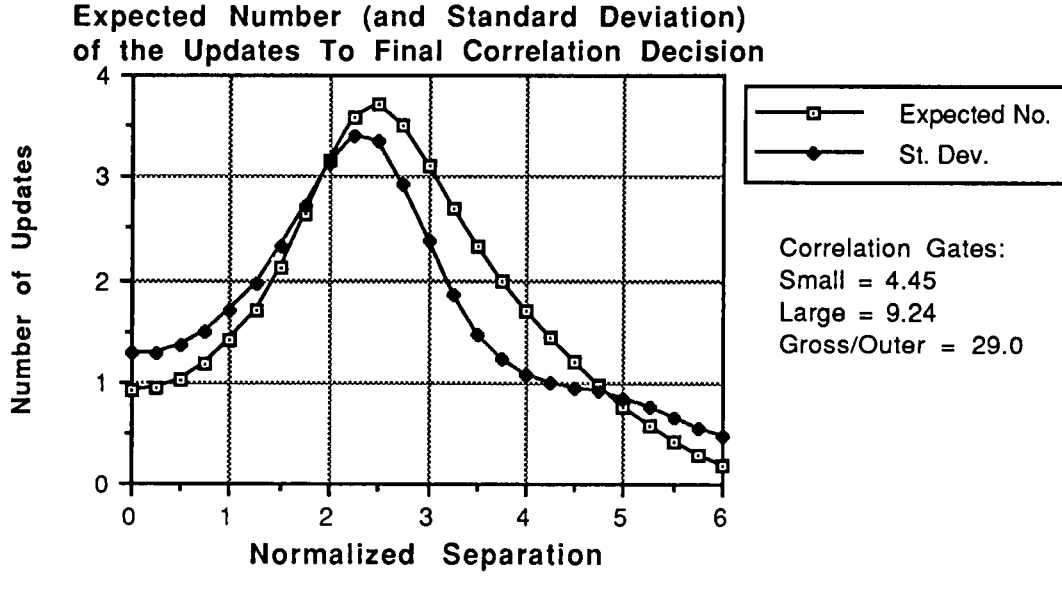


Figure 5.3: Expected Number of Transitions for a Firm Decision versus Normalized Separation. The longest decision times occur when the separation is equal to the size of the small gate.

The normal separation in Figure 5.2 can be considered in several different ways. Obviously, it can be considered as distance in the horizontal plane or the velocity (or phase) plane. It may also be considered as a fixed bias between the two tracking subsystems. For

example, if there is a 200 meter error in the knowledge of the relation positions of the two radars, then there will be an equivalent (in the FAAD case, 1 unit of normalized separation) offset in the reported track positions of a common aircraft. Similarly, a relative azimuth bias at one of the two radars will result in an apparent separation for an aircraft at given range from the biased radar. The required probability of 0.004 of missed correlations occurs for the 90/50 logic at a normalized separation of 0.5 or 100 meters. Thus, any combination of registration errors which yield a separation of less than 100 meters will be acceptable.

Now consider the performance of the sequential correlation logic versus the single trial logic discussed in Section 3. Figure 5.4 shows the performance of the 90/50 sequential logic versus the two single trial logics shown in Figure 3.1. The sequential logic yields a missed correlation error rate of 0.0025 versus a requirement of 0.004. In addition, the false track error rate for the sequential logic is 0.04 at 700 meters of separation in the horizontal plane versus the required 0.10 rate. The 0.10 false correlation rate occurs at approximately 3.15 units of normalized separation or 640 meters.

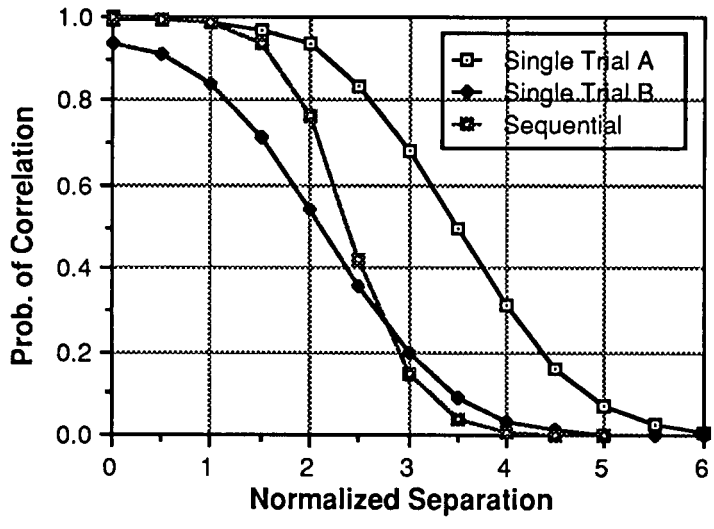


Figure 5.4: Probability of False Correlation versus Aircraft Separation. A sequential correlation decision logic can achieve simultaneously the required error rates for both false and missed correlations.

Finally, consider the situation in which one local track is to be correlated against two remote system tracks, one of which represents the same aircraft as the local track. This perhaps represents a more realistic situation than the previous cases in which the local and remote tracks represented two distinct aircraft since the ability of radars to resolve individual aircraft is a function of both range and the separation of the aircraft in the radar measurement coordinate system.

In order to evaluate performance in this case, the Markov chain model was expanded to include 16 conditional correlation states (rather than 4) and two additional absorbing states to cover the four possible results of the sequential logic. The four absorbing states are (1) no correlation, (2) correlation with a correct remote track, (3) correlation with the incorrect remote track, and (4) correlation with both remote tracks, which is the ambiguous decision case. The results of the analysis are shown in

Figure 5.5 for the cases of a perfect ambiguity resolution and resolution by a random selection, which together bound the nearest neighbor decision criterion[3]. Note that the probability of correlation shown in Figure 5.5 is the probability of correlation with the incorrect remote track, that is, case (3) and case (4) with an incorrect ambiguity resolution.

6. IMPLEMENTATION ISSUES

In order to design and implement a track correlation logic, it is necessary to describe quantitatively the sources of track error at the correlation processor, including both the random and the systematic error components in the tracks. In particular, the correlation statistic ξ defined in Equation 3.8 requires knowledge of the track covariance data for both tracks, one of which could be a track from a remote system received via a data link. Unfortunately, the standard 80-bit track message formats used on tactical data links allow only 2 to 4 bits for track quality; thus, the covariance matrix for remote tracks must be reconstructed in order to apply the standard statistical likelihood and chi-squared correlation techniques discussed in Section 3.

Reconstruction of the track covariance matrix at a remote site is, at best, only a process of approximation based on a set of assumptions about the remote system. An example of one set of assumptions is the following:

1. The Track Quality TQ in the track message corresponds to an upper bound for the standard deviation σ_{TQ} of the error in the predicted track position at the anticipated time of the next update.
2. The update rate Δt and the standard deviation σ_M of the measurement error of the remote sensor are known, at least approximately. The value of the standard deviation should be an upper bound for the errors in the coordinate system assumed in the message format rather than the actual measurement variables (e.g., range and azimuth).
3. The remote system will process the sensor measurements with some form of a linear least-squares estimation technique, for example, a Kalman filter.

The third assumption provides an approximate representation for the upper bounds of the track estimation errors for the worst-case component. Let the vector X represent one track dimension, that is position and velocity (actually, only the rate of change of x); that is $X = [x \quad \dot{x}]^t$. By assumption, X is the result of a linear least-squares estimation process; therefore the covariance matrix Σ for X (that is, the *predicted* state) has the form [1, 9]

$$\Sigma = \sigma_M^2 \begin{bmatrix} \dfrac{4n+2}{n(n-1)} & \dfrac{6}{n(n-1)(\Delta t)} \\ \dfrac{6}{n(n-1)(\Delta t)} & \dfrac{12}{n\left(n^2-1\right)(\Delta t)^2} \end{bmatrix} \tag{6.1}$$

where n is the number of measurements which were used and Δt is the approximate update rate (which may be increased from the sensor scan interval to account for a probability of detection of less than unity, if appropriate).

[3] The ambiguous case could be analyzed more precisely by use of the non-central F-distribution to approximate the probability that the correct remote track is closer to the local track than the incorrect remote track. The solution would only be approximate since the conditional probability distribution given the results of the gate checks is essentially an unsolved problem in mathematical statistics. However, the results should be sufficiently accurate to support any engineering conclusion or decision which is likely to be required.

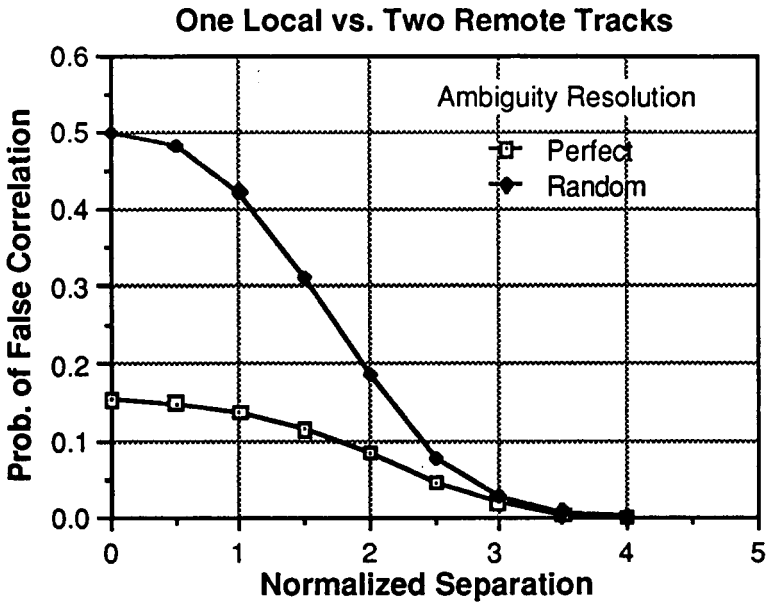


Figure 5.5: Probability of An Incorrect Correlation Decision. If one of the two remote tracks represents the same aircraft as the local track, then a correct correlation decision will occur with probability of at least 0.90 when the normalized separation greater than 2.5 units.

By the second assumption, the bound for the position variance is known; thus,

$$\sigma_{TQ}^2 = \sigma_M^2 \left(\frac{4n+2}{n(n-1)} \right) = \sigma_M^2 \, \rho \qquad (6.2)$$

Since both of the variables σ_{TQ} and σ_M are known, Equation 6.2 can be solved for the variable n; specifically,

$$n = \frac{(\rho+4) + \sqrt{\rho^2+16}}{2\rho} \qquad (6.3)$$

Given n, then the other elements of the matrix in Equation 6.1 can be computed. Finally, since this matrix represents the upper bound or worst-case for the covariance of the individual components of the track, the full track covariance can be represented as the block-diagonal matrix with each block given by Equation 6.1. Note that it may be necessary to permute the rows and columns of the block-diagonal matrix in order to match the order of the variables in the actual track vectors.

7. CONCLUSIONS

This paper has defined a method for track management in distributed, multiple sensor tracking systems, together with an analytic technique for performance evaluation with respect to random, systematic and system errors. One critical analysis remains open for a complete design and implementation approach, namely the registration requirements for participants in a distributed surveillance system. However, the analysis techniques derived in this paper can be applied in a straightforward manner to obtain the upper bound for the registration error within a system. The decomposition of the total error into the component error sources must be based on knowledge of the specific sensors involved; however, the same techniques can be applied to study the impact of the individual component errors on system performance. Finally, the final error budget for system registration must be a subjective compromise between the analysis results and the practical limits of the

data processing necessary to align the system sensors.

8. REFERENCES

[1] Bar-Shalom, Y., and T. E. Fortmann, *Tracking and Data Association*, Academic Press, San Diego, CA (1988).

[2] Waltz, Edward and James Llinas, *Multisensor Data Fusion*, Artech House, Inc., Boston, MA (1990).

[3] Anderson, W. T., *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York (1958).

[4] Ghosh, B. K., Sequential Tests of Statistical Hypotheses, Addison-Wesley Publishing Co., Reading, MA (1970).

[5] Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics*, 4th Ed., Macmillan, New York (1978).

[6] Lehman, E. L.: *Testing Statistical Hypotheses*, 2nd Ed., John Wiley & Sons, New York (1986).

[7] Mardia, K. V., J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, London (1979).

[8] Wald, A.: *Sequential Analysis*, Dover Publications, Inc., New York (1947).

[9] Blackman, S. S, *Multiple-Target Tracking with Radar applications*, Artech House, Inc., Boston, MA (1986).

[10] Anderson, B. D. O. and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Inc., Englewood Cliffs, NJ, (1979).

[11] Johnson, N. L. and S. Kotz, *Continuous Univariate Distributions-2*, John Wiley & Sons, New York (1970).

[12] Press, W. H., et. al., *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, (1986).

[13] Selected Tables in Mathematical Statistics, Vol. 1, American Mathematical Society, Providence, RI (1973).

[14] Kemeny, J. G and J. L. Snell, *Finite Markov Chains*, Springer-Verlag, Inc., New York (1976).

SECTION III

PIXEL AND SYMBOL LEVEL IMAGE FUSION, TARGET CLASSIFICATION AND RECOGNITION

INTRODUCTION

by
Dr. Carlos A. Garriga
SENER, Igeniería y Sistemas, S.A.
Aerospace Division
Parque Technologico de Madrid
c/Severo Ochoa s/n
28760 Tres Cantos (Madrid), Spain

This Section provides an overview of the fundamental and emerging techniques that are presently under research and development for their utilization in advanced real-time Automatic Target Recognition (ATR) Systems.

Image fusion, image segmentation and object classification, in conjunction with Artificial Neural Networks, constitute the basic technologies for real-time recognizers.

Low-observability conditions and the non-cooperative characteristics of military targets make multi-sensor and multi-frequency imagery data fusion very useful for target detection and recognition. The fused image obtained from multi-spectral observation of a given scenario contains most of the signatures or information of the potential targets which are present, in the range of frequencies of the observation detectors. This fact contributes positively to the detection, identification and recognition of such targets in adverse environment.

Image segmentation techniques are used to extract objects from a particular scenario. They are especially efficient for the fusion of multi-spectral imagery, since the amount of information contained in this type of imagery is much higher than in conventional ones.

Classification is the subsequent process to support object extraction by segmentation which contributes directly to the recognition of the targets. In segmentation, the utilization of fusion techniques improves the characterization of targets and, therefore, the efficiency of the classification process. Real-time processing is essential for applications in Defence Surveillance, Detection, Identification, Tracking and Guidance Systems. Neural Networks, because their architectures inherently have fine grain parallelization properties, will hopefully, provide the required real-time capabilities. On the other hand, classical numerical algorithms, due to their own direct applicability, are important for the selection,

development, training and verification of the performance of Neural Networks.

This Section presents different types of classification algorithms, Bayesian Belief Networks, and Neural Networks covering the complete Automatic Target Recognition process, including fusion, segmentation and classification, that are very promising for real-time, or quasi-real-time systems applications.

The Editor of this Section wants to thank and congratulate the authors for putting together many new ideas and expositions of techniques in the field of Automatic Recognition Systems for the near future.

CLASSIFICATION PAR FUSION DE DONNES INCERTAINES MULTI-SENSEURS

Alain APPRIOU
Chef de la Division Traitement de Données
ONERA
BP 72
92322 Châtillon Cedex
France

1. RESUME

Une analyse multi-senseurs visant à reconnaître une situation parmi un certain nombre d'hypothèses répertoriées a priori, pour l'identification d'une cible par exemple, doit être capable de tirer le meilleur parti de tout apprentissage préalable disponible. Les méthodes classiques de traitement, notamment probabilistes, restent toutefois limitées par le manque de représentativité de ces apprentissages lorsque les conditions d'observations évoluent de façon mal maîtrisée. Il convient alors de recourir aux techniques de l'incertain, et en particulier à la Théorie de l'Evidence qui procure le cadre le plus ouvert dans ce contexte. Sa mise en œuvre conduit cependant à un délicat problème d'interprétation des informations disponibles, compte tenu des notions qu'elle permet de manipuler. C'est donc une recherche axiomatique des solutions les plus pertinentes qui est proposée ici pour deux types de problèmes complémentaires, fédérés par un concept générique de traitement propre à gérer la fiabilité des données délivrées. L'intérêt de cette approche est évalué sur la base de quelques simulations simples.

2. BESOINS ET FORMULATION DU PROBLEME

L'intérêt majeur qui justifie l'association de senseurs multiples tient au bénéfice qui peut être tiré de leur complémentarité [3]. La vocation d'une telle association est donc avant tout de traiter des configurations où une partie des senseurs est plus ou moins en défaut (aptitudes pour la situation à traiter, environnement, contre-mesures, défaillances,...). Un traitement judicieux de ce type de configuration exige la prise en compte de toutes les informations disponibles, notamment celles susceptibles de nous renseigner sur la qualité des mesures, sur leur potentiel informatif, sur le contexte des relevés, et donc sur la pertinence relative des différentes évaluations qui peuvent être menées.

Ceci conduit à inférer dans un même système des données de nature particulièrement disparate, et souvent subjectives, incertaines, incomplètes, erronées, imprécises, avec des relations de dépendance très spécifiques. On devra typiquement intégrer le traitement conjoint de mesures, d'apprentissages plus ou moins représentatifs, et d'informations qualitatives sur ces données. Les théories de l'incertain procurent un cadre fédérateur séduisant dans ce contexte, mais leur mise en œuvre pratique se heurte à un certain nombre de difficultés : interprétation et modélisation des informations disponibles dans les cadres théoriques appropriés, choix d'une architecture de fusion et de règles de combinaison, principes de décision à adopter, contraintes sur la rapidité et le volume des calculs nécessaires.

Considérons plus précisément ici un système multi-senseurs conçu pour reconnaître une situation scrutée parmi un ensemble exhaustif de N hypothèses H_i répertoriées a priori ($i \in [1, N]$). Les applications concernées couvrent ainsi un panorama relativement vaste de fonctions, les hypothèses en question pouvant être, par exemple, la présence d'entités (détection, extraction), l'identité de cibles (veille, conduite de tir, autodirecteur,...) ou d'amers (recalage de la navigation d'engins), la localisation d'un vecteur ou d'une cible, l'état d'un système ou d'une situation (dommages, configuration,...). La finalité d'un tel système peut par ailleurs être le simple renseignement, ou l'intégration à un système plus global de filtrage (poursuite de cibles, recalage de navigation,...) ou d'aide à la décision en vue de la mise en œuvre de moyens d'analyse, de guerre électronique, d'intervention,...

Les senseurs sont par ailleurs réputés choisis et définis en vue d'assurer la meilleure complémentarité utile au problème traité, leurs spécificités important peu pour ce qui suit. Typiquement, ils peuvent être de type radar, imageur infrarouge, radiomètre, ESM, vidéo, acoustique, sismique, sonar. Chacun des M senseurs S_j ($j \in [1, M]$) utilisés est en outre supposé doté des traitements propres à extraire des signaux ou des images qu'il procure une mesure ou un ensemble de mesures m_j pertinent pour la fonction envisagée.

On se propose de considérer dans un premier temps le cas le plus général où chaque mesure m_j permet d'élaborer, sur la base d'un apprentissage préalable quelconque, N critères C_{ij} à valeurs dans [0, 1], propres à caractériser respectivement la vraisemblance de chaque hypothèse H_i. Un facteur de qualité q_{ij} à valeurs dans [0, 1] est en outre associé à chaque vraisemblance C_{ij}. Sa vocation est de traduire, à partir d'un apprentissage ou de connaissances exogènes, l'aptitude du critère C_{ij} à discriminer l'hypothèse H_i dans les conditions de l'observation. Il intègre notamment la confiance que l'on peut avoir dans la représentativité de l'apprentissage utilisé pour l'élaboration de C_{ij}, compte tenu de l'évolution du contexte, de la qualité, du volume, et de l'exactitude des données accessibles lors de l'apprentissage, et de la pollution éventuelle des relevés.

Deux types de problèmes particuliers d'intérêt sont ensuite développés, et fédérés par le problème générique qui vient d'être introduit. Ils correspondent, pour l'un à une élaboration particulière des facteurs de confiance q_{ij} à partir de matrices de confusion apprises pour chaque senseur, et pour l'autre à l'utilisation d'un apprentissage stochastique dans la détermination des vraisemblances C_{ij}. L'apport de l'approche préconisée est mis en évidence dans chaque cas sur la base de quelques simulations simples de synthèse.

La démarche adoptée consiste en une recherche axiomatique des solutions répondant aux différents problèmes posés. Le cadre retenu est a priori celui de la théorie de l'évidence, qui s'avère être le plus large et le mieux adapté à l'interprétation des données considérées (cf § 3). Les résultats obtenus sont néanmoins comparés aux solutions accessibles par d'autres voies de référence.

L'exposé qui suit présente une synthèse de travaux déjà partiellement discutés dans [1, 2, 4].

3. RAPPELS SUR LA THEORIE DE L'EVIDENCE

Quelques bases nécessaires à la compréhension de la suite de l'exposé sont ici rassemblées de façon informelle. Elles mettent notamment en évidence certains outils de la théorie de l'évidence particulièrement utiles en fusion de données.

3.1 Notions fondamentales

Cette théorie, développée dans [5], suppose la définition préalable d'un ensemble E de N éléments H_i ($i \in [1, N]$) exclusifs et exhaustifs, appelé cadre de discernement. 2^E désigne alors l'ensemble des $2^N - 1$ sous-ensembles A_j de E ($j \in [1, 2^N-1]$).

Une fonction de masse élémentaire m(.) est définie de 2^E sur [0, 1] par :

$$m(\emptyset) = 0 \tag{3.1}$$

$$\sum_{j=1}^{2^N-1} m(A_j) = 1 \tag{3.2}$$

Les éléments focaux sont les éléments A_j de 2^E dont la masse $m(A_j)$ est non nulle. Lorsque ces éléments focaux se réduisent aux seuls singletons

H_i, ou plus généralement à une partition de E, la notion de masse élémentaire est assimilable à celle de probabilité. L'apport de la théorie de l'évidence est donc typiquement de permettre l'évaluation conjointe d'ensembles quelconques de ces singletons H_i. Dès lors les événements considérés ne sont plus nécessairement exclusifs. Une masse $m(A_j)$ est représentative de la vraisemblance attribuable à l'un des éléments du sous-ensemble A_j, sans aucun discernement possible entre les différents éléments de A_j. En particulier m(E) désigne le degré d'incertitude totale.

Une fonction de crédibilité Cr(.) peut également être définie sur les mêmes ensembles par :

$$Cr(\emptyset) = 0 \tag{3.3}$$

$$Cr(E) = 1 \tag{3.4}$$

$$Cr(\bigcup_{j\in J} A_j) \geq \sum_{\substack{I\subset J \\ I\neq\emptyset}} (-1)^{|I|+1} Cr(\bigcap_{j\in I} A_j) \tag{3.5}$$

où l'on peut remarquer que la fonction de probabilité est un cas particulier de fonction de crédibilité, obtenue à l'égalité de (3.5).

Les fonctions de masse élémentaire et de crédibilité sont donc définies et utilisable de façon indépendante. Il existe cependant une bijection entre l'ensemble des fonctions de masse élémentaire et l'ensemble des fonctions de crédibilité, qui associe à chaque jeu de masses sur 2^E un jeu de crédibilités sur le même ensemble. Cette correspondance est formalisée par les relations :

$$Cr(B_k) = \sum_{A_j\subset B_k} m(A_j) \tag{3.6}$$

$$m(A_j) = \sum_{B_k\subset A_j} (-1)^{|A_j-B_k|} Cr(B_k) \tag{3.7}$$

La notion de fonction de plausibilité peut alors être introduite indifféremment en liaison avec la fonction de crédibilité :

$$Pl(B_k) = 1 - Cr(\neg B_k) \tag{3.8}$$

$\neg B_k$ désignant le complément de B_k dans E, ou à partir de la fonction de masse élémentaire :

$$Pl(B_k) = \sum_{A_j\cap B_k\neq\emptyset} m(A_j) \tag{3.9}$$

De façon intuitive, la crédibilité peut être interprétée comme une mesure de vraisemblance minimale d'un événement, et la plausibilité comme une mesure de vraisemblance maximale. La prise en compte d'informations concrètes dans un système peut donc se faire soit en termes de crédibilités et de plausibilités, soit plus directement sous forme de masses élémentaires.

Parmi les diverses notions manipulées par la théorie de l'évidence, signalons encore la fonction de communalité, dont l'intérêt pratique sera souligné dans la suite :

$$Q(B_k) = \sum_{A_j\supset B_k} m(A_j) \tag{3.10}$$

3.2 Combinaison de sources distinctes

L'intérêt majeur de la théorie de l'évidence en fusion de données repose sur la possibilité de construire une fonction de masse élémentaire m(.) unique, par sommation orthogonale des M fonctions de masse élémentaire $m_j(.)$ issues de M sources d'information S_j distinctes ($j\in[1, M]$), définies sur le même cadre de discernement E :

$$m(.) = m_1(.) \oplus m_2(.) \oplus \ldots \oplus m_M(.) \tag{3.11}$$

Cet opérateur a été proposé par Shafer pour réaliser la conjonction d'avis en respectant (3.1). D'autres opérateurs ont été envisagés, notamment propres à ignorer cet axiome (open world), ou relevant d'une logique de disjonction. La somme orthogonale, qui a pu être justifiée plus récemment à partir d'approches axiomatiques complexes, reste néanmoins l'outil le plus pertinent pour fusionner des sources ne présentant pas un désaccord important quant aux croyances exprimées. Elle consiste à calculer :

$$m(A) = (1-K)^{-1} * \sum_{A^1\cap A^2\cap\ldots\cap A^M=A\neq\emptyset} \{ \prod_{j=1}^{M} [m_j(A_j)] \} \tag{3.12}$$

où A_j désigne un sous-ensemble quelconque du cadre de discernement commun E, évalué par la source S_j. K est l'inconsistance de la fusion, propre à figurer le degré de contradiction entre les croyances exprimées par les différentes sources mises en jeu :

$$K = \sum_{A^1\cap A^2\cap\ldots\cap A^M=\emptyset} \{ \prod_{j=1}^{M} [m_j(A_j)] \} \tag{3.13}$$

Une telle loi de combinaison d'informations distinctes procure un certain nombre de propriétés primordiales en fusion de données multi-senseurs, notamment :

* la commutativité,

* l'associativité

* une mise en œuvre simple à partir des fonctions de communalité Q_j de chaque source S_j :

$$Q(A) = (1-K)^{-1} * \prod_{j=1}^{M} [Q_j(A)] \tag{3.14}$$

avec :

$$1-K = \sum_{B\neq\emptyset} (-1)^{|B|+1} * \prod_{j=1}^{M} [Q_j(B)] \tag{3.15}$$

3.3 Conditionnement

Une crédibilité conditionnelle Cr(.|A), relativement à un événement A, est la somme orthogonale d'une crédibilité quelconque Cr(.) et d'une crédibilité certaine $Cr_A(.)$ définie par le jeu de masses :

$$m_A(A) = 1 \tag{3.16}$$

et donc:

$$m_A(B) = 0 \ , \ \forall B\neq A \tag{3.17}$$

Autrement dit:

$$Cr(.|A) = Cr(.) \oplus Cr_A(.) \tag{3.18}$$

Les crédibilités et les plausibilités conditionnelles obéissent alors aux lois de composition suivantes :

$$Cr(B|A) = \{ Cr(B\cup\neg A) - Cr(\neg A) \} / \{ 1 - Cr(\neg A) \} \tag{3.19}$$

$$Pl(B|A) = Pl(B\cap A) / Pl(A) \tag{3.20}$$

Les crédibilités conditionnelles satisfont par ailleurs la propriété de distributivité du conditionnement par rapport à la somme orthogonale, ce qui rend en pratique indifférent l'ordre dans lequel on effectue ces deux opérations :

$$Cr^{\circ}\oplus Cr^{\circ\circ}(.|A) = Cr^{\circ}(.|A) \oplus Cr^{\circ\circ}(.|A) \tag{3.21}$$

Il convient également de noter qu'en présence d'un jeu de masse réduit à une distribution de probabilités, c'est-à-dire pour des éléments focaux formant une partition du cadre de discernement E (cf § 3.1), (3.20) assure la parfaite cohérence de la théorie de l'évidence avec l'inférence bayésienne.

3.4 Affaiblissement

L'affaiblissement $Cr^a(.)$ d'une crédibilité Cr(.) dans le rapport a ($a \in [0, 1]$) permet de réduire globalement le degré de certitude délivré par la source d'information concernée, en fonction de la confiance que l'on est amené à lui accorder par ailleurs :

$$Cr^a(E) = 1 \tag{3.22}$$

$$Cr^a(A) = (1-a) * Cr(A) \ , \quad \forall\ A \neq E \tag{3.23}$$

Ceci revient à modifier le jeu de masses correspondant selon la transformation :

$$m^a(A) = (1-a) * m(A) \ , \quad \forall\ A \neq E \tag{3.24}$$

$$m^a(E) = a + (1-a) * m(E) \tag{3.25}$$

3.5 Raffinement - Grossissement

En complément du conditionnement, deux opérations sont plus spécialement dédiées à la gestion des cadres de discernement. Elles constituent une composante vitale en fusion de données, car la somme orthogonale ne permet de fusionner que des fonctions de masse élémentaire définies sur un même cadre de discernement, alors que les sources d'information fournissent le plus souvent des évaluations sur des cadres de discernement différents.

Soit R une application d'un cadre de discernement E^1 dans un cadre de discernement E^2 :

$$E^1 = \{H_1^1, \dots, H_{N1}^1\} \xrightarrow{R} E^2 = \{H_1^2, \dots, H_{N2}^2\} \ , \qquad N2 > N1 \tag{3.26}$$

R est un raffinement de E^1 dans E^2 si $\{R(H_1^1), \dots, R(H_{N1}^1)\}$ constitue une partition de E^2. Un jeu de masses $m^2(.)$ peut alors être déterminé sur E^2 comme l'extension minimale d'un jeu de masses $m^1(.)$ donné sur E^1, par :

$$m^2(R(B)) = m^1(B) \tag{3.27}$$

pour tout élément focal B de $m^1(.)$, $m^2(.)$ étant nul ailleurs.

Il existe alors une relation inverse R^{-1}, dite de grossissement. Elle permet de déterminer un jeu de masses $m^1(.)$ sur E^1 à partir d'un jeu de masses $m^2(.)$ donné sur E^2, à l'aide de :

$$m^1(A) = \sum_{\substack{B \subset E^2 \\ A = \{H_i^1 / R(H_i^1) \cap B \neq \emptyset\}}} m^2(B) \tag{3.28}$$

3.6 Propriétés de crédibilité particulières

3.6.1 *Crédibilité bayésienne*

Les éléments focaux sont ici réduits aux seuls singletons H_i du cadre de discernement E :

$$\forall\ A_j \neq H_i,\ i \in [1, N], \quad m(A_j) = 0 \tag{3.29}$$

Dès lors les notions de crédibilité, de plausibilité, et de probabilité sont confondues :

$$\forall\ A_j \subset E, \quad Cr(A_j) = Pl(A_j) = P(A_j) = \sum_{H_i \in A_j} m(H_i) \tag{3.30}$$

D'autre part la combinaison d'un jeu de masses quelconque m(.) avec un jeu de masses bayésien $m^\circ(.)$ procure un jeu de masses bayésien $m^{\circ\circ}(.)$, différent de $m^\circ(.)$, et fourni par:

$$m^{\circ\circ}(H_i) = \{ m^\circ(H_i) * Pl(H_i) \} / \{ \sum_{k=1}^{N} m^\circ(H_k) * Pl(H_k) \} \tag{3.31}$$

où Pl(.) est la fonction de plausibilité associée au jeu de masses m(.).

3.6.2 *Crédibilité consonante*

Elle est définie par :

$$Cr(\emptyset) = 0 \tag{3.32}$$

$$Cr(E) = 1 \tag{3.33}$$

$$Cr(A \cap B) = \min \{ Cr(A), Cr(B) \} \tag{3.34}$$

On montre qu'ainsi ses éléments focaux sont emboîtés les uns dans les autres. Si ces derniers sont étiquetés de telle sorte que $A_1 \subset A_2 \subset \dots \subset A_K$ ($K \leq N$), leurs masses satisfont :

$$m(A_K) = Pl(A_K - A_{K-1}) \tag{3.35}$$

$$m(A_j) = Pl(A_j - A_{j-1}) - Pl(A_{j+1} - A_j) \ , \ \forall\ A_j \neq A_K \tag{3.36}$$

et inversement :

$$Pl(A_j - A_{j-1}) = \sum_{l=j}^{K} m(A_l) \tag{3.37}$$

La fonction de plausibilité associée à une crédibilité consonante vérifie par ailleurs la relation duale de (3.34) :

$$Pl(A \cup B) = \max \{ Pl(A), Pl(B) \} \tag{3.38}$$

De plus, tout événement A_j a nécessairement, dans ces conditions, soit sa crédibilité nulle, soit sa plausibilité égale à 1 ; autrement dit, la consonance impose à un événement d'être parfaitement plausible avant d'être crédible :

$$\forall\ A_j \subset E, \quad Cr(A_j) = 0 \ \text{ ou } \ Cr(\neg A_j) = 0 \tag{3.39}$$

Il convient de noter que les crédibilité et plausibilité consonantes sont rigoureusement équivalentes aux notions de nécessité et de possibilité de la théorie des possibilités. Au niveau de leur formalisme ces dernières peuvent donc n'être considérées que comme des cas particuliers de crédibilité et de plausibilité.

4. PRISE DE DECISION : CHOIX DE L'HYPOTHESE LA PLUS VRAISEMBLABLE

On souhaitera le plus souvent désigner l'hypothèse H_i^* la plus vraisemblable au vu de l'information élaborée. Une telle prise de décision, immédiate lorsqu'il est possible d'associer une probabilité a posteriori à chaque hypothèse, devient particulièrement délicate lorsque les évaluations sont présentées en termes de jeu de masses de la théorie de l'évidence. Toute la difficulté est alors liée à la non-exclusivité des évaluations, qui pose le problème pratique de l'interprétation et de la prise en compte relative des masses attachées aux éléments focaux de cardinal supérieur ou égal à 2, dans la désignation d'un singleton unique. Ce problème, général à la théorie de l'évidence et incontournable dans le contexte traité ici, ne donne lieu à ce jour qu'à des solutions intuitives plus ou moins satisfaisantes dans l'absolu.

Il est donc proposé dans la suite trois approches globales différentes du problème général de choix de l'hypothèse H_i* la plus vraisemblable, étant donné un jeu de masses m(.) quelconque sur un cadre de discernement $E=\{H_1,\dots, H_N\}$, lorsqu'aucun autre a priori discriminatoire entre les Hi n'est retenu.

Une synthèse des procédures dégagées visera ensuite à retenir une attitude unifiée face au problème.

4.1 Approche globale "focalisée"

Cette approche consiste à considérer N jeux de masses $m_i(.)$ certains ($i\in[1, N]$), en accord avec la définition (3.16) et (3.17), chacun étant respectivement focalisé sur une des N hypothèses H_i du cadre de discernement E. L'inconsistance K_i de la somme orthogonale entre le jeu de masses $m_i(.)$ et le jeu de masses m(.) disponible est alors représentative de leur désaccord, c'est-à-dire du conflit qui oppose l'évaluation m(.) à la certitude qu'il s'agisse de l'hypothèse H_i. L'hypothèse H_i* choisie doit donc être, dans ces conditions, celle qui correspond à une inconsistance K_i minimale. K_i étant en pratique donnée par :

$$Ki = 1 - Pl(Hi) \tag{4.1}$$

où Pl(.) est la fonction de plausibilité associée à m(.), c'est l'hypothèse pour laquelle cette plausibilité est maximale qu'il convient de retenir.

L'intérêt de ce critère d'inconsistance est en particulier conforté par la notion d'entropie qui peut être attachée à son expression [6].

4.2 Approche globale "bayésienne"

L'idée est ici de prendre en compte la donnée préalable d'un jeu de masses bayésien "équiprobable" $m_b(.)$ sur le cadre de discernement E :

$$m_b(H_i) = 1 / N \quad , \quad \forall\ i\in[1, N] \tag{4.2}$$

et donc:

$$m_b(A) = 0 \quad , \quad \forall\ A\neq H_i \quad , \quad i\in[1, N] \tag{4.3}$$

Confiant à ce jeu de masses $m_b(.)$ un rôle similaire à celui de probabilités a priori équiprobables dans l'inférence bayésienne, un jeu de masses $m_c(.)$ peut être déterminé par sommation orthogonale du jeu de masses $m_b(.)$ et du jeu de masses m(.) disponible. En vertu de la propriété (3.31), $m_c(.)$ est alors un jeu de masses bayésien défini par :

$$m_c(H_i) = Pl(H_i) / \{ \sum_{k\in[1,N]} Pl(H_k) \} \tag{4.4}$$

et donc :

$$m_c(A) = 0 \quad , \quad \forall\ A\neq H_i \quad , \quad i\in[1, N] \tag{4.5}$$

où Pl(.) est la fonction de plausibilité associée à m(.). Par référence au maximum de probabilité a posteriori, la procédure de décision consiste alors de façon immédiate à retenir l'hypothèse H_i* de masse maximale, et donc, ici encore, de plausibilité $Pl(H_i^*)$ maximale.

Conceptuellement, le principe de cette approche consiste à substituer à l'incertitude totale a priori une indifférence entre les singletons du cadre de discernement, de façon à forcer la discrimination entre ces seuls éléments.

4.3 Approche par paradigme de décision

La solution est ici cherchée en se référant à un contexte plus général de prise de décision, synthétisé par exemple dans [7]. Le but est alors de choisir une action à mener parmi Q actions a_h possibles ($h\in[1, Q]$), à partir de l'évaluation fournie par le jeu de masses m(.) sur le cadre de discernement E.

Ce choix peut être conduit en maximisant une fonction de coût $C(a_h)$ sur l'ensemble des actions possibles, connaissant le poids $G(a_h/B_k)$ que l'on affecte à chaque action potentielle a_h lorsque l'événement B_k, sous-ensemble quelconque de E, est réalisé :

$$C(a_h) = \sum_{B_k\subset E} \{G(a_h/B_k)*m(B_k)\} \tag{4.6}$$

Toute la difficulté de mise en œuvre pratique et la crédibilité d'une telle procédure restent liées à l'évaluation des poids $G(a_h/B_k)$, le plus souvent très subjective. Si l'on peut en général considérer que les poids relatifs aux seuls singletons H_i de E sont fournis par le système ou l'utilisateur, ceux relatifs aux sous-ensembles B_k de cardinal supérieur ou égal à 2 doivent en revanche faire l'objet d'une détermination intuitive, éventuellement guidée par une "attitude" privilégiée [7].

Toutefois dans notre cas, ce caractère subjectif peut être fortement atténué par la bijection que l'on est amené à établir entre l'ensemble des actions et le cadre de discernement E, chaque action a_i consistant respectivement à déclarer une hypothèse H_i comme vraie (Q=N). En effet, en l'absence de toute information complémentaire, les poids sont alors légitimement donnés par :

$$E(a_i/B_k) = 1 \quad \text{si} \quad H_i\in B_k \tag{4.7}$$

$$E(a_i/B_k) = 0 \quad \text{si} \quad H_i\notin B_k \tag{4.8}$$

de façon à respecter la notion de masse $m(B_k)$ associée, telle qu'introduite par la théorie de l'évidence, c'est-à-dire comme une évaluation d'un des éléments de B_k, sans que l'on soit en mesure de préciser de quel élément de B_k il s'agit.

Dans ces conditions, (4.6) conduit elle aussi à désigner l'hypothèse H_i* de plausibilité $Pl(H_i^*)$ maximale comme la plus vraisemblable.

4.4 Synthèse

Les trois approches globales présentées convergent toutes vers la même procédure de décision, qui consiste à retenir comme la plus vraisemblable l'hypothèse H_i* telle que :

$$Pl(H_i^*) = \max_{i\in[1,N]} \{ Pl(H_i) \} \tag{4.9}$$

5. PROBLEME GENERIQUE

Le problème très général de discrimination introduit au § 2 est considéré dans le cas d'intérêt pratique où les critères C_{ij} sont élaborés par des chaînes d'information distinctes, qui justifient la différenciation de leur pertinence respective par des facteurs q_{ij}. On suppose par ailleurs être dans le contexte le plus fréquent où les critères C_{ij} pris isolément ont toujours au moins valeur de réfutation, dans le sens où leur nullité garantit que l'hypothèse associée H_i n'est pas vérifiée.

Ceci conduit à formaliser le problème sur la base de deux axiomes :

<u>Axiome 5.1</u> : Chacun des N*M couples $[C_{ij}, q_{ij}]$ constitue une source d'information distincte, ayant pour éléments focaux H_i, $\neg H_i$, et E, où le cadre de discernement E représente l'ensemble des N hypothèses.

<u>Axiome 5.2</u> : $C_{ij} = 0$, lorsqu'il est valide ($q_{ij} = 1$), permet d'affirmer que H_i n'est pas vérifiée.

L'axiome 5.1 impose d'élaborer N*M jeux de masses $m_{ij}(.)$, à partir, respectivement, des N*M couples $[C_{ij}, q_{ij}]$. Pour chacun, la masse des éléments focaux, H_i, $\neg H_i$, et E, est dans un premier temps définie par la valeur du critère C_{ij} correspondant, qui ne peut être interprété qu'en termes de crédibilité ou de plausibilité de H_i. L'axiome 5.2 limite alors à 2 le nombre des interprétations admissibles. La première conduit à :

$$Cr_{ij}(H_i) = 0 \quad \text{et} \quad Pl_{ij}(H_i) = C_{ij} \tag{5.1}$$

et la seconde à :

$$Cr_{ij}(H_i) = Pl_{ij}(H_i) = C_{ij} \quad (5.2)$$

La prise en compte du facteur de confiance q_{ij} associé à C_{ij} permet alors d'élaborer le jeu de masses $m_{ij}(.)$ cherché, par affaiblissement dans le rapport $(1-q_{ij})$ (cf § 3.4). Ceci conduit aux deux modèles possibles :

Modèle 1 :

$$m_{ij}(H_i) = 0 \quad (5.3)$$

$$m_{ij}(\neg H_i) = q_{ij}*(1-C_{ij}) \quad (5.4)$$

$$m_{ij}(E) = 1-q_{ij}*(1-C_{ij}) \quad (5.5)$$

Modèle 2 :

$$m_{ij}(Hi) = q_{ij}*C_{ij} \quad (5.6)$$

$$m_{ij}(\neg Hi) = q_{ij}*(1-C_{ij}) \quad (5.7)$$

$$m_{ij}(E) = 1-q_{ij} \quad (5.8)$$

Un jeu de masse m(.) synthétisant l'ensemble des évaluations est ensuite obtenu par sommation orthogonale des différents jeux de masses $m_{ij}(.)$, dans le cadre de chaque modèle. Un critère de maximum de plausibilité, justifié au § 4, permet alors de dégager de m(.) l'hypothèse H_i la plus vraisemblable. Les 2 modèles introduits conduisent ainsi respectivement aux deux solutions :

Solution 1 :

$$\max_i \{ \prod_j [1-q_{ij}*(1-C_{ij})] \} \quad (5.9)$$

Solution 2 :

$$\max_i \{ \prod_j [1-q_{ij}*(1-C_{ij})]/[1-q_{ij}*C_{ij}] \} \quad (5.10)$$

Il convient de noter que la solution 1 répond également à un critère de maximum de crédibilité [4].

Le modèle 1 est par ailleurs consonant et se prête donc, mais lui seul, à une interprétation dans le cadre de la théorie des possibilités (cf § 3.6.2) :

$$N_{ij}(H_i) = 0 \quad (5.11)$$

$$\Pi_{ij}(E) = 1-q_{ij}*(1-C_{ij}) \quad (5.12)$$

Différents opérateurs de conjonction susceptibles d'assurer la combinaison des N*M sources ainsi formalisées sont alors envisageables [8]. Nous retiendrons ici à titre de référence la norme triangulaire idempotente, pour ses propriétés fondamentales qui la démarquent le plus de la somme orthogonale employée dans l'approche par la théorie de l'évidence, pour sa simplicité calculatoire, et pour son emploi le plus répandu en l'absence de choix motivés.

Un critère de maximum de possibilité pour la détermination de l'hypothèse H_i la plus vraisemblable est alors incontournable (nécessité nulle). Il conduit à la solution :

$$\max_i \{ \min_j [1-q_{ij}*(1-C_{ij})] \} \quad (5.13)$$

La détermination pratique des C_{ij} et q_{ij} est dans tous les cas un problème spécifique au type d'application traité. Deux situations différentes par la nature des informations disponibles sont présentées dans la suite. Elles sont représentatives des deux grandes classes de problèmes rencontrées dans les applications d'intérêt traitées à ce jour.

6. CLASSIFICATION AVEC MATRICE DE CONFUSION

Le problème précédent est particularisé en ce que les cœfficients de confiance q_{ij} sont élaborés à partir d'un apprentissage de la matrice de confusion relative à chaque senseur S_j testé isolement. La nature particulière de cette information conduit à considérer que chaque source $[C_{ij}, q_{ij}]$ introduite au § 5 est elle-même la somme orthogonale de N sources $[C_{kj}, q_{kij}]$ $(1 \leq k \leq N)$, strictement de même nature, où :

$$q_{kij} = P[C_{kj}=\max_h\{C_{hj}\}/H_i] \quad (6.1)$$

6.1 Solutions

Les deux solutions dégagées dans le cadre de la théorie de l'évidence s'écrivent alors simplement :

Solution 1 :

$$\max_i \{ \prod_k \prod_j [1-q_{kij}*(1-C_{kj})] \} \quad (6.2)$$

Solution 2 :

$$\max_i \{ \prod_k \prod_j [1-qkij*(1-Ckj)]/[1-qkij*Ckj] \} \quad (6.3)$$

et la solution extrapolée dans le cadre de la théorie des possibilités devient :

Solution 3 :

$$\max_i \{ \min_{k,j} [1-q_{ij}*(1-C_{ij})] \} \quad (6.4)$$

A titre de référence, la solution probabiliste adaptée au problème suppose une décision décentralisée au niveau de chaque senseur, imposée par la prise en compte des matrices de confusion :

Solution 4 :

$$\max_i \{ \prod_j q_{kij} \}, \text{ avec k tel que : } C_{kj} = \max_h \{ C_{hj} \} \quad (6.5)$$

Enfin, l'ignorance de la matrice de confusion réduit classiquement la décision à :

Solution 5 :

$$\max_i \{ \prod_j Cij \} \quad (6.6)$$

6.2 Simulations

Quelques simulations très simples permettent d'appréhender le comportement relatif de ces cinq solutions. Elles mettent toutes en scène la reconnaissance de trois cibles à l'aide de deux senseurs. Pour chaque critère C_{ij}, des valeurs aléatoires sont générées selon une loi uniforme sur $[B_{ij}, B_{ij}+0.30]$ lorsque la cible i est effectivement la cible présentée, et sur [0.35, 0.65] sinon. Les bornes B_{ij} sont choisies de façon à satisfaire les matrices de confusion données pour chaque senseur, dont on notera qu'elles présentent des taux de confusion équirépartis sur les mauvaises déclarations, et donc parfaitement définis par les seul taux de reconnaissance.

Le premier exemple présenté concerne deux senseurs identiques ayant de surcroît le même taux de reconnaissance T_r pour chacune des trois cibles. La figure 1 fournit le taux de reconnaissance moyen obtenu par les différentes méthodes lorsque T_r varie. Le meilleur comportement est obtenu pour la théorie de l'évidence, indifféremment avec l'une ou l'autre des modélisations. Le creux à $T_r \approx 1/3$ correspond à une absence d'information (matrices de confusion uniformes). A noter le bon comportement de la théorie des possibilités sur cet exemple. L'approche probabiliste est en revanche uniformément moins bonne, en dépit d'un apport bénéfique de la connaissance des matrices de confusion individuelles pour les faibles valeurs de T_r. Cette contre performance reste attachée au caractère décentralisé de la décision de classification, qui comprime abusivement l'information avant fusion. Le bon comportement de la solution 5 à T_r élevé, et sa chute à faible T_r, mettent bien en évidence l'intérêt de prendre en compte l'information qualitative donnée par la connaissance des matrices de confusion individuelles.

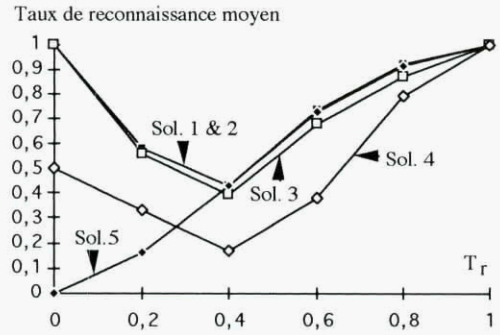


Figure 1 - Classification avec matrices de confusion : 3 cibles - 2 capteurs identiques - Incidence du taux de reconnaissance individuel T_r

Le deuxième exemple, présenté en figure 2, diffère simplement du premier en ce que l'analyse est menée en fonction du nombre M de senseurs, T_r étant fixé à 0,2 pour stigmatiser une situation défavorable. Seules les solutions issues des théories de l'évidence et des possibilités mettent à profit, de façon stable, l'effet statistique cumulatif des capteurs dans un cas aussi critique. La solution 5 montre en particulier que l'adjonction de tout nouveau senseur conduit à dégrader les performances, en l'absence d'informations qualitatives dans l'adversité. L'approche probabiliste subit quant à elle un effet de quantification particulièrement pénalisant compte tenu des matrices de confusion considérées. A noter l'impact de la gestion de l'incertain au niveau d'un seul capteur.

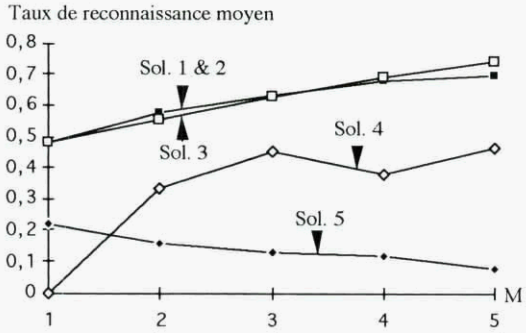


Figure 2 - Classification avec matrices de confusion : 3 cibles - Capteurs identiques - $T_r = 0,2$ - Incidence du nombre M de capteurs

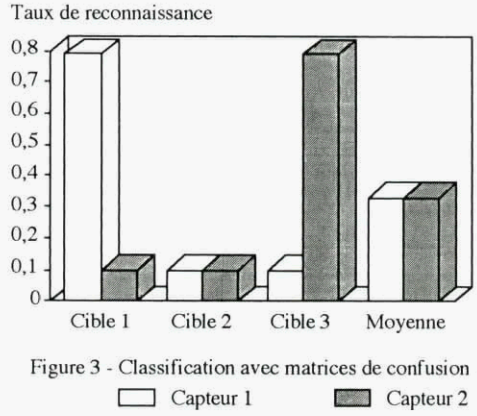


Figure 3 - Classification avec matrices de confusion

Le dernier exemple relève de la situation d'intérêt pratique ou les deux senseurs ont des capacités discriminantes complémentaires. Leurs taux de reconnaissance pour chaque cible et leur taux de reconnaissance moyen sont donnés en figure 3. Les taux similaires obtenus par les théories de l'incertain sont présentés en figure 4. Ils montrent, tant en stabilité vis-à-vis des différentes cibles qu'en valeur moyenne, l'avantage du premier modèle obtenu par la théorie de l'évidence, et la contre-performance de la théorie des possibilités due à un opérateur trop frustre. La meilleure solution dégagée peut être comparée, dans le même esprit, aux solutions 4 et 5 sur la figure 5. Il convient là encore de constater l'intérêt, dans l'ordre, de la prise en compte d'informations sur la validité des données disponibles et d'une décision centralisée, telles qu'assurées par l'approche proposée.

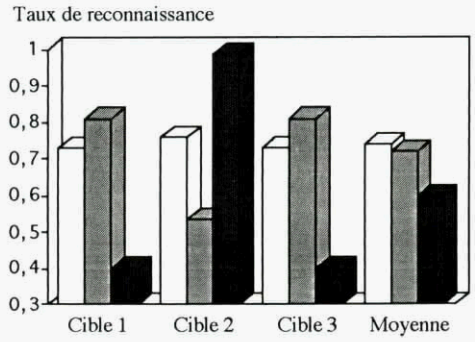


Figure 4 - Classification avec matrices de confusion
Solution 1 Solution 2 Solution 3

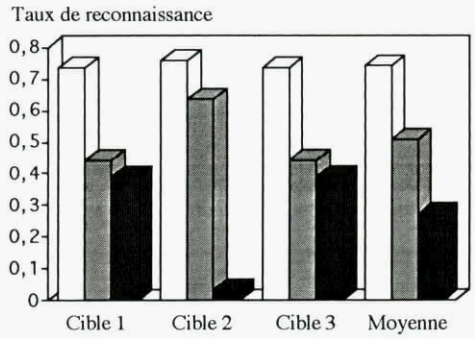


Figure 5 - Classification avec matrices de confusion
Solution 1 Solution 4 Solution 5

7. DISCRIMINATION STATISTIQUE SUPERVISEE

Le problème traité maintenant suppose que chacune des mesures m_j a pu faire l'objet d'un apprentissage préalable des distributions de probabilité a priori $p(m_j/H_i)$, sous les différentes hypothèses Hi, en vue de permettre la discrimination en phase de reconnaissance. Ce type d'apprentissage statistique est compatible avec la majorité des systèmes pour lesquels un certain nombre de mesures préliminaires peuvent être conduites dans différentes situations réelles ou simulées. Les histogrammes établis à partir de telles mesures sont en effet propres à fournir un modèle numérique ou analytique des distributions $p(m_j/H_i)$. En phase de reconnaissance, les N*M valeurs de densité de probabilité $p(m_j/H_i)$ associées respectivement aux M mesures locales m_j constituent les entrées des traitements de fusion et de reconnaissance qui vont être discutés dans la suite.

Si l'on considère le cas le plus fréquent où les mesures m_j peuvent être supposées statistiquement indépendantes, les senseurs étant en général choisis pour leur complémentarité d'information, la solution du problème posé est immédiate par l'approche bayésienne qui conduit typiquement à évaluer la probabilité a postériori $P(H_i/m_1,\dots,m_M)$ de chaque hypothèse H_i à l'aide de :

$$P(H_i/m_1,\dots,m_M) = \{[\prod_j p(m_j/H_i)]*P(H_i)\} \,/\, \sum_k \{[\prod_j p(m_j/H_k)]*P(H_k)\} \qquad (7.1)$$

où $P(H_i)$ désigne sa probabilité a priori. L'hypothèse H_i^* retenue est alors naturellement donnée par :

$$P(H_i^*/m_1,\dots,m_M) = \max_k\{P(H_k/m_1,\dots,m_M)\} \quad (7.2)$$

Une telle approche est toutefois rapidement mise en défaut lorsque les conditions d'observation réelles diffèrent des conditions d'apprentissage disponibles, ou lorsque la statistique des mesures est insuffisante à l'apprentissage. Le manque de maîtrise que l'on constate à ce niveau dans la plupart des applications amène en effet à utiliser des distributions, acquises lors de l'apprentissage, qui s'avèrent être plus ou moins représentatives des distributions réelles rencontrées. De plus, il est souvent difficile d'accéder à un jeu de probabilités a priori $P(H_i)$ capable de traduire fidèlement la réalité.

7.1 Formulation du problème

Le but poursuivi ici est donc de rechercher une procédure basée uniquement sur la connaissance des $p(m_j/H_i)$, et susceptible d'intégrer toute information sur la fiabilité des diverses distributions, telle qu'élaborée à partir d'une connaissance plus ou moins partielle des conditions d'observation.

Cette recherche est menée en considérant, pour plus de simplicité, que toute l'information qualitative disponible est synthétisée sous la forme de N*M coefficients $q_{ij}\in[0,1]$, chacun étant représentatif d'un degré de confiance placé respectivement dans la connaissance de chacune des N*M distributions $p(m_j/H_i)$. L'apprentissage des coefficients q_{ij} peut résulter d'une procédure d'élaboration contextuelle spécifique, ou d'une optimisation directe du traitement multi-senseurs complet sur des données variées préalables. Des q_{ij} nuls doivent permettre de considérer des informations incomplètes. Il est par ailleurs possible de constituer autant de jeux de coefficients q_{ij} que d'ensembles de conditions d'observation identifiables en temps réel, compte tenu de l'environnement disponible.

Traiter ce problème par la théorie de l'évidence exige de rechercher, pour chaque source S_j, une modélisation des N probabilités a priori $p(m_j/H_i)$ afférentes et de leurs N facteurs de confiance respectifs q_{ij} ($i\in[1,N]$) sous forme d'un jeu de masses élémentaires $m_j(.)$, associé à une fonction de crédibilité $Cr_j(.)$ et à une fonction de plausibilité $Pl_j(.)$. Les sources S_j étant distinctes, une évaluation globale m(.) peut alors être obtenue par sommation orthogonale des $m_j(.)$. Le cadre de discernement approprié est bien sûr l'ensemble des N hypothèses H_i répertoriées a priori.

A cette fin, il est proposé de conduire une recherche exhaustive et rigoureuse de l'ensemble des modèles susceptibles de satisfaire trois axiomes fondamentaux dans le contexte envisagé. Ces trois axiomes sont retenus a priori pour leur légitimité dans la plupart des applications visées :

<u>Axiome 7.1</u> : Cohérence avec l'approche bayésienne dans le cas où les distributions $p(m_j/H_i)$ apprises sont parfaitement représentatives des densités réellement rencontrées ($q_{ij}=1$, $\forall i,j$), et où les probabilités a priori $P(H_i)$ sont connues.

<u>Axiome 7.2</u> : Séparabilité de l'évaluation des hypothèses H_i ; chaque probabilité $p(m_j/H_i)$ doit être considérée comme une source d'information distincte donnant lieu à un jeu de masses $m_{ij}(.)$, notamment susceptible d'intégrer son facteur de confiance q_{ij} en terme d'affaiblissement (cf § 3.4). On impose ainsi à chaque jeu de masses $m_j(.)$ d'être la somme orthogonale des N jeux de masses $m_{ij}(.)$ considérés pour $i\in[1,N]$. De plus, les éléments focaux du jeu de masses $m_{ij}(.)$ ne peuvent être, compte tenu du mode d'élaboration de $p(m_j/H_i)$, que H_i, $\neg H_i$, et E, le cadre de discernement E étant toujours l'ensemble des hypothèses H_i.

<u>Axiome 7.3</u> : Cohérence avec l'association probabiliste des sources ; pour des sources Sj indépendantes et des densités $p(m_j/H_i)$ parfaitement représentatives de la réalité, les procédures de modélisation retenues doivent conduire au même résultat si l'on effectue la somme orthogonale de modélisations $m_j(.)$ élaborées à partir des $p(m_j/H_i)$, ou si l'on modélise directement les probabilités conjointes $p(m_1,\dots,m_M/H_i)$ obtenues par :

$$p(m_1,\dots,m_M/H_i) = \prod_j p(m_j/H_i) \quad (7.3)$$

La recherche des modèles satisfaisant ces trois axiomes va être conduite dans la suite en considérant la restriction progressive de l'ensemble des modèles possibles, lorsque les axiomes sont pris successivement en compte dans l'ordre de leur énoncé.

7.2 Axiome 7.1 : Cohérence avec l'approche bayésienne

7.2.1 *Développement*

Soit $m_0(.)$ le jeu de masses représentatif de la source d'information S_0 que constituent les probabilités a priori $P(H_i)$. $m_o(.)$ est donc un jeu de masses bayésien défini par :

$$m_0(Hi) = P(H_i)\,,\quad \forall\, i\in[1,N] \quad (7.4)$$

$$m_0(A) = 0\,,\quad \forall\, A\neq Hi\,,\ i\in[1,N] \quad (7.5)$$

La cohérence visée impose que la somme orthogonale des jeux de masses cherchés $m_j(.)$ et de $m_0(.)$ fournisse, dès lors que les distributions $p(m_j/H_i)$ sont parfaitement représentatives des densités réellement rencontrées, et que donc $q_{ij}=1$ pour tout i et j, un jeu de masses bayésien $m_b(.)$ en conformité avec l'inférence bayésienne (7.1). Cet axiome doit en particulier rester vrai quel que soit le sous-ensemble de sources S_j combinées, délimité par $j\in J\subset[1,M]$. Concrètement :

$$m_b(.) = \{\bigoplus_{j\in J} m_j(.)\} \oplus m_0(.) \quad (7.6)$$

doit donc vérifier dans ces conditions :

$$m_b(H_i) = \{[\prod_j p(m_j/H_i)]*P(H_i)\} \,/\, \sum_k \{[\prod_j p(m_j/H_k)]*P(H_k)\}\,,\ \forall\, H_i\in E \quad (7.7)$$

Or, compte tenu de la propriété (3.31), les équations (7.4), (7.5), et (7.6) conduisent à :

$$m_b(H_i) = \{[\prod_j Pl_j(H_i)]*P(H_i)\} \,/\, \sum_k \{[\prod_j Pl_j(H_k)]*P(H_k)\}\,,\ \forall\, H_i\in E \quad (7.8)$$

Satisfaire conjointement (7.7) et (7.8) pour tout $J\subset[1,M]$ amène finalement à définir chaque $m_j(.)$ par sa fonction de plausibilité, à l'aide des N équations :

$$Pl_j(H_i) = K_j*p(m_j/H_i)\quad,\quad i\in[1,N] \quad (7.9)$$

où K_j est un paramètre unique pour les N équations, simplement défini par :

$$K_j \in [\{\sum_i p(m_j/H_i)\}^{-1},\ \{\max_i[p(m_j/H_i)]\}^{-1}] \quad (7.10)$$

Ces bornes sur K_j sont uniquement imposées par la nature de la notion de plausibilité : cette grandeur doit notamment rester inférieure à 1, et la somme des valeurs qu'elle prend pour des événements constituant une partition de E (les H_i eux-mêmes ici) doit être supérieure à 1.

7.2.2 *Commentaires*

Le résultat ainsi dégagé de l'axiome 7.1 appelle quelques commentaires. Tout d'abord, dans le cas général où N>2, il existe pour chaque valeur de K_j autre que la valeur minimale imposée par (7.10) une infinité de jeux de masses possibles, définis par un système de N+1 équations (N équations (7.9) et la somme des masses égale à 1) à 2^N-1 inconnues.

Pour la valeur minimale de K_j, le résultat obtenu se résume dans tous les cas à un jeu de masses unique, et de surcroît bayésien :

$$m_j(H_i) = p(m_j/H_i) \,/\, \sum_k p(m_j/H_k)\ ,\ \forall\ i\in[1,N] \qquad (7.11)$$

$$m_j(A) = 0\ ,\ \forall\ A\neq H_i\ ,\ i\in[1,N] \qquad (7.12)$$

Parmi les diverses solutions obtenues pour la valeur maximale de K_j, il existe une solution consonante, unique sur l'ensemble des solutions dégagées. Elle correspond au modèle proposé par G. SHAFER lui-même, sur la base de cette seule caractéristique, pour un contexte similaire à celui du présent axiome 7.1 [5]. Pour donner l'expression pratique de cette solution, supposons les $p(m_j/H_i)$ rangés de telle sorte que $p(m_j/H_1)\geq p(m_j/H_2)\geq\ldots\geq p(m_j/H_N)$. Les éléments focaux sont alors les N sous-ensembles de E :

$$A_i = \bigcup_{k\leq i} H_k\ ,\ i\in[1,N] \qquad (7.13)$$

et les masses correspondantes sont obtenues conformément à (3.35) et (3.36) :

$$m_j(A_N) = K_j{*}p(m_j/H_N) \qquad (7.14)$$

$$m_j(A_i) = K_j{*}\{p(m_j/H_i)-p(m_j/H_{i+1})\}\ ,\ \text{pour } 1\leq i\leq N-1 \qquad (7.15)$$

Il convient toutefois de remarquer que cette dernière solution ne satisfait pas les axiomes 7.2 et 7.3, et qu'en conséquence elle ne pourra pas être retenue dans la suite.

Notons enfin que dans le cas idéal où les distributions $p(m_j/H_i)$ sont parfaitement représentatives des densités réellement rencontrées, une procédure de maximum de vraisemblance conduit à retenir l'hypothèse H_i qui maximise $p(m_1,\ldots,m_M/H_i)$, c'est-à-dire le produit des $p(m_j/H_i)$ fournis par les M sources S_j supposées indépendantes. Par ailleurs, les notions de plausibilité et de communalité étant identiques pour les singletons H_i du cadre de discernement E, la loi de combinaison (3.14) est applicable aux $Pl_j(H_i)$ pour obtenir la plausibilité $Pl(H_i)$ après fusion des sources S_j. La relation (7.9) amène dans ces conditions :

$$Pl(H_i) = K_f{*}p(m_1,\ldots,m_M/H_i)\ ,\ \forall\ i\in[1,N] \qquad (7.16)$$

où le coefficient K_f, indépendant de H_i, intègre les K_j et l'inconsistance de la fusion. Pour rester cohérente avec ce cas particulier, toute procédure de décision visant à désigner l'hypothèse la plus vraisemblable devra pour notre problème exclusivement maximiser une fonction monotone croissante de la plausibilité $Pl(H_i)$ obtenue après fusion des sources S_j. Ce résultat est en parfait accord avec les conclusions de l'approche générale présentée au § 4.

7.3 Axiome 7.2 : Séparabilité de l'évaluation des hypothèses

Cet axiome consiste à considérer que chaque jeu de masse $m_j(.)$ cherché est lui-même le résultat d'une fusion entre N jeux de masses $m_{ij}(.)$ ($i\in[1,N]$) :

$$m_j(.) = \bigoplus_i m_{ij}(.) \qquad (7.17)$$

Un jeu de masses $m_{ij}(.)$ possède par ailleurs trois éléments focaux (H_i, $\neg H_i$, et E) dont les masses ne dépendent que de la valeur $p(m_j/H_i)$ et du facteur q_{ij} correspondants.

La règle (3.14) peut être directement appliquée aux plausibilités associées $Pl_{ij}(H_i)$, pour fournir la plausibilité $Pl_j(H_i)$ résultant de la fusion (7.17), puisque les notions de plausibilité et de communalité sont identiques pour les singletons H_i du cadre de discernement E. Les plausibilités $Pl_j(H_i)$ peuvent ainsi être mises sous la forme :

$$Pl_j(H_i) = Kf_j{*}\{m_{ij}(H_i)+m_{ij}(E)\}/\{m_{ij}(\neg H_i)+m_{ij}(E)\},\ i\in[1,N] \qquad (7.18)$$

où le facteur Kf_j est indépendant de l'hypothèse H_i concernée.

Le respect de la contrainte (7.9) imposée par l'axiome 7.1 ne permettra alors d'associer au seul jeu de masses $m_{ij}(.)$ la seule probabilité $p(m_j/H_i)$, pour $q_{ij}=1$, que si :

$$\{m_{ij}(H_i)+m_{ij}(E)\}/\{m_{ij}(\neg H_i)+m_{ij}(E)\} = R_j{*}p(m_j/H_i) \qquad (7.19)$$

où R_j est une constante de normalisation indépendante de H_i, dont les valeurs possibles ne dépendent que des distributions $p(m_j/H_i)$ effectivement prises en compte, comme nous le verrons dans la suite. Cette constante permet en pratique de considérer le cadre général où les $p(m_j/H_i)$ ne sont connues que de façon relative, c'est-à-dire à un gain de normalisation près.

Exprimée de façon paramétrique en fonction du niveau d'incertitude $m_{ij}(E)$, (7.19) procure le jeu de masses cherché :

$$m_{ij}(H_i) = \{R_j{*}p(m_j/H_i)-m_{ij}(E)\}/\{1+R_j{*}p(m_j/H_i)\} \qquad (7.20)$$

$$m_{ij}(\neg H_i) = \{1-R_j{*}p(m_j/H_i){*}m_{ij}(E)\}/\{1+R_j{*}p(m_j/H_i)\} \qquad (7.21)$$

$$m_{ij}(E) = f[R_j{*}p(m_j/H_i)] \in [0,\ R_j{*}p(m_j/H_i)] \qquad (7.22)$$

où f est une fonction quelconque vérifiant simplement (7.22).

Cette condition (7.22) est imposée par la notion de masse (comprise entre 0 et 1) qui limite également les valeurs possibles de R_j en fonction des distributions $p(m_j/H_i)$ utilisées, et ceci indépendamment des mesures m_j effectivement observées :

$$R_j \in [0,\ (\max_{m_j,i}\{p(m_j/H_i)\})^{-1}] \qquad (7.23)$$

Il est par ailleurs possible de montrer que ces conditions suffisent à ce que le coefficient K_j de l'expression (7.9), calculé pour la fusion (7.17), vérifie bien la contrainte (7.10). Ceci peut en particulier se faire simplement en mettant en évidence que l'expression de K_j est alors une fonction monotone croissante de chaque $m_{ij}(E)$, dont les valeurs extrémales permettent de satisfaire l'intervalle (7.10).

Si l'on introduit le facteur q_{ij} dans les expressions (7.20), (7.21), et (7.22) en terme d'affaiblissement, conformément à (3.24) et (3.25), les $m_{ij}(.)$ sont finalement données par :

$$m_{ij}(H_i) = q_{ij}{*}\{R_j{*}p(m_j/H_i)-A_i\}/\{1+R_j{*}p(m_j/H_i)\} \qquad (7.24)$$

$$m_{ij}(\neg H_i) = q_{ij}{*}\{1-R_j{*}p(m_j/H_i){*}A_i\}/\{1+R_j{*}p(m_j/H_i)\} \qquad (7.25)$$

$$m_{ij}(E) = 1-q_{ij}+q_{ij}{*}A_i \qquad (7.26)$$

où R_j reste défini par (7.23), et A_i par :

$$A_i = f[R_j{*}p(m_j/H_i)] \in [0,\ R_j{*}p(m_j/H_i)] \qquad (7.27)$$

l'expression générale des modèles $m_j(.)$ qui satisfont les axiomes 7.1 et 7.2 est ainsi obtenue par (7.17) appliquée à (7.24), (7.25), et (7.26). Une infinité de solutions répondent donc encore au problème.

7.4 Axiome 7.3 : Cohérence avec l'association probabiliste des sources

Compte tenu de la structure particulière (7.17) des jeux de masses $m_j(.)$ répondant aux axiomes 7.1 et 7.2, et compte tenu de l'associativité de la somme orthogonale, l'axiome 7.3 sera satisfait pour les modèles tels que, si les q_{ij} sont égaux à 1, le jeu de masses $m_i(.)$ défini par :

$$m_i(.) = \bigoplus_j m_{ij}(.) \tag{7.28}$$

$$m_{ij}(.) = F[R_j * p(m_j/H_i)] \tag{7.29}$$

est identique au jeu de masses $m'_i(.)$ obtenu par la modélisation directe, à l'aide de la même fonction F(.) :

$$m'_i(.) = F[\prod_j \{R_j * p(m_j/H_i)\}] \tag{7.30}$$

Les $m_{ij}(.)$ vérifiant (7.20), (7.21), (7.22), et (7.23), procurent dans la fusion (7.28) :

$$m_i(H_i) = (V*X-Y*W)/(V*X+X-Y*W) \tag{7.31}$$

$$m_i(\neg H_i) = (X-Y*W)/(V*X+X-Y*W) \tag{7.32}$$

$$m_i(E) = Y*W/(V*X+X-Y*W) \tag{7.33}$$

avec les définitions :

$$V = \prod_j \{R_j * p(m_j/H_i)\} \tag{7.34}$$

$$W = \prod_j \{1+R_j * p(m_j/H_i)\} \tag{7.35}$$

$$X = \prod_j \{1+m_{ij}(E)\} \tag{7.36}$$

$$Y = \prod_j m_{ij}(E) \tag{7.37}$$

et les contraintes :

$$m_{ij}(E) = f[R_j * p(m_j/H_i)] \in [0, R_j * p(m_j/H_i)] \tag{7.38}$$

$$R_j \in [0, (\max_{m_j,i}\{p(m_j/H_i)\})^{-1}] \tag{7.39}$$

Parallèlement le jeu de masses $m'_i(.)$ s'écrit :

$$m'_i(H_i) = \{V-m'_i(E)\}/\{1+V\} \tag{7.40}$$

$$m'_i(\neg H_i) = \{1-V*m'_i(E)\}/\{1+V\} \tag{7.41}$$

$$m'_i(E) = f[\prod_j \{R_j * p(m_j/H_i)\}] \in [0 , \prod_j \{R_j * p(m_j/H_i)\}] \tag{7.42}$$

où V est toujours donnée par (7.34), la contrainte sur les R_j étant par ailleurs satisfaite par (7.39).

La comparaison des jeux de masses $m_i(.)$ et $m'_i(.)$ obtenus peut être abordée en imposant $m_i(E)=m'_i(E)$ dans (7.33). Dès lors les expressions (7.40) et (7.41) sont respectivement équivalentes aux expressions (7.31) et (7.32), entraînant en toutes circonstances $m_i(H_i)=m'_i(H_i)$ et $m_i(\neg H_i)=m'_i(\neg H_i)$. En revanche (7.38) et (7.42) ne seront équivalentes pour une même fonction f, au travers de (7.33) toujours contraint par $m_i(E)=m'_i(E)$, que pour les deux fonctions f suivantes :

$$f(x) = 0 , \ \forall x \tag{7.43}$$

$$f(x) = x \tag{7.44}$$

Après examen de l'axiome 7.3, les modèles satisfaisant simultanément les trois axiomes sont donc réduits au nombre de deux. Tous deux sont définis par (7.24), (7.25), (7.26). Ils se différencient par le fait que pour l'un $A_i=0$, alors que pour l'autre $A_i=R_j*p(m_j/H_i)$, avec dans les deux cas R_j contraint par (7.23).

7.5 Synthèse des modèles obtenus

Les modèles satisfaisant conjointement les trois axiomes souhaitées sont donc finalement au nombre de deux. Tous deux répondent à la décomposition :

$$m_j(.) = \bigoplus_i m_{ij}(.) \tag{7.45}$$

Le <u>modèle 1</u> est particularisé par :

$$m_{ij}(H_i) = 0 \tag{7.46}$$

$$m_{ij}(\neg H_i) = q_{ij} * \{1-R_j * p(m_j/H_i)\} \tag{7.47}$$

$$m_{ij}(E) = 1-q_{ij}+q_{ij} * R_j * p(m_j/H_i) \tag{7.48}$$

et le <u>modèle 2</u> par :

$$m_{ij}(H_i) = q_{ij} * R_j * p(m_j/H_i)/\{1+R_j * p(m_j/H_i)\} \tag{7.49}$$

$$m_{ij}(\neg H_i) = q_{ij}/\{1+R_j * p(m_j/H_i)\} \tag{7.50}$$

$$m_{ij}(E) = 1-q_{ij} \tag{7.51}$$

Dans les deux cas le facteur de normalisation R_j est simplement contraint par :

$$R_j \in [0, (\max_{m_j,i}\{p(m_j/H_i)\})^{-1}] \tag{7.52}$$

Il est à remarquer que la méconnaissance totale d'une distribution $p(m_j/H_i)$, caractérisée par $q_{ij}=0$, revient bien dans les deux cas à ignorer le jeu de masse $m_{ij}(.)$ correspondant, puisqu'alors celui-ci est trivial ($m_{ij}(E)=1$) et donc élément neutre de la somme orthogonale.

7.6 Solutions du problème

Moyennant le fait que les notions de plausibilité et de communalité sont équivalentes pour les singletons H_i du cadre de discernement E, la loi de combinaison (3.14) peut être directement appliquée aux plausibilités $Pl_j(H_i)$ définies par (7.18) pour notre problème, en vue de fournir la plausibilité $Pl(H_i)$ qui devra être maximisée par la procédure de décision établie au § 4 et confortée au § 7.2.2 :

$$Pl(H_i) = \prod_j \{[m_{ij}(H_i)+m_{ij}(E)]/[m_{ij}(\neg H_i)+m_{ij}(E)]\} \tag{7.53}$$

La procédure de décision correspondant à chacun des deux modèles dégagés au § 7.5 est alors obtenue pour leurs jeux de masses $m_{ij}(.)$ respectifs, en appliquant (4.9) à (7.53). Ceci conduit à retenir l'hypothèse H_i qui satisfait, <u>pour le modèle 1</u> :

$$\max_i\{\prod_j [1-q_{ij}+q_{ij} * R_j * p(m_j/H_i)]\} \tag{7.54}$$

et <u>pour le modèle 2</u> :

$$\max_i\{\prod_j [\{1-q_{ij}+R_j * p(m_j/H_i)\}/\{1+(1-q_{ij}) * R_j * p(m_j/H_i)\}]\} \tag{7.55}$$

Si l'on cherche à comparer ces deux procédures en considérant les expressions optimisées comme deux fonctions des variables q_{ij} et $[R_j * P(m_j/H_i)]$, il apparaît qu'elles sont numériquement très proches ;

elles satisfont les mêmes valeurs limites et les mêmes monotonies, avec une différence maximale de quelques % pour des valeurs médianes particulières ; les valeurs prises par la fonction relative au modèle 2 sont par ailleurs toujours supérieures à celles de la fonction établie sur le modèle 1. Il y a donc tout lieu de penser qu'en pratique ces deux formulations mèneront à des performances en général équivalentes. Ceci est notamment vérifié dans le cas des simulations présentées dans la suite, pour lesquelles aucune tendance significative n'a pu être dégagée en faveur de l'une ou l'autre des deux approches. Dans une telle situation, il est clair que notre préférence doit aller à la formulation la plus simple, à savoir celle associée au modèle 1. Seule la procédure (7.54) sera donc discutée dans la suite.

Notons enfin que lorsque tous les q_{ij} valent 1, c'est-à-dire lorsque les distributions $p(m_j/H_i)$ sont parfaitement représentatives de la réalité, les deux approches se ramènent bien à une procédure de maximum de vraisemblance.

7.7 Lien avec le problème générique

Le problème de discrimination à partir d'un apprentissage statistique, tel que traité dans ce § 7, est en fait un cas particulier du problème générique discuté au § 5. A l'inverse de l'application à la classification avec matrices de confusion menée au § 6, les cœfficients de confiance qij conservent ici toute leur généralité. En revanche les critères Cij sont maintenant particularisés en ce qu'ils assurent la prise en compte des distributions p(mj/Hi).

Les deux procédures (5.9) et 5.10) obtenues au § 5 sont rigoureusement équivalentes aux deux procédures (7.53) et (7.54) dégagées ici, si l'on adopte pour C_{ij} les définitions respectives suivantes :

Pour le modèle 1 : $C_{ij} = R_j * p(m_j/H_i)$ (7.56)

Pour le modèle 2 : $C_{ij} = R_j * p(m_j/H_i)/[1+R_j * p(m_j/H_i)]$ (7.57)

où R_j reste bien sûr le gain de normalisation contraint par (7.52).

Ce résultat est en fait légitime si l'on note que l'axiome 5.1 est directement exprimé par l'axiome 7.2, et que les solutions contraintes par les axiomes 7.1 et 7.3 vérifient automatiquement l'axiome 5.2. Les axiomes 7.1 et 7.3 permettent simplement de préciser la prise en compte des informations particulières p(mj/Hi) dans l'expression du critère C_{ij}.

Il convient cependant de noter que la nature probabiliste des contraintes qui permettent de définir plus précisément le problème traité ici (cohérence avec Bayes) interdit tout développement formel dans le cadre de la théorie des possibilités. Les notions de crédibilité bayésienne et de crédibilité consonante sont en effet incompatibles (cf § 3.6).

7.8 Simulations

Comme au § 6, l'analyse de quelques simulations très simples va nous permettre de mieux appréhender le comportement des procédures dégagées, d'évaluer leur potentiel, et de préciser leur emploi privilégié, indépendamment du domaine d'application envisagé.

Pour cela considérons un problème de discrimination entre deux hypothèses H_1 et H_2 représentatives, par exemple, de l'identité d'une cible dans un contexte de reconnaissance. Les distributions $p(m_j/H_i)$, apprises au préalable par les senseurs S_j utilisés, sont des lois normales réduites :

$$p(m_j/H_1) = N(0,1) \quad (7.58)$$

$$p(m_j/H_2) = N(Sa_j,1) \quad (7.59)$$

figurant la mesure d'attributs déterministes, entachée d'un bruit additif gaussien. Les mesures m_j simulant les relevés rencontrés dans la réalité sont alors générées à partir de lois normales réduites différentes :

$$p_r(m_j/H_1) = N(0,1) \quad (7.60)$$

$$p_r(m_j/H_2) = N(Sr_j,1) \quad (7.61)$$

propres à traduire une éventuelle dérive Sr du signal attendu Sa, par exemple liée à l'évolution de la signature (radar, infrarouge,...) de la cible en fonction de l'environnement, de la météorologie, des conditions de la mesure,...

Différents scénarios d'intérêt pratique vont être balayés dans ce cadre. Compte tenu de la remarque exprimée au § 7.6, seuls les résultats obtenus à l'aide de la procédure basée sur le modèle 1 sont présentés ; ils sont représentatifs des deux procédures dégagées.

7.8.1 *Cas 1 : 1 capteur et 1 donnée incertaine*

La figure 6 présente les résultats obtenus avec un senseur unique S_1 lorsque le signal Sr_1 réellement présenté prend diverses valeurs S, plus ou moins différentes du signal attendu $Sa_1=6$. Corrélativement, l'évaluation $p(m_1/H_1)$ est considérée comme certaine ($q_{11}=1$), alors que l'évaluation $p(m_1/H_2)$ est sujette à incertitude, q_{21} pouvant prendre différentes valeurs paramétriques q. Le taux de bonne reconnaissance de l'hypothèse H_i effectivement présentée est obtenu en générant des mesures m_1 en nombre égal selon la loi $p_r(m_1/H_1)$ et selon la loi $p_r(m_1/H_2)$.

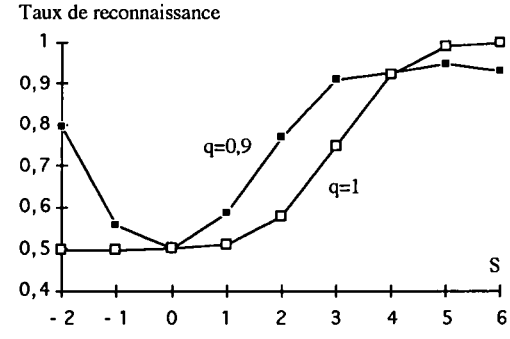


Figure 6 - Classification avec apprentissage statistique : 1 senseur - $Sa_1=6$; $Sr_1=S$; $q_{11}=1$; $q_{21}=q$

Rappelons qu'en vertu de la remarque faite à la fin du § 7.6, la courbe obtenue pour q=1 correspond à l'approche probabiliste traditionnelle qui suppose dans tous les cas la parfaite représentativité des distributions $p(m_j/H_i)$ ($Sr_1=Sa_1$). Par rapport à cette courbe, toute diminution de q tend progressivement à réduire légèrement les performances dans le cas d'un apprentissage cohérent ($Sr1 \approx Sa1$), pour les améliorer sensiblement dans les cas d'information défectueuse ($Sr_1 << Sa_1$), y compris pour une inversion de contraste ($Sr_1<0$) par exemple représentative de l'évolution de signatures infrarouges. Notons que le taux de 0,5 obtenu pour S=0 est incontournable par quelque méthode que ce soit, puisqu'alors les deux distributions $p_r(m_1/H_1)$ et $p_r(m_1/H_2)$ sont parfaitement confondues.

Au-delà d'une optimisation fine des facteurs de qualité q_{ij} pour des conditions particulières d'erreur dont une modélisation paramétrique permettra bien souvent une approche probabiliste compétitive, bien que plus complexe, ces résultats mettent en lumière l'intérêt majeur d'un fonctionnement robuste à deux états : q=1 si l'erreur ne peut que rester faible (typiquement $Sa_1-Sr_1<2$), et q de l'ordre de 0,9 ici si elle a de fortes chances d'être plus importante, et cela sans pouvoir être caractérisée.

7.8.2 *Cas 2 : 2 capteurs et 1 donnée incertaine*

L'intérêt de cette configuration est essentiellement de pallier l'incertitude susceptible d'affecter un capteur de bonne qualité, en lui associant un capteur de qualité moins bonne, mais sûre. La figure 7 présente donc le résultat obtenu lorsqu'on adjoint au senseur S_1 du cas 1 un senseur S_2 tel que $Sa_2=Sr_2=2$ (et donc $q_{12}=q_{22}=1$).

Les conclusions dégagées dans le cas 1 restent parfaitement valables ici. De surcroît, la valeur de q voisine de 0,9 permet de bénéficier pleinement des performances du senseur S_2 dans le cas le plus défavorable ($Sr_1=S=0$) avec un taux supérieur à 0,8, alors que l'approche classique,

stigmatisée par q=1, n'arrive pas à décoller de 0,5. D'une façon plus globale, la procédure élaborée procure ici un gain en performances particulièrement important dès que l'erreur commise n'est plus négligeable (typiquement Sa_1-Sr_1>2), grâce au parti qu'elle permet de tirer des informations disponibles.

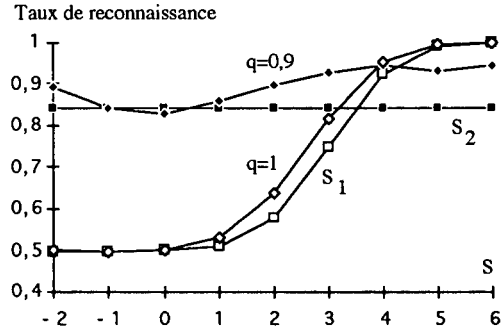


Figure 7 - Classification avec apprentissage statistique : 2 senseurs - Sa_1=6 ; Sr_1=S ; Sa_2=Sr_2=2 ; q_{11}=q_{12}=q_{22}=1 ; q_{21}=q

Le fonctionnement à deux états introduit pour le cas 1 (q=0,9 ou 1 selon que Sa_1-Sr_1 risque d'être important ou non) permet en outre d'assurer dans chacune des deux circonstances identifiées des performances multi-senseurs au moins égales à celles du meilleur senseur pris isolément, que l'on considère le taux de reconnaissance minimal ou le taux moyen sur l'intervalle des valeurs de S alors respectivement concernées. Cette faculté est importante en ce qu'elle correspond à une finalité indispensable pour les systèmes multi-senseurs, le plus souvent délicate à satisfaire en présence d'erreurs mal maîtrisées.

La figure 8 met par ailleurs en évidence une bonne stabilité des résultats obtenus pour q=0,9 lorsque cette valeur décroît, compte tenu de la précision statistique des essais menés. Le fonctionnement robuste dégagé plus haut est donc peu sensible au choix de q qui peut par conséquent être fixé de façon relativement arbitraire, sans optimisation fine.

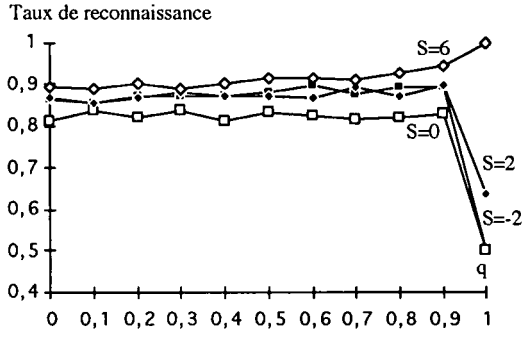


Figure 8 - Classification avec apprentissage statistique : stabilité des performances en fonction du choix du facteur d'incertitude q, pour le cas 2

7.8.3 *Cas 3 : 2 capteurs et 2 données incertaines*

Il s'agit ici de chercher à pallier l'incertitude du senseur S_1, toujours inchangé, en lui associant cette fois un senseur S_2 de même bonne qualité, mais aussi de même incertitude. On considère donc que Sa_1=Sa_2=6, que q_{11}=q_{12}=1, et que q_{21}=q_{22}=q, facteur commun d'incertitude. En effet, compte tenu des configurations déjà couvertes par le cas 2 (au moins un capteur "certain"), on ne s'intéresse maintenant qu'aux situations où l'erreur risque d'être importante simultanément sur les deux senseurs (Sa_1-Sr_1 et Sa_2-Sr_2 grands tous les deux).

Les résultats présentés en figure 9 sont significatifs des performances accessibles. Ils correspondent à Sr_2=2 et à une évolution paramétrique S de Sr_1. La courbe q=0,9 fait de nouveau apparaître un gain substantiel pour les situations visées par le présent exemple, confortant encore une fois les meilleures performances garanties par la procédure proposée face à chaque capteur pris isolement ou face à la fusion probabiliste classique (q=1), que l'on considère le taux de reconnaissance minimal ou le taux moyen sur l'intervalle des valeurs de S alors concernées.

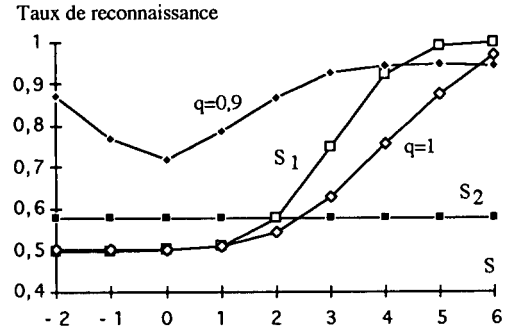


Figure 9 - Classification avec apprentissage statistique : 2 senseurs - Sa_1=Sa_2=6 ; Sr_1=S ; Sr_2=2 ; q_{11}=q_{12}=1 ; q_{21}=q_{22}=q

8. APPORTS DE L'APPROCHE PROPOSEE

L'intérêt de la solution générale proposée au § 5 a été dégagé dans le cadre de deux types de problèmes complémentaires, représentatifs d'une majorité d'applications pratiques. Dans le premier cas la solution est basée sur la connaissance préalable des matrices de confusion de chaque senseur, alors que dans le deuxième cas elle prend en compte un apprentissage statistique plus ou moins représentatif des mesures.

Le gain en performances et en robustesse mis en évidence dans les deux cas sur quelques simulations très simples est notamment lié à la meilleure exploitation des particularités du domaine d'emploi que permet l'approche proposée : mise en œuvre d'une fusion la plus centralisée possible au niveau le plus riche en information, gestion de l'incertitude sur les données utilisées, prise en compte d'informations exogènes éventuellement subjectives sur la validité des observations ou leur pouvoir informatif et sur le contexte, valorisation de la complémentarité des capteurs, rentabilisation de l'apport de toute source supplémentaire,...

L'approche proposée est en particulier bénéfique face à différents problèmes d'apprentissage : pollution des banques de données par des informations parasites, défaut de représentativité des modèles, notamment sous des conditions d'observation évolutives, insuffisance des relevés disponibles,... Elle permet également de gérer avec succès la disparité intrinsèque du potentiel informatif des différentes sources.

Il convient également de rappeler que la démarche adoptée conduit naturellement à traiter des informations incomplètes : la méconnaissance totale d'une évaluation C_{ij} (q_{ij}=0) est élément neutre de la fusion, une méconnaissance partielle peut être contrôlée par le facteur q_{ij}, et un cadre de discernement incomplet peut être géré directement à partir des modèles élémentaires, ceux-ci étant simplement définis sur $\{H_i, \neg H_i\}$ sans présager du contenu de $\neg H_i$.

Les algorithmes présentés n'exigent par ailleurs qu'une charge de calcul très limitée, parfaitement compatible avec les contraintes opérationnelles, et en particulier négligeable devant les prétraitements d'extraction propres au sources utilisées (traitement d'images,...).

Les applications traitées sont multiples : reconnaissance radar de cibles aériennes, maritimes, ou terrestres, à l'aide de formes d'onde multiples (fréquences, types de résolution, polarimétries, discriminants multiples extraits d'une même réponse,...) ; reconnaissance de sources sonar ou radar en écoute passive, par analyses multiples des formes d'onde émises ; reconnaissance de cibles terrestres par imagerie bimode radar et optronique, dans un contexte d'observation satellitaire ou aéroporté, ou à des fins de guidage terminal d'engins ; alerte bimode radar et infrarouge en défense aérienne ; reconnaissance automatique d'amers pour le recalage de la navigation de véhicules aériens à l'aide de senseurs

millimétriques et infrarouges ; interprétation des données dans des systèmes d'aide à la décision.

Un certain nombre de travaux ont par ailleurs été menés pour intégrer ce formalisme au sein de méthodes de pistage de cibles, de façon à bénéficier de la meilleure synergie entre les fonctions classification et poursuite.

Il reste maintenant à en étendre le domaine de mise en œuvre du concept présenté à la prise en compte d'informations plus variées, par exemple en recherchant une méthodologie plus systématique d'élaboration des critères Cij et des cœfficients qij, et ceci notamment sur la base des cadres théoriques compatibles (probabilités, ensembles flous, théorie des possibilités, théorie des fonctions d'utilité, mesures d'information,...).

9. BIBLIOGRAPHIE

1. A. APPRIOU, "Formulation et traitement de l'incertain en analyse multi-senseurs", Conférence invitée, 14ème Colloque GRETSI, Juan-les-Pins, 13-16 septembre 1993.

2. A. APPRIOU, "Probabilités et incertitude en fusion de données multi-senseurs", Revue Scientifique et Technique de la Défense, n°11, 1991-1, pp 27-40.

3. A. APPRIOU, "Perspectives liées à la fusion de données", Science et Défense, Paris, mai 1990.

4. A. APPRIOU, "Intérêt des théories de l'incertain en fusion de données", Conférence invitée, Colloque International sur le Radar, Paris, 24-28 avril 1989.

5. G. SHAFER, "A mathematical theory of evidence", Princeton University Press, Princeton, New Jersey, 1976.

6. R. R. YAGER, "Entropy and specificity in a mathematical theory of evidence", International Journal General Systems, Vol. 9, 1983, pp 249-260.

7. R. R. YAGER, "A general approach to decision making with evidential knowledge", Uncertainty in Artificial Intelligence, L. N. Kanal & J. F. Lemmer éd., Elsevier Science Publishers, B. V. North-Holland, 1986.

8. D. DUBOIS, H. PRADE, "Théorie des possibilités : application à la représentation des connaissances en informatique", Masson, Paris, 1988.

THE USE OF BAYESIAN BELIEF NETWORKS TO FUSE CONTINUOUS AND DISCRETE INFORMATION FOR TARGET RECOGNITION, TRACKING, AND SITUATION ASSESSMENT

Leland Stewart & Perry McCarty, Jr.
Lockheed Palo Alto Research Laboratories
3251 Hanover St.
Palo Alto, CA 94304
USA

1. SUMMARY

This paper describes the use of Bayesian Belief Networks for the fusion of continuous and discrete information. Bayesian belief networks provide a convenient and straightforward way of modeling the relationships between uncertain quantities. They also provide efficient computational algorithms. Most current applications of belief networks are restricted to either discrete or continuous quantities. We present a methodology that allows both discrete and continuous variables in the same network. This extension makes possible the fusion of information from, or inferences about, such diverse quantities as sensor output, target location, target type or ID, intent, operator judgment, behavior profile, etc.

2. MODELING USING BELIEF NETWORKS

Target recognition and tracking performance is improved by taking advantage of all information available. This includes sensor information (target signatures, lines-of-bearing, etc.), intelligence information and historical information about target behavior profiles. Knowledge about target behavior profiles, for example, may identify behaviors that are related to target modes of operation. Sensor contacts provide samples from these behavior profiles that allow inference about a mode of operation. Fusing all of these factors in order to improve target recognition and tracking requires the ability to accurately model the dependencies among all factors.

Belief networks provide a principled approach for constructing and reasoning with models in problem domains that require managing dependency relationships among a large number of variables. An example belief network for target recognition is shown in Figure 1. Ovals (or nodes) represent variables relevant to the domain, while arcs represent dependencies. The belief network also includes the specification of the probability distribution of each variable conditional on the values of its parent variables.

The variable *ID* represents the target type. The states of this variable are the different types of targets. The target recognition problem is to identify the actual state of this variable which is not observed directly, only inferred. The variable M_n represents operating mode (e.g. transit or attack). The composite variable $\{A_n, \tau_n\}$ depicts the current activity and time in that activity. This variable can be used to encode behavior profiles that are known to be related to the operating mode.The inclusion of velocity and location variables, V_n and L_n, shows that a tracker is an integral part of the identification model. Finally, the D_n variables are the only variables that are actually directly observed. In general, these may be contact reports such as lines-of-bearing or position fixes, signature data from sensors that measure parameters that can be used to discriminate among target IDs, or other types of intelligence information.

Dependencies exist among all of the variables in the belief network. Model construction requires identifying dependencies as being either direct, or indirect. An arc connecting two variables shows a direct dependency. In general, a model could have direct dependencies among all variables, and would be shown as a fully interconnected graph. In practice, variables either exist or can be introduced in a model to reduce the number of direct dependencies and simplify the problem of quantifying dependency relationships. For example, L_{n+1}, the location of the target at time $n+1$, is shown as dependent on the target location at time n and on target velocity. Location L_{n-1} also is relevant to L_{n+1} but provides no further information once velocity and location at t_n are known, so no arc is shown from L_{n-1} to L_{n+1}.

The ability to construct a model that has a large number of indirect dependencies is critical to the successful application of belief networks to large problems. The probability distribution of each variable is assessed conditionally for each combination of values of predecessor nodes. Nodes that have no predecessors are specified as prior probabilities. The mathematical foundations of belief networks shows that the joint distribution of all of the variables in the network is equal to the product of the conditional distributions at each node (given the dependency constraints that are stated in the network). This is important to understanding the semantics of belief networks, however it could be very difficult computationally to work with the full

E-mail: stewart@rddvax.decnet.lockheed.com,
mccarty@aic.lockheed.com

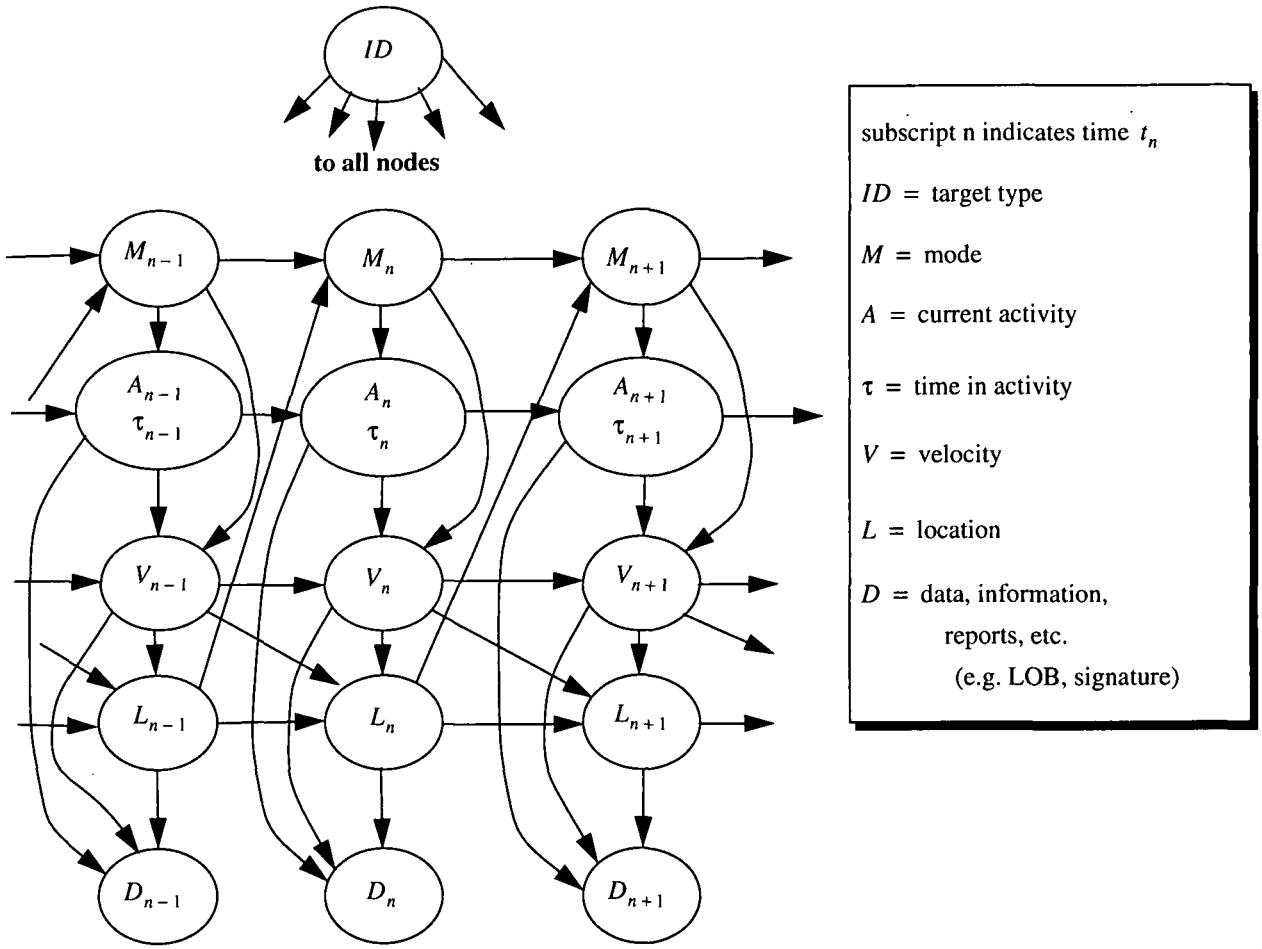


Figure 1. Target recognition and tracking belief network

joint probability directly. Instead, a significant body of research has investigated methods for using local computations to achieve the same effect as working with the joint, but at lower computational cost [1].Using these techniques, networks of over 1000 variables have been constructed that perform probabilistic reasoning in a few seconds on Sun 4 workstations [2].

Most belief network research has been conducted using models composed of discrete variables. More recently, the methods have been extended to construct Gaussian belief networks, belief networks for multivariate normal densities [3]. For certain cases these methods have been combined to allow mixed discrete and continuous models [4]. However, these analytic methods impose the constraint that the continuous parameters have a Gaussian density.

The approach taken in this paper is to use stochastic methods, rather than analytic methods for reasoning with belief networks. Applications such as target recognition require the ability to construct models that combine discrete quantities such as ID, mode, and activity, with continuous quantities such as velocity and location. However, in many instances the continuous densities will not be Gaussian, and analytic methods are difficult to apply. Our approach is to use belief networks to construct the dependency model, then use Monte Carlo sampling from the model, in order to fuse information about both continuous and discrete quantities to improve target recognition. We will have a brief description of Monte Carlo reasoning, then illustrate the application of this approach to a scenario using the belief network shown in Figure 1.

3. MONTE CARLO COMPUTATION IN BAYESIAN ANALYSIS

In this paper we use a Monte Carlo computational procedure. With this approach both continuous and discrete variables as well as non-Gaussian distributions can be handled in a straightforward manner.

Bayesian analysis using Monte Carlo integration has been used successfully for difficult problems in traditional statistical analysis. We use the following notation to describe the methodology. (The symbol $p(Y|X)$ denotes the probability density of Y conditional on the value of X.)

D - the data

θ - the vector of parameters

$p(D|\theta)$ - the likelihood function. D is the data actually observed. The likelihood equals the "probability" of getting the data that was actually observed as a function of θ.

$p(\theta)$ - the prior probability density which represents knowledge and uncertainty about θ before seeing the data.

$p(\theta|D)$ - the posterior probability density which represents the knowledge and uncertainty about θ based on both the data and the prior.

To be consistent with certain reasonable axioms of inference and decision making, the prior and posterior probability densities and the likelihood function must be related according to Bayes' Theorem:

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{\int p(\theta^*)p(D|\theta^*)d\theta^*} \propto p(\theta)p(D|\theta) \quad (1)$$

The information in the data enters through the likelihood function, information external to the data through the prior. The posterior density represents the combination of the two. All inferences and decisions are based on the posterior density.

Bayesian inference, decision making and the computation of uncertainty limits require the evaluation of integrals of the form

$$I = \int h(\theta)p(\theta|D)d\theta \quad (2)$$

where the $h(\theta)$ are real valued functions. To evaluate these integrals by Monte Carlo we proceed as follows: θ_m, $m = 1, 2, \ldots, MC$, are generated independently from a multivariate density ("importance function") $g(\theta)$. Then, weights $W(\theta_m)$, $m = 1, 2, \ldots, MC$, are computed from

$$W(\theta_m) = \frac{p(\theta_m)p(D|\theta_m)}{g(\theta_m)} \quad (3)$$

To minimize computational error for a fixed Monte Carlo sample size, MC, $g(\theta)$ should usually be chosen to approximate the posterior density. This is known as importance sampling. It is more convenient however, and in some problems adequate, to choose $g(\theta)$ to equal the prior density, $p(\theta)$. We will do that in this paper. Importance sampling methods for the type of problems discussed in this paper are being developed.

Since $\{\theta_m, W(\theta_m) : m = 1, 2, \ldots, MC\}$ is a weighted sample from the posterior distribution, the Monte Carlo approximation to I is given by

$$\hat{I} = \frac{\sum_{m=1}^{MC} h(\theta_m)W(\theta_m)}{\sum_{m=1}^{MC} W(\theta_m)} \quad (4)$$

Further discussions of Bayesian statistical analysis using Monte Carlo integration and examples of different types of problems where this approach has been used can be found in Stewart [5,6,7,8].

The problem of interest in this paper, fusing information to monitor and track a developing situation and to classify a target, requires an extension of the Monte Carlo procedure just described. Here we wish to update our current prior distribution to a posterior distribution as new information becomes available and to update (evolve) that posterior distribution to a prior distribution at a later time for the next information update.

Let

$$\theta^n = (ID, M_n, A_n, \tau_n, V_n, L_n)$$

D_n = Data (information, measurement, report, etc) at time t_n

$$D_n^* = (D_1, D_2, \ldots, D_n)$$

The information update is as described above: An unweighted Monte Carlo sample $\{\theta_m^n\}$ from the prior distribution $p(\theta^n|D_{n-1}^*)$ at time t_n will be converted to a weighted sample $\{\theta_m^n, W_m\}$ from the posterior distribution, $p(\theta^n|D_n^*)$, by weighting each θ_m^n by the likelihood, i.e. $W_m = p(D_n|\theta_m^n)$. (This can be seen from equation (3) when $g(\theta)$ equals the prior.)

The evolution update, converting the weighted sample from the posterior, $p(\theta^n|D_n^*)$, to an unweighted sample from the prior, $p(\theta^{n+1}|D_n^*)$, uses the Monte Carlo technique of splitting. Each θ_m^n is split into J_m different $\theta_{m'}^{n+1}$ where J_m is a random variable with expected value equal to $W_m/\sum W_m$ times the desired Monte Carlo sample size and each θ^{n+1} is generated from $p(\theta^{n+1}|\theta_m^n)$. The resulting sample will be the desired unweighted Monte Carlo sample from the prior density $p(\theta^{n+1}|D_n^*)$. (Note that the $\theta_{m'}^{n+1}$ will not be statistically independent.)

All inferences and decisions are based on the current Monte Carlo sample. It is used to compute

Figure 2. Target path and Monte Carlo samples of location following every third line-of-bearing update.

probabilities, to display uncertainties, to predict the future and to evaluate possible decisions.

4. EXAMPLE

In this example the objectives are to track a target, infer target type (three possible types) and to infer current operational mode. Four types of sensor input are available:

1. Line-of-bearing contacts from two sensors, A & B, at different locations. There are 24 line-of-bearing updates alternating between the two locations. Each contact has a standard deviation of error equal to 2.5 degrees.

2. Accompanying each line-of-bearing contact is discrete information about activity, which is related to target speed.

3. Signature information that depends only on target type, reported as the most likely target type. In this example there are four such reports, one of which is in error.

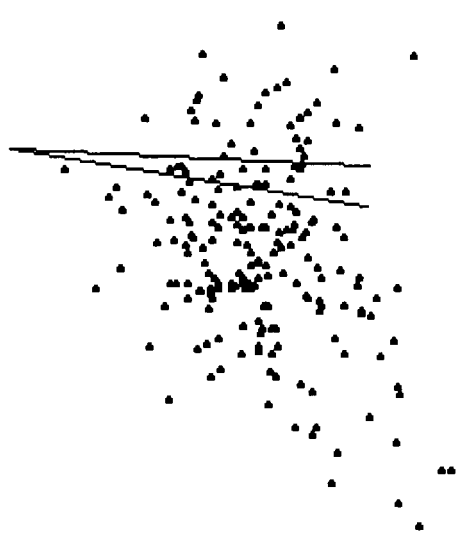


Figure 3. Distribution of location immediately before a line-of-bearing update.

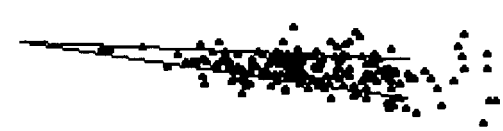


Figure 4. Non-Gaussian distribution of location resulting from the line-of-bearing contact.

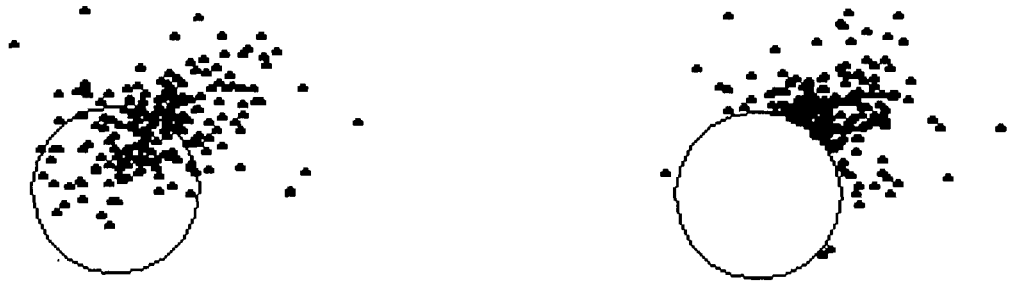


Figure 5. Before and after "negative information" update.

4. Negative information obtained when the area within a circle is thoroughly searched and no target is found.

Figure 2 shows the target path, where movement is from right to left. Also shown are Monte Carlo samples from the posterior distributions of location following every third line-of-bearing update. These Monte Carlo samples show the uncertainty in the location of the target. Each Monte Carlo sample can be thought of as a sample of possible target locations that are compatible with all the information available at that time.

Figures 3 and 4 illustrate the uncertainty distributions of location before and after a line-of-bearing update. The distribution in Figure 4 is non-Gaussian. (Contours of Gaussian densities are elliptical in shape.) Figure 5 shows the distributions of location before and after a "negative information" (target not in circle) update. Notice that the distribution after the update is very different than a Gaussian. One of the reasons for using the Monte Carlo methodology is its ability to handle and display non-Gaussian distributions

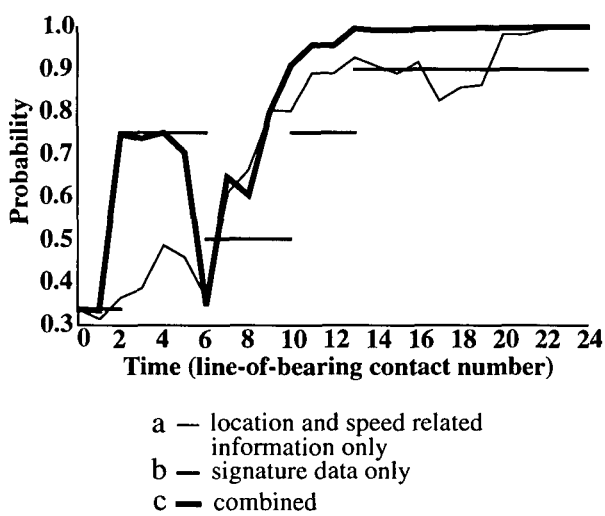


Figure 6. Probability of target type II (correct type)

Figure 6 shows the probability computed for target type II, which in fact is the correct target type, as a function of time. The three curves correspond to (a) using location and speed related information only, (b) using signature data only, and (c) using all available information. Curve (c) illustrates the benefits of fusing continuous and discrete information to improve target recognition.

Figure 7 shows the probability of being in operating mode one as a function of time. The actual time of transition from operating mode two to operating mode one is indicated.

5. CONCLUSION

The performance of a system for target tracking, recognition, and situation assessment is improved by using as much information as possible. Bayesian belief networks provide a convenient and straightforward way of modeling the relationships between uncertain quantities and of fusing information. The Monte Carlo computational approach offers great versatility and can accommodate models that would be difficult to handle with other approaches.

6. ACKNOWLEDGEMENTS

This research was supported by the Information Fusion; Uncertainty Reasoning; and Reliability, Experimental Design, and Discrimination projects of the Lockheed Independent Research Program. These results were presented at the SPIE Conference on Signal Processing, Sensor Fusion and Target Recognition, April, 21, 1992.

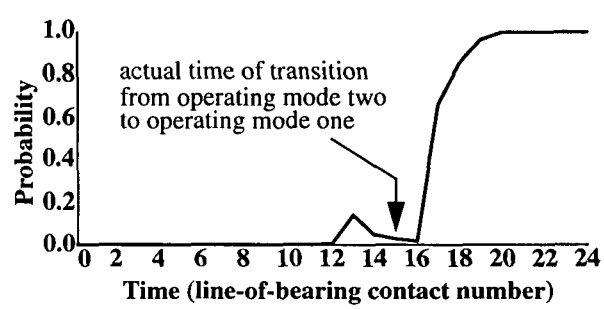


Figure 7. Probability of being in operating mode one

7. REFERENCES

1. S. L. Lauritzen and D. J. Spiegelhalter, "Local Computation with Probabilities in Graphical Structures and Their Applications to Expert Systems," *Journal of the Royal Statistical Society B*, Vol. 50, No. 2, 1988.

2. S. K. Andersen, K. G. Olesen, and F. V. Jensen, "HUGIN - a shell for building belief universes for expert systems," in *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 1080-1085, 1989.

3. R. D. Shachter and C. R. Kenley, "Gaussian Influence Diagrams", in *Management Science*, Vol. 35, No. 5, pp. 527-549, May 1989.

4. S. L. Lauritzen, "Propagation of Probabilities, Means and Variances in Mixed Graphical Association Models," Aalborg University Technical Report, April 17, 1990.

5. L. T. Stewart, "Bayesian analysis using Monte Carlo integration," LMSC/D560641, Lockheed Palo Alto Research Laboratories, 1977.

6. L. T. Stewart, "Multiparameter univariate Bayesian analysis," *Journal of the American Statistical Association*, 74, pp. 684-93, 1979.

7. L. T. Stewart, "Bayesian analysis using Monte Carlo integration - a powerful methodology for handling some difficult problems," *The Statistician*, 32, pp. 195-200, 1983.

8. L. T. Stewart, "Hierarchical Bayesian analysis using Monte Carlo integration: computing posterior distributions when there are many possible models," *The Statistician*, 36, pp. 211-219, 1987.

GREY LEVEL SEGMENTATION WITH SELECTABLE NUMBER OF DISCRIMINATION LEVELS USING HOPFIELD-LIKE NEURAL NETWORKS WITH CONSTRAINT SATISFACTION CRITERIA

Alberto DOMINGO, Javier SANTAMARIA
Aerospace Division
SENER Ingeniería y Sistemas, S.A.
Severo Ochoa s/n, PTM, 28760 Madrid
Spain

SUMMARY

A method is proposed for grey-level image segmentation which combines the benefits of an Artificial Neural Network (ANN) approach with the definition of simple, intuitive constraints that govern the network behaviour. Several schemes are introduced for the definition of the constraints, as well as for the network potential initialization criteria. A new neural network model, derived from the Hopfield Neural Net, is proposed, adapted to the defined constraints. Obtained results are shown and compared with those achieved with classical algorithms.

Image Segmentation Neural Networks
Constraint Satisfaction Hopfield Networks

1. INTRODUCTION

Image segmentation tasks, though difficult to precisely define, are usually understood as procedures to cluster each pixel within an image into one of a number of different categories. A desirable goal for such procedures is to yield segmented images where the resulting clusters are grouped into well defined regions, those regions are uniform (in the sense that they avoid single, mis-classified pixels, usually caused by the presence of noise and/or fault detectors), and they represent the contents of the original image in such a way that this information can be passed along on to the next stage of the image process.

The pixel clustering into regions, as performed by any segmentation algorithm, must be carried out on a classification scheme based on some attribute [1], such as the pixel grey level, or the texture.

Though texture segmentation has proven to be very useful in the case of infra-red images [2], and as a way to complement those results obtained from grey-level segmentation,there is no precise definition of what constitutes a texture (texture models do not necessarily adapt well to some types of natural scenes nor do they provide much intuition about human texture perception [3]).

To perform an image grey level segmentation, some transformation of the original intensity information of each pixel must be used, given that the intensity of the image may not always be a good representation of the underlying physical variation of the scene [4]. In this sense, statistical techniques such as modelling image data with Markov Random Fields (MRF's) have been commonly applied, using a Maximum A Posteriori (MAP) criteria to cluster the pixels. However, the minimization of the energy function derived is exceedingly difficult, and it is usually a non-convex function, and the search space is much too large to perform a direct minimization [5].

While this approach has been shown to yield good results in many cases, specially in the presence of noise, in the case of natural scenes the segmented image lacks both visual coherence and homogeneity, particularly if the image to be segmented presents a complex set of contents, textures and illumination conditions, as most real images do.

Furthermore, the MRF approach seems to be better suited to segmentation to only two-level regions (that is, segmenting targets from the background). The performance decreases strongly in the case of segmenting to multilevel regions, mainly in the combined presence of both noise and overlapping targets, or when leakage zones are present.

In this paper we propose a set of empirical constraints that determine to which of the available segmentation levels each pixel should belong. These

constraints depend not only on the brightness of the pixel, but also on the region to what the neighbouring pixels belong, so as to minimize the probability of a mis-classification based on the alteration of the grey-level of the pixel caused by noise or a fault detector.

The set of constraints are joined to form an energy function representing the difference between the optimally segmented image and the presently achieved one. When minimized, they should yield a resulting image good enough to perform further vision processes (such as detection and classification), regardless of the contents of the original image and the vision conditions, and allowing manual selection of the desired number of levels.

For the minimization part of the algorithm, we have selected a severely modified version of the Hopfield Neural Network model, where several neurons are assigned to each image pixel. This network architecture has been proven to achieve suitable results in a short number of iterations of the energy minimization algorithm (commonly 40 to 50 "epochs"), and is easily hardware implementable so as to achieve real-time performances. Speed of convergence is mainly determined by the number of selected classes, and not by the particular contents of the image.

Evaluation of several strategies to minimize the energy function, to initialize the neuron input biasses and to update the neuron potentials have also been considered.

2. CONSTRAINT SATISFACTION FOR MULTILEVEL, GREY LEVEL SEGMENTATION

The energy function to be minimized has been formulated in terms of imposing certain spatial restrictions to the segmentation label assigned to each pixel within the image. These restrictions are the so-called *constraints*.

Let us assume that, for every pixel (x,y) in the image, we assign to it a value $U_{k,x,y}$, called *potential*, that represents the probability of that pixel belonging to the k_{th} class of pixels (that is, the higher this potential, the greater the probability that this is a class *k* pixel).

If there are *n* possible different classes, then there will be *n* different U_k values per pixel, ranging from U_0 to U_{n-1}. We will later assign *n* neurons (representing such potentials) per pixel in the original image.

A combination of constraints has been defined to calculate, in each moment, the value of any of the *k* possible potentials assigned to a certain pixel. While the number of different constraints defines the degree of precision in the task to be accomplished (in this case, grey-level segmentation), the need to keep the number of them relatively low may arise from the ability of the network to converge in a reasonable amount of time [6]. Those constraints, defined in terms of spatial restrictions, have been formulated as follows:

CONSTRAINT 1: avoidance of mis-classified pixels within a region.

In a normal scene, the probability of an isolated pixel within a region, classified as belonging to another region is very low. The first constraint tries to ensure that the class already assigned to the neighbours of a pixel is taken into account when deciding the class to which the pixel belongs. That is, if all neighbouring pixels are assigned to class "*i*", the probability of the pixel belonging to any other class should be lowered, while the probability of being assigned to the same class should be increased.

This can be formulated as a partial contribution to the global potential variation, given by the following expression:

$$U^{1}_{k,x,y}(t+1) = \sum_{x'y' \in W} \alpha_{x,x',y,y'} \, U^{G}_{k,x',y'}(t) \qquad [1]$$

where $U^G(t)$ represents the kth potential assigned to pixel (x,y) at time *t*, and $U^1(t+1)$ represents the contribution to that potential by constraint 1 at time t+1. *W* contains all the pixels within the excitatory neighbourhood of the (x,y) pixel defined for this constraint, and $\alpha_{x,x',y,y'}$ is the excitatory factor as a function of the distance between the two pixels.

While the excitatory factor can be as complex as desired, it has been found that a step function in terms of the Manhattan distance between the two pixels is sufficient for our purposes.

CONSTRAINT 2: clustering regions similar to the pixel grey-level.

This constraint intends to cluster pixels according not only to their spatial location, but also to their brightness, so that different regions of the same class (in different positions of the scene) imply similar brightness in the original image.

This constraint can also be formulated as a contribution to the global potentials assigned to each pixel. Of all the available k potentials (segmentation levels) per pixel, only one will receive a positive feedback (increase), while the others will receive a negative one (decrease). The one receiving a positive feedback is determined by assigning a central level to each possible segment, and finding the one which achieves the lowest difference with the normalized grey level of the pixel. This can be expressed as:

$$U^2_{k,x,y}(t+1) = \left(pix_{x,y} + \frac{k \cdot (n+1) - 2}{k \cdot (n+1)} \right)^2 \qquad [2]$$

where $U^2(t+1)$ represents the contribution to that potential by constraint 2 at time $t+1$, $pix_{x,y}$ is the original, normalized, grey-level value of the pixel at location (x,y), and n is the number of segmentation levels. In this case, the neighbourhood of a neuron (k,x,y) is composed of the rest of neurons belonging to the same pixel, (k',x,y).

CONSTRAINT 3: Inter-inhibition of same-pixel potentials.

The third constraint tries to ensure that only one of the n available potentials per pixel will increase its value well above the rest, then imposing only one of the possible class assignments to the pixel.

The formulation of the third constraint can be implemented as follows:

$$U^3_{k,x,y}(t+1) = \beta \sum_{j=0}^{n-1} \left[U^G_{k,x,y}(t) - U^G_{j,x,y}(t) \right] \qquad [3]$$

where $U^3(t+1)$ represents the contribution to the neuron potential from constraint 3, and β is a weight constant.

3. POTENTIAL UPDATE RULES. NETWORK DYNAMICS

Once the three contributions to the neuron potential corresponding to each of the three constraints have been calculated, the expression is globalized to determine the extent of the variation that it causes on the global neuron potential.

The potential update rules determine the influence of both the classification constraints and the recent history of class assignment for each pixel, along with the present classification obtained for every surrounding pixel within a well defined neighbourhood.

On each iteration of the globai potential estimation, obtained classification value for each pixel should converge to a final one, where further iteration should, ideally, represent no variations in the segmented resulting image. Convergence, then, should be understood as the situation where pixel classifications remain stable through subsequent potential updates, rather than the condition of stable potential values.

It has been found that this convergence is only achieved if several conditions are simultaneously satisfied (causing the network energy to always decrease, regardless of the possible pixel classification variations according to the potential update rules). Among them, constraint influences over the potential should be both upper and lower bounded and mutually normalized, even if weight factors are used to increase the effect of any one of them over the others.

It has also been found that using a threshold function with the values yielded from the constraints influences over the global potential both improves the segmentation results and reduces the convergence time. To implement such a threshold, a sigmoid function has been used, denoted by:

$$F(x) = \frac{2}{1+e^{-(\delta x)}} - 1 \qquad [4]$$

where $F(x)$ is the thresholded value of the input x, bounded between -1 and 1, and δ is a gain term, initially set to a value such as 0.1, which can be slowly increased during the operation of the network.

The global function that defines the neuron potential for each neuron in the network can, then, be defined

as follows:

$$U^G_{k,x,y}(t+1) = U^G_{k,x,y}(t) - \Delta t\,[\,C_1 \cdot sig(U^1_{k,x,y}(t)) + C_2 \cdot sig(U^2_{k,x,y}(t)) + C_3 \cdot sig(U^3_{k,x,y}(t)) - U^G_{k,x,y}(t)\,] \quad [5]$$

where $U^i(t)$ represent the evaluation of constraint "i" at time t, C_i is the influence or weight of constraint "i" over the neuron potential update, and Δt is a time constant that represents the system memory and allows slow convergence while avoiding strong potential oscillations. It also serves as a limiting factor which guarantees that potentials will not overflow through many iterations of the algorithm, even if they continue growing.

The importance of properly selecting the weight constants resides in their direct influence over the resulting image, that is, over which of the existing constraint effects will eventually take control over the others. A trade-off between the three (or more) constants allows a significant image segmentation, avoiding nonsense results such as segmentation of the image into one single region (first constraint supremacy), segmentation of each pixel to the region closer to its grey-level (second constraint supremacy) or image segmentation to initialization seed (supremacy of the third constraint).

4. PROPOSED HOPFIELD-LIKE NEURAL MODEL

We now can identify the expression in expression (5) as an objective function whose global extreme, maximum for the potential corresponding to a minimum of the system energy, must be found. The segmented image will be formed by taking only one of the potential values assigned to each pixel, specifically the one with the highest value (in a "winner-takes-all" strategy), and assigning the pixel to that class.

Once assumed that the expression is a good representation of the segmented image, a relaxation method must be found to minimize the global energy solution from the objective function in expression (5), as efficiently as possible [7]. However, most conventional and classical minimization methods are not valid when applied to non-convex functions such as this one, with very large search spaces.

Hopfield neural networks, where a single neuron is assigned to every pixel, is connected only to its neighbour and has negative self feed-back [8], have been shown to solve complex minimization problems with relatively good results.

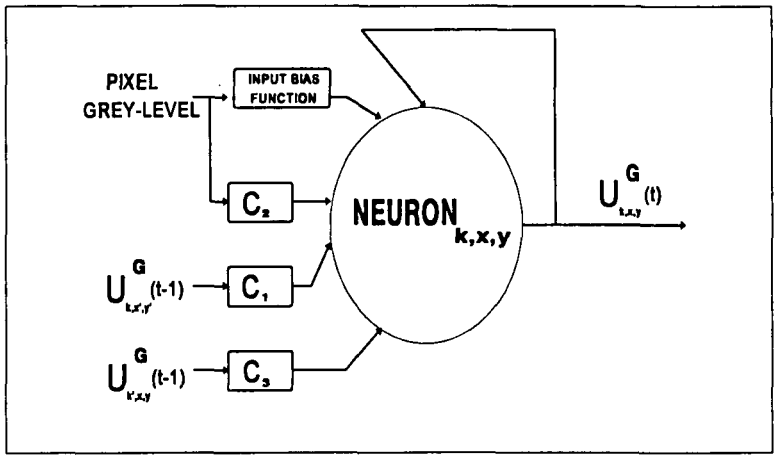


FIGURE 1. Schematic representation of each neuron of the ANN. Each of the constraints is shown as a black box, C_i, whose input is a combination of neighbourhood potentials for each of the constraints.

The advantage of such methods is that the approach is highly robust and noise insensitive, even when the input data is incomplete or defective. Some other derived benefits are the ability to be easily hardware implemented for real-time operation through parallel and distributed computing, and their fault tolerance and graceful degradation [9].

In our model, the Hopfield neural network has been modified so as to directly map the potential function as defined in expression (5). K neurons are assigned to each pixel, each representing the potential for the k class of the pixel. Thus, there are as many as MxNxK neurons in the network, if MxN is the image size in pixels.

Each neuron in this network is a modified version of those corresponding to a Hopfield ANN, in the sense that the neuron potential depends on the recent history of the potential, the potentials assigned to neighbouring pixels and the input bias, but with a different, slightly more complex relationship between these magnitudes. A graphical representation of such behaviour is shown in Figure 1.

The potential of the neuron is computed from the last obtained value, which represent the recent history of the network, and from a modified version of the original pixel grey level. The rest of the influences,

derived from the different defined constraints, affect the computation of the potential grouped in as many different sites as constraints are defined. Each site groups a number of neurons which represents the neighbourhood of influence for each constraint and, so, can be viewed as a function of certain neuron potentials affected by a weight constant.

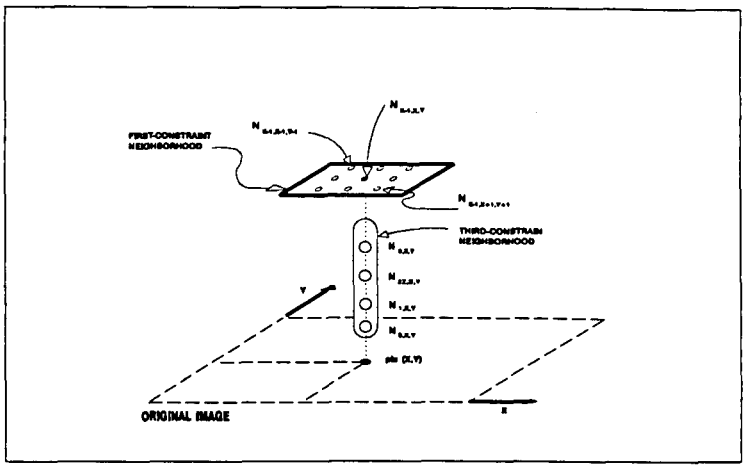


Figure 2. Neighbourhood representation for the $N_{k-1,x,y}$ neuron of the network. Neighborhood for the first constraint has been selected to be of size 3x3.

In our model, site 1 represents the first of the constraints (avoidance of mis-classified pixels within a region), computed from neighbouring potentials belonging to the same "*k*" class, and affected by weight constant C_1. Site 2 represents the second constraint (clustering regions similar to their pixel grey-level), which depends only on the pixel grey-level and the neuron class. Finally, site 3 stands for the last of the previously defined constraints (inter-inhibition of same-pixel potentials), which is computed only from potentials belonging to all the neurons assigned to that pixel, affected by weight constant C_3.

The definition of the scope is independent for each of the neighbourhoods. For the first constraint, the size of the neighbourhood must be chosen as a trade-off between a large value, which avoids small-isolated regions but with distortionate sharp edges, and small values, which keep clean edges while being less efficient with single, mis-classified pixels.

For the second constraint the definition of a neighbourhood is meaningless, given that the potential influence is derived solely from the pixel grey-level and the class that the neuron is assigned to. Finally, for the third constraint, it is obvious that the neighbourhood is composed of all the neurons assigned to the pixel (one for each segmentation class), excepting the one whose potential is being modified.

Figure 2 shows the neighbourhoods corresponding to the first and third constraints for neuron $N_{K-1,x,y}$, assigned to the pixel of coordinates *(x,y)*. The second constraint neighbourhood is limited to the pixel value and the class of the neuron, *K-1*.

The dimensions of the neighbourhood, especially the one defined by the first constraint, determine the effect of pixel clustering changes throughout the entire image. Though only changes in those pixels included in the neighbourhood directly affect the computation of the potential for the central neuron during the present iteration, any change propagates itself though computations in adjacent neighbourhoods. After *N/(2m-1)* iterations of the minimization algorithm (where N is the image side size in pixels, and m is the neighbourhood side size, also in pixels), this change, though very attenuated, is translated to the opposite side of the image.

5. EFFECTS OF NEURAL POTENTIAL INITIALIZATION

It has been shown that the strategy used for the neurons potential initialization plays a very important role in the number of iterations which are needed to achieve the segmented image, in the evolution of the segmentation process and in the visual aspect of the resulting segmented image [10].

For this work, several strategies werw tested for the initialization of the neuron potentials, which are detailed as follows:

- ✓ Initialization of potentials to a value proportional to the pixel normalized grey level. This method accelerates the convergence of the minimization process considerably, achieving a presegmented image in less than 10 iterations, but presents inertia in the sense that noise corrupted pixels remain mis-classified for as many as 40 to 50 iterations.

- ✓ Initialization of potentials to a low, random value, uniformly distributed around the statistical mean of the normalized input data range. This strategy achieves the best results in both visual appearance and evaluation figures of all the tested methods, but requires many more iterations of the minimization algorithm than the other strategies (usually between 100 and 130 iteration for a 6 level segmentation).

- ✓ Initialization of potentials to a presegmented value obtained by dividing the grey-scale range into the number of desired discrete levels, and assigning each pixel to the center value of the grey-scale segment to which it belongs. This method has achieved good results only with synthetic images, but not with natural scenes.

Two other strategies have been tested, as a combination of the three introduced previously. They are as follows:

- ✓ Initialization of potentials to a presegmented value, which is then increased with a small noise term.

- ✓ Initialization of potentials to a value proportional to the pixel normalized grey level, to which is then added a small noise term.

From all the proposed methods, the first one usually achieves good enough results in a reasonable number of iterations (usually between 50 and 60). If the incoming scene presents a high level of noise corruption, or contains many details manifested as very small, insignificant regions, the last initialization method (value proportional to the pixel normalized grey level, plus a small noise term) has proven to achieve better results, though it usually requires from 15 to 25 iterations more to converge to good results.

Figure 3 shows a comparison of the three main (non-combined) strategies for potential initialization, in terms of percentage of mis-classified or un-classified pixels versus the number of iterations of the minimization algorithms, for the test image.

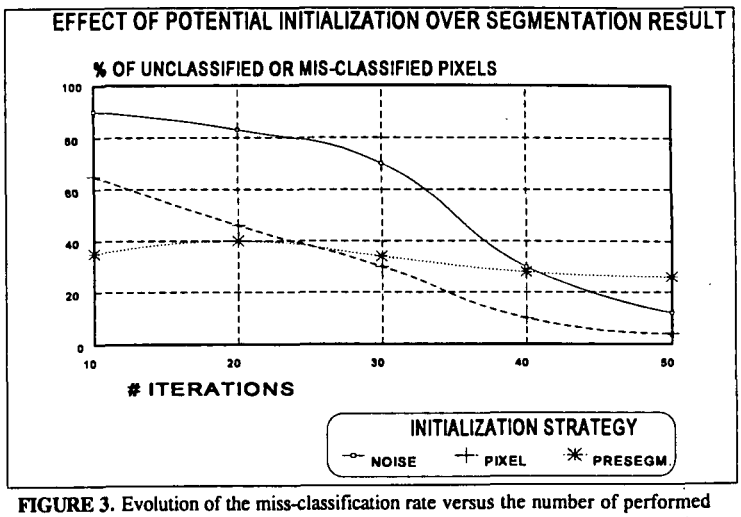


FIGURE 3. Evolution of the miss-classification rate versus the number of performed iterations of the segmentation algorithm, for three different potential initialization strategies.

6. OBTAINED RESULTS

Figure 4 shows a comparison of results obtained by segmenting an image with different methods. The first row of images include the original scene, with 256 different input levels, and the segmented image obtained by two classical methods: a Perkins segmenter and a pyramidal segmenter.

The second, third and fourth rows show the same image segmented by the method proposed in this paper for obtaining 3, 4 and 6 segmentation regions (depending from the parameter K). Each of these rows shows the evolution of the achieved result, in terms of iterations of the algorithm.

For all the ANN segmentations shown in the figure, neuron potentials have been initialized to a value proportional to the pixel normalized grey level (the first of the proposed initialization strategies), the time constant of the potential update circuit (Δt) has been set to 10^{-4}, and the neighbourhood size for the first constraint has been set to 3. Some improvements have been achieved by updating the neuron potentials following an asynchronous (random) scheme, rather than a sequential one.

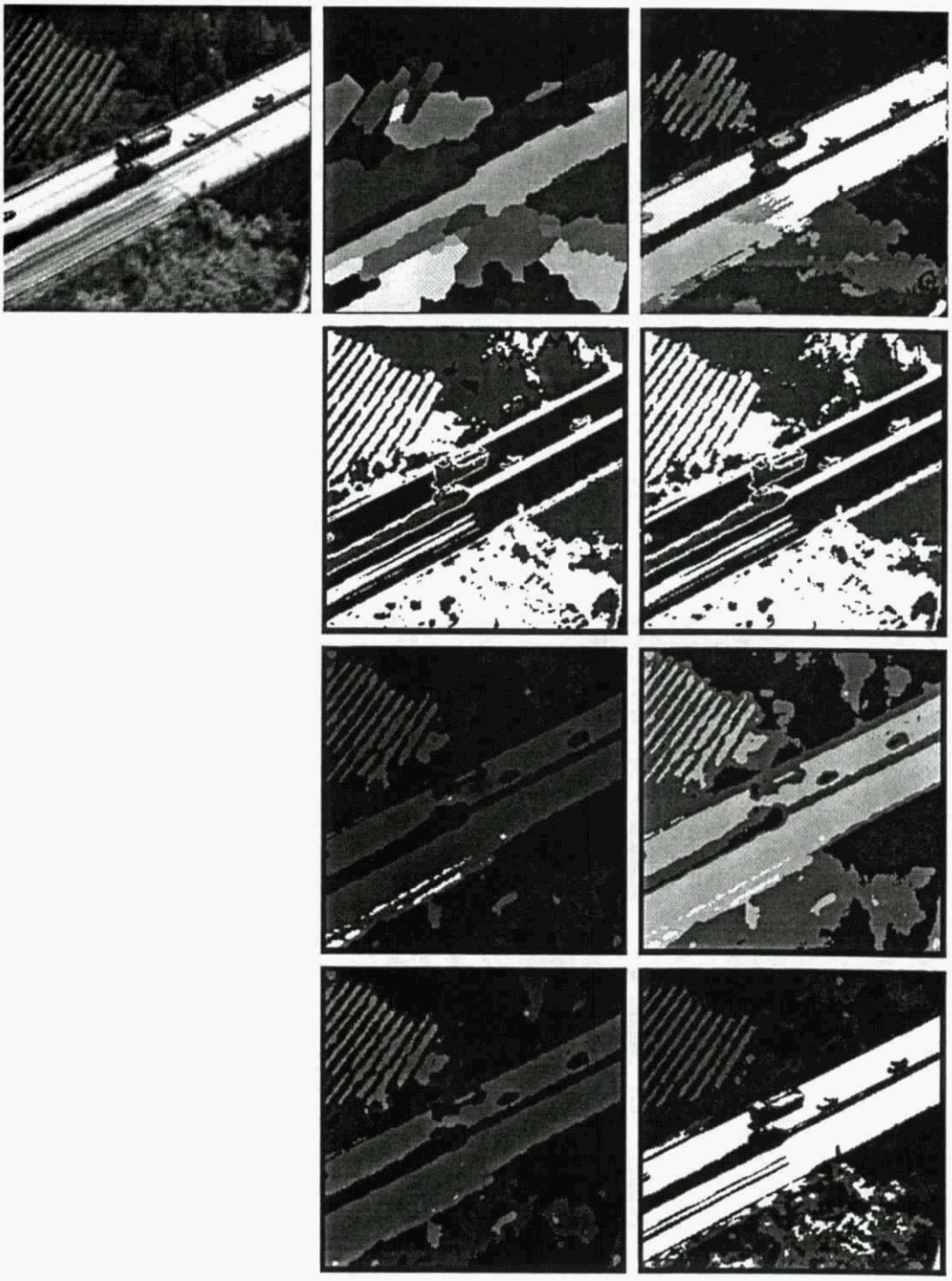


FIGURE 4. From left to right, first row, original image to segment (256 levels), segmentation result using a Perkins segmenter, and segmentation result using a pyramid segmenter. Second, third and fourth rows show the results of segmenting the same image using the neural network as defined in the text, with three, four and six segmentation levels, respectively, being the left one showing the segmented image after 25 iterations of the algorithm, and the right one after 50 iterations.

7. CONCLUSIONS

The proposed method for grey-level image segmentation using Neural Networks has been shown to achieve good results with nearly all of the tested images, both synthetic and natural scenes. Its main advantage is the independence of the scene contents and nature, in the sense that the algorithm does not necessarily have to be optimized for the expected incoming scenes. Both Perkins and pyramidal segmenters require several input parameters, usually derived from the contents of the image or the image type.

Another advantage of the proposed methods is its capability for allowing the selection of the resulting number of segmentation levels, thus allowing the user, in manual intervention operations, to adapt the algorithm to the required level of detail.

Also, an architecture such as the one suggested here is very susceptible to hardware integration, making use of both paralelization and distributed computing, thus providing a direct path for attempting real-time image segmentation.

Obtained results have been shown to be very promising, most of the time both visually and numerically better than those obtained by different classical segmentation tools. However, further research mustbe carried out, both for the optimization of obtained results and for the definition of new constraints (such as a Markov Random Field modelling) which could even improve the results shown in this paper.

8. ACKNOWLEDGEMENTS

The authors wish to thank Dr. Octavio Nieto Taladríz (Universidad politécnica de Madrid) and Mr. Carlos Miravet (SENER Ingeniería y Sistemas) for their support and collaboration during the preparation of this manuscript.

9. REFERENCES

[1] Wei-Chung Lin et Al. *Constraint satisfaction neural networks for image segmentation*. Pattern Recognition, Vol. 25, No. 7, pp 679-693, 1992.

[2] Desmouceaux, J., Derycke, N. *Discrimination de textures infrarouges par réseaux de neurones*. Revue technique Thomson-CSF, Vol 22, N°4, December 1990.

[3] Manjunath, B.S. *Perceptual Grouping and Segmentation using Neural Networks*. USC-SIPI Report #188. December 1991.

[4] Tarr, G.L. *Multi-Layered Feedforward Neural Networks for Image Segmentation*. AFIT/DS/ENG, December 1991.

[5] Codrington, C.W., Tenorio, M.F. *Experiments in Target Segmentation Using Markov Random Fields*. Remote Sensing Reviews, 1992.

[6] Malki, H.A. *Image Segmentation using Multilayer Neural Network*. International Joint Conference on Neural Networks, Vol 4, pp 354-360. 1992

[7] Yungsik, K., Rajala, S.A. *Image Segmentation using an annealed Hopfield neural network*. SPIE Vol. 1766 Neural and Stochastic Methods in Image and Signal Processing, pp 333-342, 1992.

[8] Gosh, A., Pal, N.R., Pal, S.K. *Image Segmentation using a Neural Network*. Biological Cybernetics, Vol 66, pp 151-158. 1991.

[9] Jones, M.J. *Using Recurrent Networks for Dimensionality Reduction*. Master of Science Thesis. DEECS, Massachusetts Institute of Technologie. September 1992.

[10] Youngjk Lee, Sang-Hoon Oh, Myung Won Kim. *The effect of Initial Weights on Premature Saturation in Back-Propagation Learning*. IJCNN, Orlando, 1991.

SCENE/OBJECT CLASSIFICATION AND SEGMENTATION USING MULTISPECTRAL DATA FUSION

L.E. Lazofson, T.J. Kuzma, H.C. Choe, E.B. Preston and J.D. Chovan

Battelle
505 King Avenue
Columbus, OH 43201 USA

1. SUMMARY

Near-simultaneous, multispectral imagery of ground target and background signatures were collected over a full diurnal cycle in visible, infrared, and ultraviolet spectrally filtered wavebands using Battelle's portable sensor suite. The imagery data were coregistered and processed using a variety of classical statistical algorithms, artificial neural networks and data clustering techniques to classify pixels and objects in the imaged scenes. Imagery collected at different times throughout the day were employed to verify algorithm robustness with respect to temporal variations of spectral signatures. Sensor fusion hardware was also designed and built to accompany the sensor suite to provide real-time pixel classification capability. In addition, research is being performed to advance the state of the art using differential absorption lidar as an active remote sensing technique for spectrally detecting, identifying, and tracking hazardous emissions. These investigations support a wide variety of multispectral signature discrimination applications, including automated target search and landing zone detection. Battelle's sensor suite has also been used to quantify the extent of natural gas leaks and record the thermal characteristics of smoke grenades.

2. INTRODUCTION

The spectral distribution of energy reflected and/or emitted from an object is a unique characteristic that may be exploited to discriminate the object of interest from the local background. Multispectral data fusion techniques can be used to solve a variety of classification and discrimination problems. Battelle's portable multispectral sensor suite has been used to collect simultaneous multispectral scene imagery in several data collection episodes. Implementing tailored data fusion algorithms, Battelle has processed some of the recorded imagery to solve classification problems including automatic detection of ground targets and location of aircraft landing zones. Upcoming investigations using active lidar technology include the detection and identification of gaseous emissions for environmental monitoring and chemical/biological weapons treaty verification.

Algorithms implemented in the study included unsupervised maximum likelihood and fuzzy clustering algorithms along with Multilayer Perceptron and Learning Vector Quantization (LVQ) artificial neural networks. Supervised clustering of the data was also used. The algorithms were tailored to perform pixel-level classification of scene imagery using simultaneous multispectral measurements covering the UV, visible, near-IR, MWIR, and LWIR wavebands. Imaged outdoor scenes were comprised of tactical targets, buildings, roads, runways, and vegetation. Imagery were spatially coregistered with an RMS error on the order of 0.5 pixels. Outdoor scenes processed through the data fusion algorithms were displayed with artificial color to demonstrate algorithm classification performance.[1]

Waveband saliency analyses were performed to determine which spectral bands contained the bulk of the discriminating information for discerning objects in the scenes. These analyses help determine the optimum subset of wavebands for discriminating spectral signature characteristics for an object of interest. Histograms and scatter plots of the multispectral data were generated with the Geographic Resources Analysis Support System (GRASS).[2] For system implementation, the potential benefits of waveband saliency analyses include a reduction in the number of system

sensors, ultimately minimizing cost, weight, complexity, and processing requirements.

To further verify classification robustness, algorithms were tested on outdoor scene imagery recorded over broad periods of time throughout the day. Results were excellent, indicating that scene classification is achievable despite temporal signature variations.

3. MULTISPECTRAL IMAGERY COLLECTION AND PROCESSING

Battelle's portable sensor suite consists of two high-resolution, high-sensitivity thermal imagers and five charge-coupled-device (CCD) cameras. A variety of spectral filters and telescopic lenses accompany the sensor suite to enable rapid system reconfiguration for supporting many unique imaging and data collection requirements. The attached photograph (Figure 1) of the multispectral sensor suite shows one system configuration example using telescopic lenses on the thermal imagers for long distance, outdoor imaging. However, the sensor suite may be easily reconfigured to support close-up multispectral imaging applications. With its collection of spectral filters, the Battelle sensor suite images over a large selection of wavebands in the visible, infrared, and ultraviolet regions of the spectrum. The CCD cameras generate 640 x 480 pixel images while the thermal imagers spatially sample a coarser resolution of 207 x 260 pixels in the MWIR and 207 x 344 pixels in the LWIR.

During two outdoor data collection episodes, Battelle's multispectral sensor suite was positioned looking downward from a 110-foot tower location. Imagery were then recorded over a period of several days. Pixel-level fusion was accomplished by feeding the coregistered multispectral pixel intensity measurements into data fusion algorithms.

In the initial multispectral data collection episode conducted in June 1992 at Wright-Patterson AFB, Ohio, Battelle recorded scene imagery of a mobile missile launcher amidst a background containing roads, buildings, trees, and grass over a full diurnal cycle. Figure 2 depicts a 35 mm photograph of a scene containing the mobile missile launcher and ground clutter imaged simultaneously in six wavebands by the sensor suite. The multispectral imagery of this scene, processed through an artificial neural network, is displayed in Figure 3 with artificial colors designating the different classes of objects (target, road, building, trees, and grass) identified by the network algorithm. To further verify robustness of this classification algorithm, it was tested on imagery recorded over broad periods of time throughout the day. Results were excellent, indicating that scene classification is achievable despite temporal signature variations.

Ground truth data covering a variety of meteorological parameters and site thermocouple measurements on targets and background were collected. This critical data, in conjunction with the multispectral imagery, provides a valuable record of weather and atmospheric conditions to support future modeling and analysis efforts using the imagery.

There are clear tactical advantages in implementing a data fusion algorithm to automatically perform multispectral data fusion for detecting potential targets. Chaotic and obscured operational environments pose a challenge for human observers looking for targets and threats. Even in more placid scenarios, system operators may become task- and data-saturated to the point where potential threats go unnoticed. Automatic target detection/cuing can leverage an operator's effectiveness by reducing workload and increasing probability of target detection.[3] Battelle's investigation implementing sensor fusion of multispectral target and background signatures supports the concept of automated target search.

In conjunction with the Federal Aviation Administration's Runway Detection Program, Battelle collected additional multispectral imagery in a separate measurement episode and processed the data using analysis and fusion techniques to detect a runway at Wright-Patterson AFB, discriminating it from other objects in the area using spectral characteristics. Figure 4 shows a 35 mm photograph of a scene imaged in multiple wavebands consisting of a runway, roads, vegetation, and tactical targets at approximately 3 km. Figure 5 displays, in two artificial colors, the binarized result of a data fusion algorithm merging

the multispectral data to segment, or detect, the runway.

4. DATA CLUSTERING ALGORITHMS

As a baseline, the study began with an investigation using an unsupervised maximum likelihood statistical algorithm for clustering the multispectral data. Displaying the clustered pixel classes with artificial color indicated that fusion of data from all six wavebands successfully distinguished the classes of interest. Using only the visual and near-infrared bands, the camouflage-green mobile missile launcher was difficult to discriminate from background trees. Employing data from the two thermal bands, the clustering algorithm confused the mobile missile launcher with paved road, but successfully separated vegetation from man-made objects. Combining the data from all six wavebands successfully clustered the classes of interest, distinguishing target pixels from background pixels as well as differentiating between vegetation and man-made objects. Pruning combinations of sensor inputs indicated that the green-filtered visual band and the 8-12 micron thermal band together contained most of the key information for distinguishing the classes. Clustered classes were displayed with artificial color.

A fuzzy clustering algorithm was also applied to the multispectral scene imagery containing the mobile missile launcher target. Similar to other techniques, this algorithm carved different object regions within the multispectral feature space, except that it allowed for "overlapping" class possibilities among the data clusters. In other words, a specific point in the multidimensional pattern recognition feature space may have been simultaneously designated as belonging to more than one cluster or object class, with weighted possibility factors pertaining to the "degree of belonging" to each class. However, the researcher must ultimately select a defuzzifying threshold to apply when making a final classification decision.

In addition to clustering algorithms, a supervised classification algorithm was applied to the multispectral data set obtained during the first imagery collection episode. The supervised algorithm generated a pixel intensity histogram in each waveband for pixels sitting on a user-designated object (i.e., target) within an imaged scene. Thresholds were established near the tails of the histograms. Pixels were then classified as target pixels if the intensity values in each waveband fell within established thresholds on the histograms.

5. NEURAL NETWORK ARCHITECTURES AND TRAINING RESULTS

The scene employed in the initial investigation consisted of a mobile missile launcher parked on a grassy area in front of a grove of trees (Figure 2). Additional objects within the scene consisted of paved areas and buildings. Imagery from all measured spectral bands were spatially coregistered, or aligned, with an RMS error on the order of 0.5 pixels. To assess classification robustness, the tailored neural networks were trained and tested on scenes imaged at different times of day. Learning coefficients, number of computational nodes, and node transfer functions were varied to optimize performance of a Multilayer Perceptron network and a modified Learning Vector Quantization (LVQ) network employing "conscience" with added training noise. Both architectures trained successfully, converging within several thousand training iterations to a 98% pixel classification accuracy on a separate set of test data. As shown in Figure 3, trained network outputs of classified pixels from a full scene were displayed with artificial color to pictorially convey the near-perfect segmentation and classification of the five class regions (mobile missile launcher, buildings, paved road, trees, and grass). Network architectures were developed and tailored with the NeuralWorks Professional II/PLUS software package by NeuralWare, Inc.[4] Image pixel data were processed with the Geographic Resources Analysis Support System (GRASS), a Geographical Information System (GIS) possessing some image processing capabilities.[2]

The initial training data set consisted of 750 spectral vectors (pixels) comprised of 150 vectors for each of the five classes. The corresponding test data set contained 250 vectors including 50 pixels for each class. The trained neural networks were later used

to classify all 233,000 pixels imaged in a full scene, of which approximately 5600 were target pixels. Results were excellent, indicating that the trained neural networks served as robust target segmentation algorithms. With only limited training, approximately 90% of the on-target pixels were correctly classified and over 99% of the background pixels were correctly classified as not being on target. The contiguity of the on-target pixels offers an advantage in that a few misclassified target pixels will not detract from the target segmentation display. Also, image processing techniques, such as windowed convolution with a median filter, may serve to "clean up" many of the misclassified pixels.

To further verify the classification robustness of the LVQ neural network for target detection, the network was trained and tested on imagery recorded over broad periods of time throughout the day. Results were excellent, indicating that automatic target detection is achievable despite target and background signature variations over time. Imagery collected at dawn and dusk were the most difficult for the neural network classifier, apparently due to the rapidly changing signature dynamics at crepuscular times of the day. Adding a time-of-day input to the neural network, along with the spectral intensity measurements for each pixel, significantly improved the classification accuracy.

Waveband saliency analyses were performed to determine which spectral bands contained the bulk of the discriminating information for discerning targets amidst cluttered background. These analyses investigated the waveband combinations providing complementary information for finding targets. Spectral inputs to the networks were ultimately reduced to include only the green, near-infrared, and LWIR wavebands. For system implementation, the benefits of reducing the number of system sensors include minimized cost, weight, complexity, and processing requirements.

6. HARDWARE IMPLEMENTATION

To accompany the passive electro-optical sensor suite as it collects multispectral imagery, Battelle designed and built a real-time digital data collection and fusion processing system. The system incorporates state-of-the-art high speed video, computer graphics, memory and programmable logic technologies into a hardware architecture that enables data fusion algorithms (i.e., look-up tables derived from trained neural networks) to be executed at video pixel rates. The system classifies all pixels in an imaged scene by combining inputs from three separate cameras imaging in different wavebands. The classification result is visualized in a video format on a full-color, nine-inch, active matrix liquid crystal display (LCD). The sensor fusion board is an I/O addressable standard full-length PC/AT bus-compatible circuit card assembly comprised primarily of VLSI surface mount components. The host processor is used to configure and program the hardware functions.

The sensor fusion board accepts three line-locked composite video inputs from three spatially registered cameras. These inputs are digitized to eight bits each and are processed to form a 24-bit-per-sample input to the video classifier and the video output select circuitry. The classifier uses 23 bits to produce an eight-bit data word containing the two possible four-bit classifications. The four-bit classification results in up to 16 distinct membership classes. The video output select circuitry receives the eight-bit classifier output data word, 24-bit pixel input data (32 bits total) and user configuration data. Using these inputs, it selects one eight-bit video input and the final synchronous four-bit classifier data. The color space converter performs 3-by-3 matrix multiplication at pixel rates to accomplish 24-bit-to-24-bit color space coordinate conversion. This device is used to create transparent pseudocolor overlays on grayscale image data. The video RAMDAC converts the 24-bit color output of the color space converter into an RS-170 compatible RGB video output for display. The entire classification process is pipelined, with all stages operating simultaneously. The process does not require input video storage other than the use of registers required to temporarily store data at each stage of the pipeline. As a result, the overall input-to-output latency is just a few pixels, and the output is line-locked to the input.

The sensor fusion board is designed to operate at a top pixel rate of 18.75 MHz. In most cases, the

pixel rate used will be less than 15 MHz, as multisensor pixel registration becomes more of a challenge with high resolution cameras.[5]

7. DIFFERENTIAL ABSORPTION LIDAR (DIAL) APPLICATIONS

Battelle is using a state-of-the-art mobile sensor suite incorporating active, multiwavelength, laser remote sensing technologies for applications such as field remote sensing of chemical/biological warfare (CBW) agents and environmental monitoring. The test bed system consists of the U.S. Army Dugway Proving Ground lidar system van currently on loan to Battelle. The mobile system provides a lidar capability for evaluating and testing improved hardware designs and components while serving as a field test support system for a variety of industrial, regulatory, environmental, DOE and DoD applications. Specific areas of use for the system include atmospheric boundary layer profiling, aerosol detection, hazardous air pollution monitoring and CBW species identification and detection in support of treaty compliance and nonproliferation monitoring.

The ground-based, direct detection, tunable CO_2 system is designed for rapid detection and concentration measurement of various molecular constituents within the atmosphere using range-resolved and column-content DIAL techniques. Modifications and refurbishment of the system are currently being conducted at Battelle for the lidar (laser sources) and receiver subsystem.

The lidar system is housed in a 30-foot 1988 Frontrunner manufactured by Wolverine Western Corporation. The van houses a climate control system to maintain a comfortable operating environment for both the operator and equipment. A 30 kW Kohler generator mounted in the rear of the van provides electric power. Leveling and stabilization are performed by hydraulic jacks located at each corner of the vehicle.

The lidar package received from Dugway Proving Ground contains four Questek Series 7000 TEA CO_2 lasers with a nominal 4 Joule transmitted energy at 10P(20). The infrared receiver uses a 16-inch f/2.5 telescope along with a liquid-nitrogen-cooled HgCdTe detector. The scanner, mounted on top of the vehicle, allows pointing over a full hemisphere. Data acquisition and processing is accomplished with a Digital Equipment Corporation Microvax 3200 and CAMAC data acquisition module. A color graphics display, magnetic tape storage and a laser printer are also on board the vehicle.

8. CONCLUSION

The spectral distribution of energy reflected and/or emitted from an object is a unique characteristic that may be exploited to discriminate the object of interest from the local background. Near-simultaneous, multispectral imagery of ground target and background signatures were collected over a full diurnal cycle in visible, infrared, and ultraviolet spectrally filtered wavebands using Battelle's portable sensor suite. The imagery data were coregistered and processed using a variety of classical statistical algorithms, artificial neural networks and data clustering techniques to classify pixels and objects in the imaged scenes. Imagery collected at different times throughout the day were employed to verify algorithm robustness with respect to temporal variations of spectral signatures. The investigation supports a wide variety of multispectral signature discrimination applications, including automated target search, landing zone detection, and chemical/biological agent tracking.

Results of the study were excellent and indicated that the chosen data fusion algorithm was not critical to solving the object classification problem. Rather, the system solution lies in choosing the best set of features (in this case wavebands) for discriminating the object of interest (i.e., target, runway, etc.) from its background surroundings. With a set of wavebands judiciously chosen for discriminating the object of interest under established conditions, any of the mentioned algorithms can be tailored to use the multispectral measurements to effectively classify the imagery.

9. REFERENCES

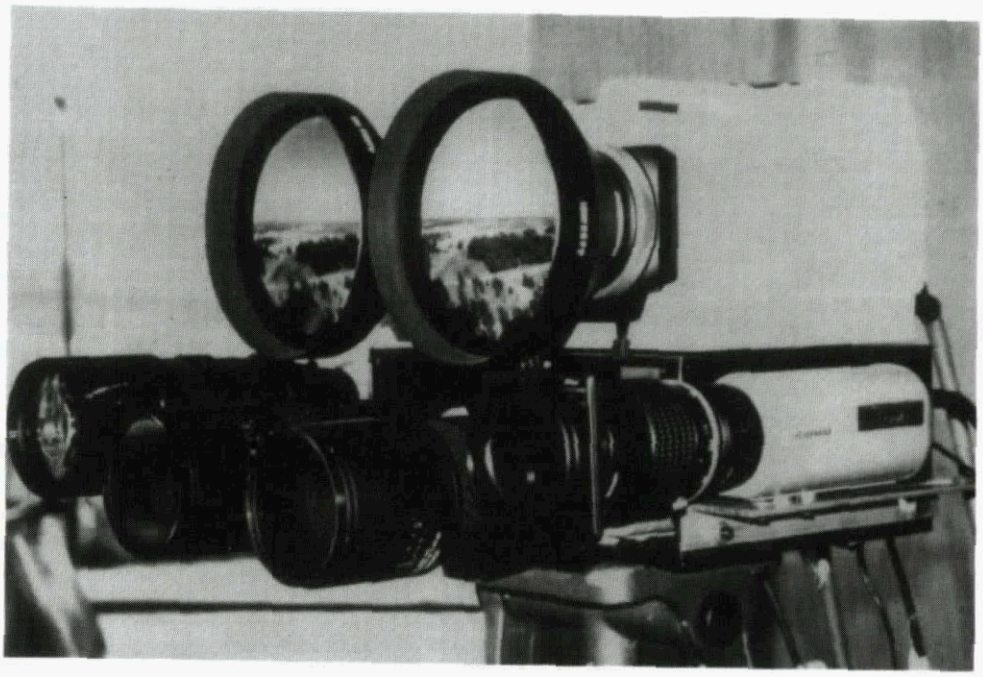
[1] Lazofson, Laurence E. and Kuzma, Thomas J., "Scene Classification and Segmentation Using Multispectral Sensor Fusion Implemented with Neural Networks," *SPIE International Symposium on Optical Engineering and Photonics in Aerospace and Remote Sensing*, Orlando, FL, April 1993.

[2] Geographic Resources Analysis Support System (GRASS) 4.0 (Software), U.S. Army Construction Engineering Research Laboratory, 1991.

[3] Kuzma, Thomas J. and Lazofson, Laurence E., "Automatic Target Detection Using Multispectral Sensor Fusion Implemented with Neural Networks," *1993 Automated Mission Planning Society Symposium*, San Antonio, TX, 24-27 May 1993.

[4] NeuralWorks Professional II/PLUS Software, NeuralWare, Inc., Pittsburgh, PA, 1992.

[5] Preston, Evan et al., "Development of a Field-Portable Imaging System for Scene Classification Using Multispectral Data Fusion Algorithms," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 9, No. 9, September 1994.



Battelle Imaging Sensor Suite Bands:

UV 0.2 - 0.4 μm

blue 0.42 - 0.49 μm
green 0.51 - 0.56 μm
red 0.61 - 0.66 μm

NIR 0.81 - 0.95 μm (vegetation)

Subband MWIR 3.5 - 4.1 μm
Subband MWIR 4.5 - 5.0 μm
Broadband MWIR 3.0 - 5.3 μm

Subband LWIR 7.81 - 9.8 μm (CW)
Subband LWIR 9.98 - 11.41 μm (CW)
Subband LWIR 9.14 - 9.68 μm (CW)
Subband LWIR 10.497 - 10.857 μm
(CO_2 laser/CW)
Broadband LWIR 8.0 - 13.0 μm

Fig. 1. Battelle multispectral sensor suite.

Fig. 2. Multispectral signature test - tower view.

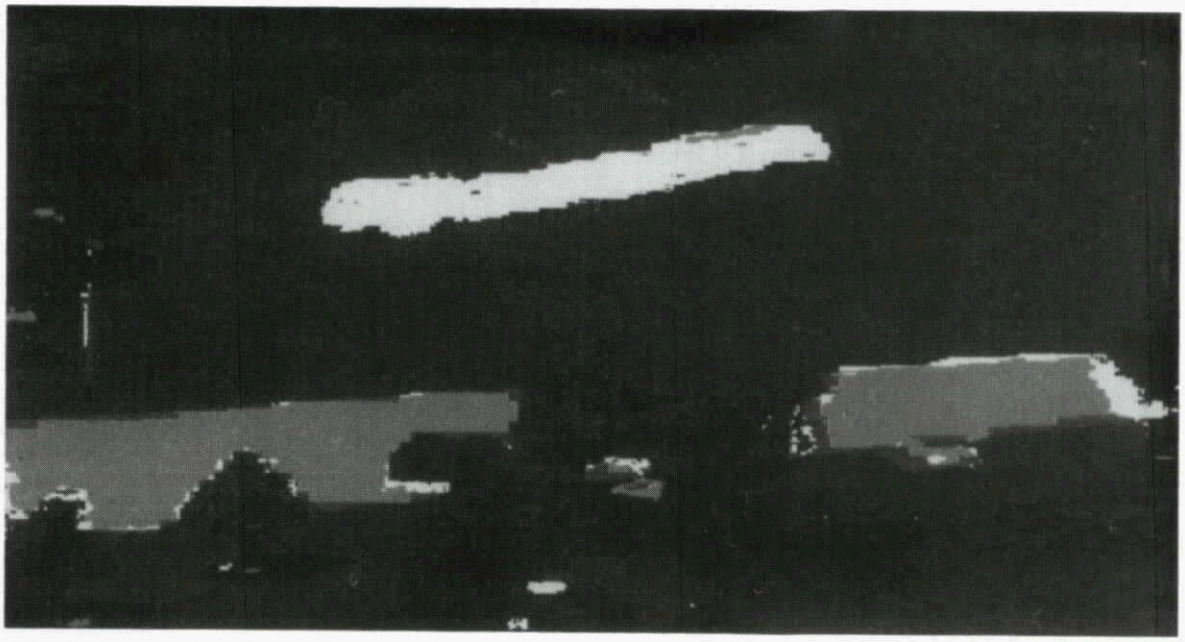


Fig. 3. Image classified using neural network.

Fig. 4. FAA landing program data collection - imaged scene.

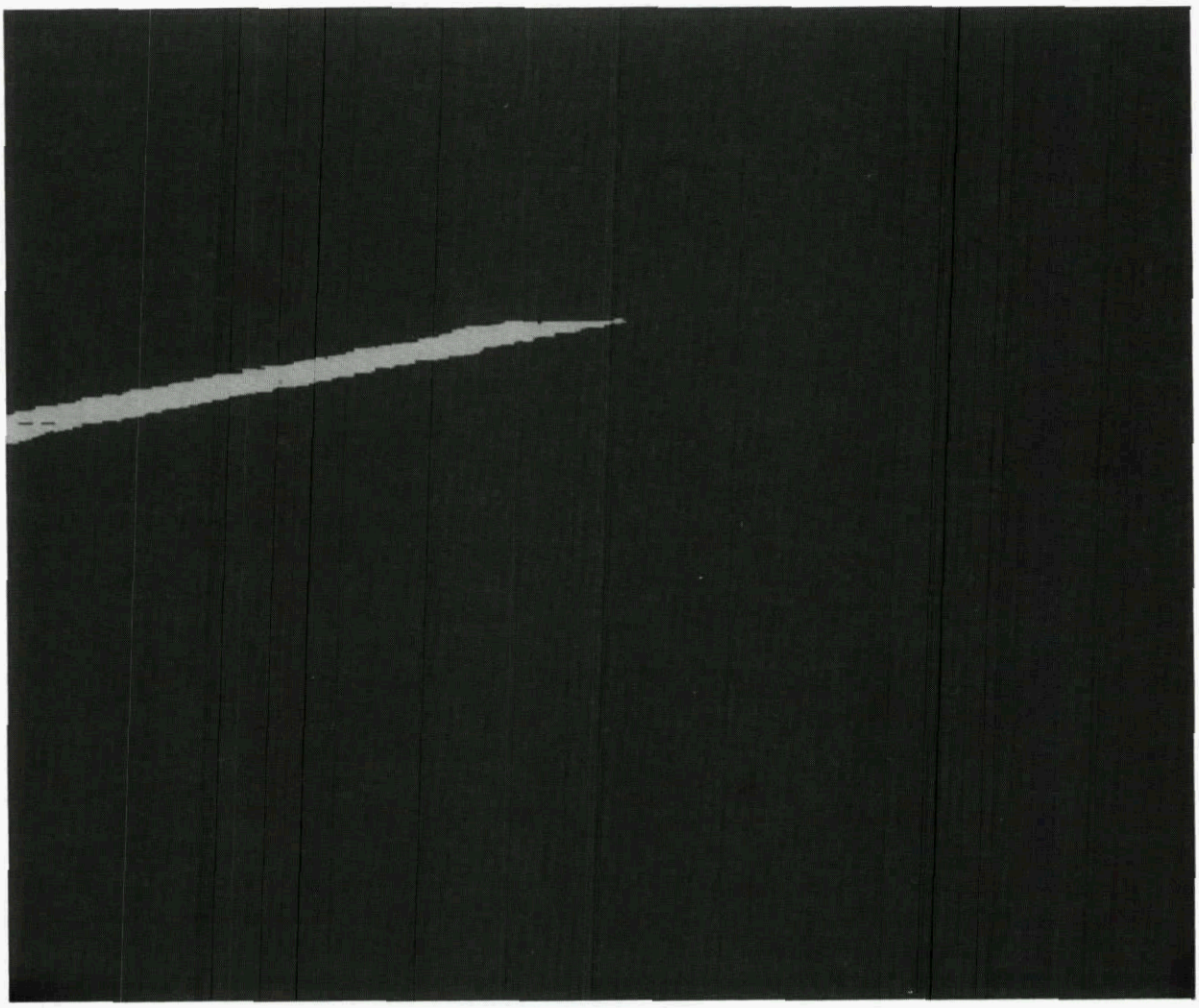


Fig. 5. Segmented runway using unsupervised classification - 1800 hours, 3km.

SECTION IV

SIMULATION AND PERFORMANCE EVALUATIONS

INTRODUCTION

by
Dr. Bruno Mazzetti
ALENIA
Preliminary Design Manager
Defence Aircraft Division
Corso Marche 41
I-10146 Torino
Italy

This Section contains three papers that support the simulation and performance evaluation of multi-spectral data fusion systems. An effective simulation and performance evaluation is essential to rapid prototyping and algorithm selection to support numerous MS/MTT applications. In this Section, we cover the handling of Automatic Target Recognition (ATR) test data, an effective tool to support the development of precision guided munitions and a study of target acquisition and sensor cueing in air-to-air environment.

The first paper presents a method for bench-marking ATR data. The bench-mark for performance measurement of this statistical classification rule was compared with other classification algorithms. By means of statistical considerations it is possible to derive the optimal classification rule that, on the average, its application yields the lowest probability of committing classification errors. This technique becomes central in real-time Automatic Target Recognition Systems that employ multi-sensor and multiple-look evidence accumulation strategies.

The second paper describes the Modular Algorithm Concept Evaluation Tool (MACET) to quickly configure and evaluate multi-spectral sensor fusion algorithms. Target acquisition in a high clutter, all weather day/night condition represents a rather demanding capability for air-to-surface attack mission. The MACET is being developed by Wright Laboratory to evaluate acquisition and aim point selection algorithms. This will ensure that the critical elements will allow the weapon system to operate independently and accurately in the operational scenarios.

The last paper used a simplified flight simulator in a dynamic environment to evaluate sensor fusion algorithms and cueing processes. The operational capability of fighter aircraft in a very demanding air-to-air threat scenario represents one of the future challenges for aircraft

designers. The use of multi-sensor data fusion, sensor management, situation assessment and attack management techniques is the envisaged solution to the problem. Accurate tracking, short response time and a sufficient level of confidence in identification were recognized as critical requirements. Therefore, a simulation study was aimed at tailoring the algorithms to tracking and fusion problems, so that there would be more confidence in the phenomena and in the performance of the adopted techniques in the operational situation.

Benchmarking ATR Test Data*

Andrew Hauter, Vince Diehl, and Arnold Williams
SAIC, Arlington, VA 22203
vdiehl@rsta.saic.com

Geoffrey Orsak and Matthew Sorell
College of Engineering, George Mason University
Fairfax, VA 22030 gorsak@tejas.gmu.edu

SUMMARY

By means of statistical considerations it is possible to derive a classification rule which is optimal in the sense that, on an average basis, its use yields the lowest probability of committing classification errors. This statistically optimal classification rule is a generally accepted standard against which the performance of other classification algorithms are often compared. As is true in many fields which deal with measuring and interpreting physical events, statistical considerations become central in Automatic Target Recognition because of the randomness under which classes are distributed in sensory data. There has been considerable investment in real time Automatic Target Recognition Systems that employ multiple-sensor and multiple-look evidence accumulation strategies. These ATR's exploit predetermined models of the targets and target states using correlation matching processes to declare targets or target features and various detector forms and graph matching techniques to accumulate evidence. Due to the nature of the sensors and environment, it is not clear that these algorithms can be cast in an optimal fashion; however, given a selected training and test set much can be said about the respective performance. It is this benchmark by which this paper measures performance. To this end, techniques for estimating the probability of errors inherent in the sampled data sets and the calculations of bounds on this measured performance are investigated.

1.0 INTRODUCTION

The common approach for evaluating ATR's has been to provide a common data set, formulate an ATR competition, and select the best performer by process of elimination. This testing process can be beneficial if (1) the test data selected does not bias one algorithm over another, (2) the data represents a robust challenge to the algorithms and (3) there exists enough data that there is sufficient confidence for design and testing. However, the shortcoming of this approach is that it doesn't provide a stopping point. The resolution of this issue or even the methodology to arrive at an answer is not altogether clear in the literature, yet it is critical. This paper takes a two-fold approach, first from an optimal classification point of view in which the optimal classifier is derived and its performance estimated by simulation, and second from a minimum achievable error probability approach in which the minimum probability of error is bounded by analytical expressions which incorporate sufficient statistics derived from the data. The topics covered here are grouped under the following headings:

2. SUMMARY PROBLEM DESCRIPTION
   - 2.1 Description of Error Measures
   - 2.2 Decision Criteria
   - 2.3 The Receiver Operating Characteristic
   - 2.4 Summary
3. BOUNDS ON THE MINIMUM PROBABILITY OF ERROR
   - 3.1 Problem Statement
   - 3.2 Distance Measures and f-Divergence
   - 3.3 A class of upper bounds on min P_e based on f-divergence's
   - 3.4 A class of lower bounds on min P_e based on f-divergence's
4. SAMPLE SIZE AND ERROR ESTIMATION TECHNIQUES
   - 4.1 Monte Carlo Methods
   - 4.2 Error Counting Methods
   - 4.3 Importance Sampling
   - 4.4 Application to ATR
   - 4.5 Challenges for the ATR Problem
5. APPLICATIONS
   - 5.1 Likelihood Ratio Test
   - 5.2 Parzen Kernel LRT
   - 5.3 Asymptotically Optimal Detector
   - 5.4 Asymptotic Equivalence and Conclusions
6. REFERENCES

* Work Sponsored By ARPA Contract #F33615-91-C-1801

2.0 SUMMARY PROBLEM DESCRIPTION

Before proceeding to details and specific examples, we give a descriptive account of how two probability distributions determine the detectability of a signal, or the discernibility between two signals. Optimally one wants to determine the best possible performance given some set of sensor data. This performance can be predicted by a bound on the probability of error. The error represents those cases where we have missed a target (probability of miss, P_M) together with those where we have designated clutter as a target (probability of false alarm, P_{FA}) The existence of tight bounds on these quantities will offer a meaningful measure of relative performance of different detection schemes. If we can determine this fundamental performance for the data we will be able to judge how well algorithms can operate on the input sensor data. In turn, this will also allow a determination of the adequacy of a sensor design to fulfill a specified requirement.

In evaluating ATR performance, two obstacles prevent an easy solution: firstly, the underlying distributions are incompletely characterized, and secondly, the number of training samples (or feature sets) required for performance estimation is usually smaller than desired (sample size determines the accuracy of the performance estimates). To this end, pattern recognition research has considered various questions concerning the relationship between the training set size, the number of features, and the estimation of performance [1].

2.1 Description of Error Measures

The Neyman-Pearson lemma indicates that the optimal test between two hypotheses is the likelihood ratio test. To illustrate, let us consider the two class detection problem based upon an N-dimensional observation vector with the following corresponding hypotheses:

$$\begin{aligned} H_2&: \ X \sim P_2(X) \text{ target} \\ H_1&: \ X \sim P_1(X) \text{ clutter} \end{aligned} \qquad (1)$$

In the case that the joint distributions are known precisely, the likelihood ratio test becomes: (2)

$$l(X) = \frac{P_2(X)}{P_1(X)} \underset{H_1}{\overset{H_2}{\gtrless}} T$$

where p_1 and p_2 are the corresponding joint density functions under H_1 and H_2 respectively. It is well known that the likelihood ratio, $l(x)$, is the sufficient statistic under the most widely used optimality criteria in detection and estimation theory. These include minimum probability of error, optimal Neyman-Pearson, and mini-max (P_{FA}, P_M). Because of the generality of the LRT, it is to be expected that "good" ATR solutions will attempt to approximate this statistic.

For the binary hypothesis problem, the interaction between the likelihood ratio test, the probability of errors, and the threshold setting T are depicted in Figure 1.

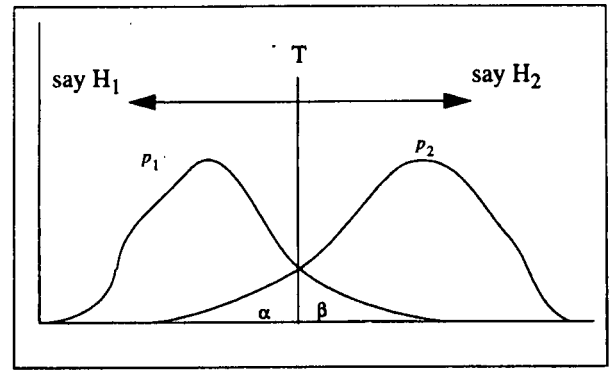


FIGURE 1. Testing Between Two Hypothesis

The probability of errors are defined on the conditional probabilities:

$$P_{FA} \equiv \beta = \Pr\langle say\, H_2 \mid H_1\, true\rangle = \int_T^{\infty} P_1(x)\,dx$$

$$P_M \equiv \alpha = \Pr\langle say\, H_2 \mid H_1\, true\rangle = \int_{-\infty}^{T} P_2(x)\,dx \qquad (3)$$

2.2 Decision Criteria

As alluded to above, the three decision criteria of interest in this document are the Bayes min P_e, Neyman-Pearson, and so-called mini-max. We examine each of these in detail beginning with the Bayes min P_e.

2.2.1 Bayes min Probability of Error

The Bayes Probability of error is described by the following relationship:

$$P_e = P[say\, H_1 \mid H_2]P[H_2] + P[say\, H_2 \mid H_1]P[H_1] \qquad (4)$$

As discussed in the previous section, the optimal detection rule under this criterion is the likelihood ratio tested against the threshold $P(H_2)/P(H_1)$. If the threshold T is not set to this quantity, then an increase in the P_e will be incurred. Unfortunately, a

difficulty of using a Bayes detector for the ATR problem is lack of knowledge of the prior probabilities. Nevertheless, many of the approaches in detection theory and pattern recognition have derived their error estimation procedures based on the Bayes detector; as such we will examine this classification scheme closely. Moreover, the basic Bayes error concept given above can be generalized to include loss functions where the Bayes error then becomes the Bayes risk [2].

2.2.2 Neyman-Pearson Criterion

In the absence of the knowledge of the prior probabilities, researchers have typically chosen to use the Neyman-Pearson Criterion:

$$\min P[\text{say } H_1|H_2] \text{ such that } P[\text{say } H_2|H_1\} < \alpha \qquad (5)$$

In this case we seek decision rules with minimum Probability of miss while imposing a level constraint on the false alarm rate. As in the minimum P_e optimality criteria, the LRT against a suitably chosen threshold is the optimal classification rule. It is a popular practice in the ATR community to standardize a specification for a P_{FA} (that is a); and correspondingly to find solutions which minimize P_M [$P_D = 1 - P_M$]. It is interesting to note that the minimum Bayes error decision rule and the optimal Neyman-Pearson solution are both the Likelihood Ratio Test with a different threshold T (see Section 2.3, the Receiver Operating Characteristic).

2.2.3 Minimal Test

In addition to the min P_e and N-P criteria, detection theorists often consider the so-called minimal criterion:

$$\min \max \{P[\text{say } H_2|H_1], P[\text{say } H_1|H_2]\} \qquad (6)$$

Once again, under this criterion, the optimal classification rule is the likelihood ratio tested against a threshold T. However, in this case, the threshold is determined by the worst case set of a priori probabilities. Thus one can be assured that the performance resulting from the environment will always be superior to the performance of the mini-max test.

2.3 The Receiver Operating Characteristic

As previously mentioned, the Neyman-Pearson test completely specifies the minimum values of P_{FA} and P_D (probability of detection) as the test threshold varies. It is worth noting that for the likelihood ratio test each class can be made up of an N-dimensional joint density and that the ratio function is a multidimensional reduction onto a one-dimensional decision axis. Since the Bayes error (4) follows easily if P_D and P_{FA} are known, then it is logical to focus on calculating P_D and P_{FA} as a function of the chosen threshold. A convenient method for displaying this information is in the form of a Receiver Operating Characteristic (ROC) curve, in which a plot of P_D vs. P_{FA} is constructed for all thresholds of the likelihood ratio test.

We begin by considering equations (1) and (2) in Section 2.1, the likelihood ratio test is

$$l(X) = \frac{p(X \mid H_2)}{p(X \mid H_1)} \underset{H_1}{\overset{H_2}{\gtrless}} T \qquad (7)$$

The variable $l(X)$ is a random variable whose probability density depends on which hypothesis is true. If the two conditional densities, $p(l|H_2)$ and $p(l|H_1)$ are known then P_{FA} and P_D are given by:

$$P_D = \int_T^{\infty} p(l \mid H_2)\, dl \qquad (8)$$

$$P_F = \int_T^{\infty} p(l \mid H_1)\, dl \qquad (9)$$

In Fig. (2) we show representative conditional density functions $p(l|H_2)$ and $p(l|H_1)$ which have been transformed by a 5x5 likelihood ratio filter (the details of which will be explained in Section 5.1). Clearly, one of the main obstacles in evaluating the optimal achievable ATR performance is lack of detailed knowledge of these functions. However, if sufficient amounts of data are available, empirical estimates may be formed. Based on these quantities, estimated ROC curves can be constructed. In Figure 2, we see both the empirical data densities and the corresponding densities associated with the LRT. From these we have formed the empirical ROC. The difficulty is that it is often hard to find $p(l|H_i)$. Clearly, determining accurate estimates of $p(l|H_i)$ is a shortcoming of this approach. Nevertheless, given sufficient data there are many statistical methodologies for developing consistent density estimators.

The receiver operating characteristic for the likelihood ratio summarizes a specific relation between two probability distributions. The coordinate system shows hit rates (P_D) as ordinates and false-alarm rates as abscissas. When the probabilities are plotted linearly, the values of the coordinates run from zero to one, and therefore all possible ROC curves are bounded by a unit

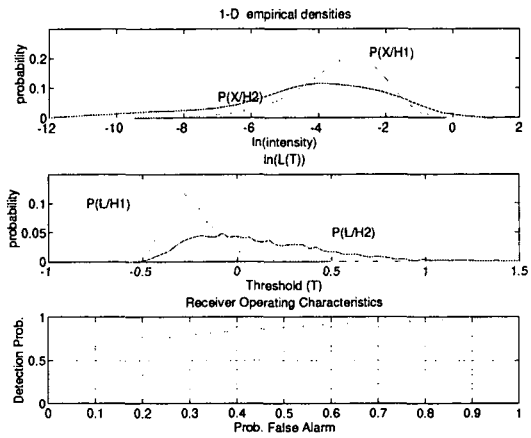


FIGURE 2. Typical densities

square. As shown in Fig. (3), the chance line (a) is the positive diagonal, which occurs where $P_D=P_{FA}$ This line is the ROC curve for a detection scheme which ignores the data given and simply guesses at random. Any ROC curve above this diagonal line, such as (c), is termed "proper" as the performance is better than guessing. A proper ROC curve is a monotonically increasing convex down function from (0,0) to (1,1). Any ROC curve below the diagonal line such as (d) is "improper" as better performance is achieved by guessing; however, such a poor classifier can be made proper by inverting its decision so that the classifier says "target" when it now says "no target", and vice versa. The intersection of the negative diagonal (b) where $P_D=1-P_{FA}$ with the ROC curve is the minimax operating point, and the intersection of the line P_{FA}=a with the ROC curve is the Neyman-Pearson operating point. For a more detailed treatment of the properties of ROC curves based upon likelihood ratios, see [3, 4].

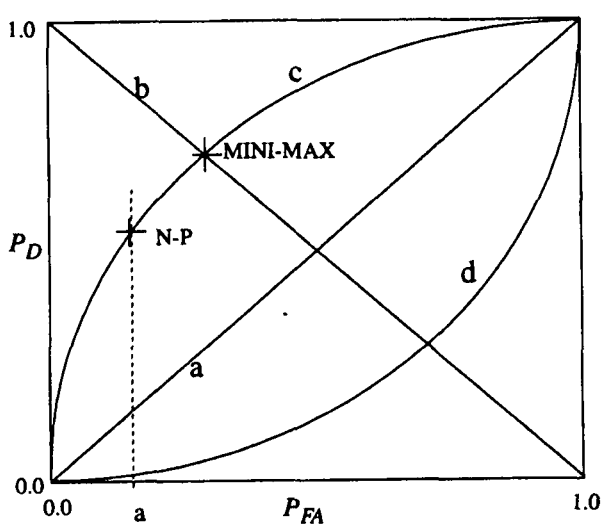


FIGURE 3. The coordinate system for the ROC curve

2.4 Summary

Summarizing several key concepts for the binary decision problem:

1. From the Neyman-Pearson lemma, we find that the optimum test is a likelihood ratio test. Thus, regardless of the dimensionality of the observation space, the test consists of comparing a scalar variable with a threshold.

2. A complete description of the $l(X)$ performance can be obtained by plotting the conditional probabilities P_D and P_{FA} as the threshold T is varied. However, if only the minimum average error is desired it is generally easier to calculate the Bayes for a specific threshold setting.

3. For many cases, construction of the $l(X)$ can be simplified by using a sufficient statistic. This solution, however, has been evasive for the ATR designer because of the non-parametric and non-stationary nature of imagery.

The classical Bayes, minimax, and Neyman-Pearson strategies have been known for some time (hypothesis testing was formally introduced in 1933)[1]. The challenge lies in requiring a complete (parametric) statistical description of the data to specify the optimum decision rule. It is frequently demonstrated that algorithms designed around a particular model may perform poorly when actual data differs from the assumed model. This suggests two methods for solving this dilemma: either develop accurate models of all class distributions or develop a non-parametric technique that is proven robust for parametric cases. The first option is more desirable because the question of "true" performance then becomes an analytic exercise (not necessarily an easy one). Unfortunately, this option is not pragmatic due to the difficulty of deriving and validating the large number of classes that would be required. Alternatively, the non-parametric approach is attractive because the test becomes universal in the sense that the test can be constructed in partial ignorance of the probability distributions governing the system ["partial" is used because blindly applying a non-parametric approach without some knowledge of the data can cause unnecessary errors]. The disadvantage is that proof of robustness over a wide range of cases is not a general proof.

It should be clear from the preceding discussion, that evaluation of the error probability is in general a very difficult task. The analytical bound approach seeks an approximate expression for the error probability in the form of upper and/or lower bounds. These bounds are often easier to compute than the P_e itself. The second approach is facilitated through the use of detectors which attempt to approximate the LRT through the use of non-

parametric approaches. One non-parametric approach employs a Parzen kernel estimate and the second employs a finite state Markov models. Minimum error rate estimates are then obtained by estimating the performance of these detectors offering a tight upper bound on the min P_e.

This paper outlines each procedure and the process of testing them on a common data set, which in this case corresponds to high frequency data from the Lincoln Laboratory ADTS SAR System. Although targets are available testing was restricted to two classes of clutter—grass and trees, in order to stay at the unclassified level.

3.0 BOUNDS ON THE MINIMUM PROBABILITY OF ERROR

To evaluate the performance of a detector, it is necessary to establish the minimum achievable probability of error for comparison. Direct calculation of this figure requires both knowledge of the optimal detector and integration over the full decision space. Since we cannot define the optimal detector in the ATR problem (the distributions of the hypotheses are never precisely definable) and the decision space is large, direct computation is infeasible. Although it is possible to construct an asymptotically optimal detector, and estimate the probability of error of this detector, for finite training sets there is no guarantee that this detector is optimal, hence there is no guarantee that we have estimated the minimum achievable probability of error. It is therefore desired to characterize the minimum achievable probability of error by means of tight upper and lower bounds. These bounds can then be compared with the performance of the classifiers under test to assess their relative merit.

In this section, a family of bounds are presented for the binary decision or detection case. These and other bounds for the multiple hypothesis classification problem may be found in the literature, see for example.[23]

The bounds presented give various degrees of tightness of fit to the true minimum probability of error for a given data set. The choice of the specific bound depends upon the amount of data and the type of sufficient statistics available, and on assumptions made in the derivation of the bound, such as dependence, symmetry or distribution.

It is well known that all good bounds require the knowledge of the true probability distributions underlying the data. In the ATR problem domain, this information is rarely available, thus there is a gap between theory and practice. Our aim here is to present classical theoretical bounds and show how they may be used with simple models. They are equally applicable to real sensor data, but more sophisticated models are required. Signal modeling is not addressed by this report, though it is critical for the ATR problem.

The motivation for finding the minimum achievable probability of error is to evaluate how well the classifier under test performs against the best possible classifier for a given data set, which by definition must achieve the minimum probability of error. Rather than determining the exact form and performance of the optimal classifier, we merely estimate its performance based on the data set. For example, if the minimum probability of error against a given data set were bounded by say 0.009 and 0.012, we would consider a classifier with an error probability of 0.011 to be performing very well and possibly optimally, whereas another classifier with an error probability of 0.020 would be suboptimal. On the other hand, we would be dubious of our measurement techniques and/or the detector algorithm if an error probability of 0.001 were measured.

In this section we assume our data to be independent and identically distributed. Such an assumption is rarely justified in ATR as there are certainly dependencies between neighboring pixels. Nevertheless theory may be extended to cover the case of dependent data, although this will not be discussed here. It should also be noted that the data vector size depends both on the window size in the image and the number of looks within the window - increasing the window size is of no use if the target is already completely encapsulated, rather the window should be of the optimal size to enclose the target, the classification error could then be reduced by increasing the number of looks at the target - from different angles, polarizations, radar frequencies *etc.*

3.1 Problem Statement

We seek to find the minimum Bayes error-probability defined by [4]:

$$\min P_e = 1 - \int_{R^n} \max_i \, [\pi_i p(x \backslash H_i)] dx \tag{10}$$

where the integral is taken over the entire observation space R_n. As stated, computing this integral is elusive even if the conditional densities are known precisely. To circumvent this, information theoretic and other quantities will be utilized to bound the minimum probability of error both from above and below.

3.2 Distance Measures and *f*-Divergence

The purpose of a distance measure between two hypotheses is to offer some meaningful measure of

the dissimilarity between two density functions. Based upon this, one may attempt to quantify how distinguishable the two hypotheses are from one another and hence lead to an estimate of the minimum achievable probability of error. Our motivation is then to seek some distance measure with a close relationship to the true minimum probability of error, so that the distance measure can be used to bound the minimum probability of error from above or below.

Many distance measures have been derived for this purpose by various authors - see for example.[4] To illustrate, we will describe a family of distance measures parameterized in terms of some function. Some well known distance measures which happen to belong to this family off-divergences will also be given. Lower and upper bounds on the minimum probability of error derived from this family will then be presented, again including some classical bounds.

We proceed by defining $f(x)$ to be a convex-up function, following the conventions in 17, 91, which is well behaved as x approaches zero and infinity. We define the f-divergence between two joint density functions P_1 and P_2 by:[5]

$$D_f(P_1,P_2)=\int_{R^n} f\left(\frac{P_2(x)}{P_1(x)}\right)P_1(x).dx \tag{11}$$

Use of the f-divergence allows the construction of a general family of bounds on the probability of error which includes many classical bounds such as the Chernoff bound. It should be noted however that this family, though broad, does not include all common bounds.

3.3 A class of upper bounds on min P_e based on f-divergences

It has been shown by Boekee and van der Lubbe in [3] that the probability of error for the binary hypothesis problem with probabilities P_1 and P_2 is upper bounded by

$$P_e \le \frac{f_o\pi_2 + f_\infty \pi_1 - \overline{D_f}(H_1,H_2)}{f_2 - f_1} \tag{12}$$

where

$$f_o = \lim_{\mu\downarrow o} f(\mu), \quad f_1 = f(1),$$

$$f_2 = f_o + f_\infty, \quad f_\infty = \lim_{u\to\infty}\frac{f(u)}{u} \tag{13}$$

provided f_2 is finite - this places further constraints on our choice of f. In this way we have an upper bound on the probability of error. We illustrate this upper bound with the classical Bhattacharyya upper bound.[6]

From the average Bhattacharyya coefficient,

$$\bar{\rho} = \int_{R^n} \sqrt{P(H_1 \mid x)P(H_2 \mid x)}\, p(x)dx \tag{14}$$

we have

$$f(u) = -\sqrt{u}$$
$$f_1 = -1$$
$$f_0 = f_2 = f_\infty = 0 \tag{15}$$

From which it follows that

$$\min P_e \le -\overline{D_p}(H_1,H_2) = \bar{\rho} \tag{16}$$

which is a well known result. This bound is illustrated in Figure 3.1 for the following example, in which we assume a constant signal of amplitude s is present or absent in Laplacian noise and is plotted as a function of s with the true minimum probability of error for comparison.

$$H_0: \quad P_0(x) = \frac{1}{2}\exp(-|x|)$$

$$H_1: \quad P_1(x) = \frac{1}{2}\exp(-|x-s|) \tag{17}$$

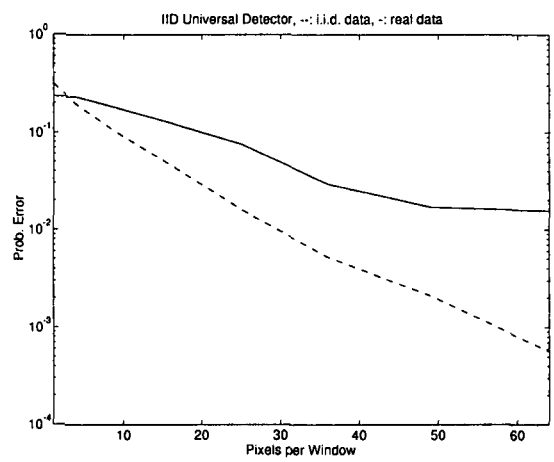


FIGURE 4. Minimum Pe and the corresponding Bhattacharyya upper bound

Note that for the special case of the multi-dimensional Gaussian distribution, the Bhattacharyya distance is given in terms of mean vectors m_i and covariance matrices S_i by:

$$\bar{\rho}(P_1,P_2)=\sqrt{\frac{\pi_1\pi_2\,|\Sigma_1+\Sigma_2|}{2\sqrt{\Sigma_1\Sigma_2|}}}\;2^{\left[\frac{1}{4}(\mu_2-\mu_1)^T(\Sigma_1+\Sigma_2)^{-1}(\mu_2-\mu_1)\right]}$$

(18)

3.3.1 A clas of lower bounds on min Pe based on f-divergences

Boekee and Ruitenbeek[7] have extended a class of lower bounds on the minimum Bayes probability of error by considering the f-divergence between two hypotheses. Following the approach in [Boekee81], we introduce a modified divergence function.

$$f*(u)=u\cdot f\left(\frac{1-u}{u}\right) \tag{19}$$

and set u = $u(x) = P(H_2|x) = 1 - P(H_1|x)$, such that

$$\overline{D_f}(H_1,H_2)=E_x\left\{f*(P(H_2\mid x))\right\} \tag{20}$$

This is a just a special-case f-divergence where we have chosen to use $f*$ as our parameter function. We further simplify our analysis by only considering f functions such that a symmetric modified function is obtained, that is

$$f_*(u)=f_*(1-u),\quad (0\le u\le 1) \tag{21}$$

which is symmetric about $u = \frac{1}{2}$. The following bound can be obtained:

$$P_e\ge u\left(\overline{D_f}(H_1,H_2)\right) \tag{22}$$

where $u = u(\overline{Df}(H_1,H_2))$ is the solution of

$$f_*(u)=\overline{D_f}(H_1,H_2) \tag{23}$$

This equation can be solved analytically for special cases and numerically in general. As an example, consider the Bhattacharyya coefficient as given in (3.15), we have:

$$f(u)=-\sqrt{u} \tag{24}$$

and

$$f_*(u)=\sqrt{u(1-u)} \tag{25}$$

Noting that since $\overline{Df}(H_1,H_2) = -\bar{\rho}$, we obtain the following bound:

$$P_e\ge\frac{1}{2}\left(1-\sqrt{1-4\,\bar{\rho}^2}\right) \tag{26}$$

Other upper and lower bounds have also been derived for specific distance measures, for example for the symmetric Kullback-Leibler bound:

$$\frac{1}{2}\min(\pi_1,\pi_2)\exp(-J)\le P_e\le\sqrt{\pi_1\pi_2}\left[\frac{J}{4}\right]^{-1/4}$$

$$J(P_1,P_2)=\int_{R^n}(p_2)-p_1)\log_2\frac{p_2}{p_1}\,dx \tag{27}$$

We illustrate the Bhattacharyya and Kullback-Leibler bounds with the following example:

Grass and tree SAR image data were used to construct histograms under the i.i.d. assumption. The histograms were constructed from chips of size 192 by 128 pixels and quantized into 25 bins. The bin widths were chosen such that each bin contained approximately the same number of pixels from both distributions together - in this way differences in the distributions tend to be highlighted. Synthetic i.i.d. data was then constructed. The minimum probability of error for this data set was estimated by importance sampling using a likelihood ratio detector trained on the source histograms (infinite training) - since we know that the detector will achieve optimal performance and that importance sampling will result in an unbiased, small variance estimate.

We then computed the Bhattacharyya and symmetric Kullback-Leibler distances for the two histograms and hence constructed the upper and lower bounds for each distance. The number of data elements was varied from 1 pixel to 100, representing up to a 10x10 pixel window. The result is shown in figure 5.

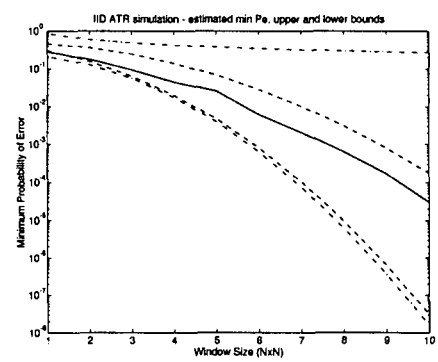


FIGURE 5. Probability of Error Estimate for the Universal Detector, compared with upper and lower Bhattacharyya (dashed) and Kullback-Leibler (dash-dot) bounds on the minimum probability of error

It will be noted that real data is highly structured and therefore contains considerable dependency, hence these results should not be taken as representative of true performance against empirical data. However by introducing known dependencies into the synthetic data, representative of true structure, it is possible to extend this approach to estimate performance against more realistic data.

3.4.1 Conclusion

We have presented a variety of upper and lower bounds for the minimum Bayesisn error probability. These bounds were presented primarily in the context of f-divergent distance measures together with the motivation for this approach. Although these bounds cannot be computed analytically except for very special cases, computer numerical integration makes the calculation trivial, provided that the underlying distributions are known or can be approximated to sufficient accuracy. Such approximation will affect the tightness of the bound, nevertheless for small error probabilities an order-of-magnitude estimate is often informative.

4.0 SAMPLE SIZE AND ERROR ESTIMATION TECHNIQUES

Despite the many bounds available, the net result of extensive investigations on bounds for P_e seems to be that one should try to estimate the error probability itself in some direct way. The relative "tightness" afforded by bounds can only be assessed by examining the true performance. To address this, in this section we consider properties of performance estimation techniques. The topics considered are Monte Carlo, error counting methods, and Importance Sampling.

4.1 Monte Carlo Methods

As before, for the sake of simplicity, let us consider the classical binary hypothesis testing problem where one choose between two alternate hypotheses:

$$H_0: \ \mathbf{X} \sim \mathbf{P}_o$$
$$H_1: \ \mathbf{X} \sim \mathbf{P}_1 \qquad (28)$$

where X is an N-dimensional observation vector and the P_i are arbitrary joint distribution functions governing the data. Assuming a non-randomized decision rule, we may define $Z_0 \subseteq R^N$ to be the region where the classifier chooses H_0 and $Z_1 = Z_0^c$ to be the region where H_1 is chosen. From this, the probability of false alarm for any classification system can be written, as

$$P_F = \Pr[say\, H_1 \mid H_0] = \int_{Z_1} P_0(x)dx \qquad (29)$$

where $p_0(x)$ is the joint probability density function of **X** conditioned on the hypothesis H_0. Observe from the above that the false alarm rate can be written [Hereafter, all expected values are with respect to P_0 unless otherwise indicated] as $E_{P_0}[I_{Z_1}(\mathbf{X})]$ where $I_{Z_1}(x)$ is the indicator function [$I_{Z_1}(x) = 1 \quad \forall x \in Z_1$; 0 otherwise] over Z_1. We form the standard Monte Carlo estimate for P_F as simply a sample average of $I_{Z_1}(\cdot)$, i.e.,

$$\hat{P}_F = \frac{1}{T}\sum_{i=1}^{T} I_{Z_1}(X_i), \qquad (30)$$

where T is the number of trials of the Monte Carlo simulation, end X_i are independent and identically distributed realizations of X generated from distribution P_0. It is easily verified that $\hat{P}_F$ is an unbiased estimator for P_F with associated variance given by

$$\mathrm{var}\left[\hat{P}_F\right] = \frac{P_F(1-P_F)}{\mathrm{T}}$$

Typically one chooses the number of simulation trials T based upon a predefined confidence interval on the estimator $\hat{P}_F$. More specifically, T is chosen so that

$$P\left[\left|\hat{P}_F - P_F\right| > kP_F\right] \le \delta \mathrm{P[IP\ p}$$

By use of Chebyshev's inequality, one can readily show that a sufficient condition for the above confidence interval to hold is for

$$T \ge \frac{1-P_F}{\delta k^2 P_F}$$

If we consider the standard 95% confidence interval defined by $\left[\frac{2}{5}P_F, \frac{8}{5}P_F\right]$, that is with probability 0.95 $\hat{P}_F \in \left[\frac{2}{5}P_F, \frac{8}{5}P_F\right]$ then the number of simulation trials must exceed the quantity

$$T \ge \frac{55}{P_F}$$

To appreciate the scale of this number, consider for the moment an ATR operating at a false alarm or miss rate of 10^{-5}. The above analysis states that one must simulate with at least 5.5 million observation frames (not to be confused with pixels). This number becomes even more prohibitive as the probabilities of interest diminish.

As an alternate means of evaluating this result. one may consider that the number of trials essentially establishes that one must observe either 55 false alarms or 55 missed targets over the simulation to

achieve the desired level of estimator accuracy. For highly effective ATR's, one will hopefully observe far fewer error events than this. As such' standard Monte Carlo methods will suffer in these critical cases.

4.2 Error Counting Methods

The Monte Carlo technique assumes an unlimited supply of random variables, in real system testing the number of samples is frequently too small to merit a reliable estimate. Over the years practitioners in pattern recognition have developed methods for small sample sets to improve the estimate on performance. Moreover, because they make a tradeoff between the optimistic bias of the design set and the weaker bias against an independent test set they also represent a form of upper and lower bound on the true performance. In and of themselves, they do not represent a bound on the optimal performance, which can only be achieved (on average) with the most powerful decision criteria. But if used in conjunction with a classifier designed for optimal performance these error counting methods offer a reliable upper bound range to the optimal performance. We now present four of these techniques.

1. The Resubstitution Method

The resubstitution method is simply training and testing with the same data sets. Ideally, if enough data has been collected and set aside then there will exist a sufficient amount of data to train the classifier. In so doing, the designer frequently tests the performance on this training set and if necessary will proceed to modify or optimize the algorithm to obtain better performance.

However, there are perils involved with drawing conclusions with this resubstitution estimate. For example, let us suppose that the number of samples is not larger than twice the dimensionality of the sample space. Then it can be shown[8] that a hyperplane exists that separates the samples perfectly. Then while testing the detector with the training set, the designer would observe an over-optimistic error rate. Since it is not always possible to precisely determine the dimensionality of the data, one way to avoid this peril is to reserve enough samples for testing independent of the training set, i.e., the hold out method, which would provide a relative bound to the resubstitution estimate.

In general, the estimated P_e using the resubstitution method will tend to be less than the hold out method, depending on the test and training data set used. The difference between these estimates can provide a measure of variability between the training and test data. This measure becomes more reliable as the data sets increase in size and the variance of the error estimates decrease.

2. The Hold Out Method

In the hold-out testing process, the density estimates in training are obtained from data sets independent from the data sets used in testing. Since the detector will be seeing new and different data the hold out estimate provides an upper bound bias. Even though it is conceivable the performance could improve if the test data were easier to discriminate, one could argue that the test set was not chosen properly to match the design set. The importance of testing on data independent of the training data is that this degradation in performance provides a more realistic estimate of how well the classifier will perform, avoiding some of the pitfalls explained in the previous section. However, for severe degradation the important insight is the fact that the classifier was poorly designed, some causes being: poor feature selection, poor model selection, wrong decision criteria, etc. If the loss is extreme, re-thinking the classifier design may be necessary. Hence, the hold out method is important for testing the validity of the resubstitution estimate.

To illustrate this, assume the optimal Bayes classifier is being used.

Let P_d = the design set distribution and P_t = the test set distribution. We know in general given two models P and P' that &e probability of error will be bounded such tied:

$$Pe = Pe(P,P) \leq Pe\,(P,P^l) \quad \forall P^l \neq \tag{31}$$

Therefore the training samples P_d and P_t provide an upper bound on the true performance.

3. Leave-One-Out Method

The leave-one-out method gets around the problems of the hold-out method when sufficient samples for training and testing are not available. The technique trains on all samples but one and tests against the remaining sample. The classifier is then retrained leaving out a different sample and again tested using the remaining sample. The process is repeated for all samples in the data set. This method has been found experimental to be approximately unbiased, however the bias reduction is achieved at the expense of an increase in the variance of the estimator.

4. Rotation Method

The rotation method is a compromise between the hold-out and leave-one-out methods.[9] The data is partitioned into some number of blocks (more than 2 and less than the data set size), and the leave-one-out method is used on a block-by-block (rather than a sample-by-sample) basis. The rotation method reduces both the bias inherent to the hold-out

method and, to some extent, the computational burden associated with the leave-one-out method.

In summary, the average resubstitution error rate provides a lower bound on the true error probability while the others yield upper bounds. The performance is conditioned on the quality of the non-parametric estimate. The better the estimate the tighter the upper bound.

4.3 Importance Sampling

The use of Monte Carlo simulations can produce estimates for probabilities which are arbitrarily accurate, provided the number of trials (observations) is sufficiently large. Unfortunately, in applications where the probability of interest is small, the number of simulation trials required to obtain the desired accuracy is often unacceptable. A modified Monte Carlo technique known as *Importance Sampling* has been shown to provide a significant reduction in the required number of simulation trials [1,3,5,6, and references contained therein]. Given proper implementation, this technique will reduce the variance of the estimate, thereby reducing the number of simulation trials required for a specified level of accuracy. However, before elaborating on the method of Importance Sampling, we begin by reviewing the critical features of the Monte Carlo and error counting methods.

To circumvent the limitation of the Monte Carlo estimate, simulation theorist have utilized the very powerful technique often referred to as *Importance Sampling* to "speed up" the simulations. Importance Sampling is a technique for significantly reducing the number of trials in Monte Carlo simulations. This is achieved by first generating the system input **X*** from a *biased* or modified probability density function, p_0^*, rather than from the original density function. This "biasing density" is chosen such that the random vector (data), **X***, is more likely to come from the regions in the observation space which cause errors. Since in this implementation simulation errors occur with a greater frequency than in standard Monte Carlo simulations, each error in the simulation is scaled by a weighting function determined such that the resulting performance estimate is unbiased. The proper biasing and weighting of events in the simulation will render a reduced-variance estimate of P_F.

Mathematically speaking, the Importance Sampling estimator described above for estimating the false alarm rate is given by

$$P_F^* = \frac{1}{T}\sum_{i=1}^{T} W(\mathbf{X}_i^*) I_{Z_1}(\mathbf{X}_i^*) \tag{32}$$

where $\mathbf{X}_i^*$ are sample distribution vectors generated from the biasing distribution p_0^*, and where $W(\mathbf{X}_i^*)$ is the so-called weighting function on the observation vector. It is easily shown that when the weighting function is chosen as

$$W(x) = \frac{P_0(x)}{P_0^*(x)}$$

and P_0 is absolutely continuous with respect to p_0^*, the Importance Sampling estimator P_F^* is unbiased with variance given by

$$\mathrm{var}\left[P_F^*\right] = \frac{\overline{W} - P_F^2}{T}$$

where $\overline{W} = \left[W(X) I_{Z_1}(X)\right]$.

By computing the Monte Carlo and Importance Sampling variances, we observe that if $\overline{W} < P_F$, then Importance Sampling will result in a reduced-variance estimator for P_F. This critical inequality can be satisfied through the proper choice of biasing strategy, that is through the proper choice of "biased" input data. Fortunately, a number of biasing strategies have been shown to afford significant variance reduction for a variety of general detection and classification problems. These are nominally constructed through some modification of the original density functions underlying the data.

The originating approach was introduced by Shaurnugarn and Balaban[10], where the input data was simply modified by scaling each data value. In this way, the modified density function becomes $p^*(x) = \frac{1}{a} p\left(\frac{x}{a}\right)$. Shortly after this important work, an improved Importance Sampling biasing strategy was concurrently introduced by Orsak and Aazhang[11] and Lu and Yao[12] in which the data was modified by simply adding a constant to each sample. The resulting biasing density is therefore $p^*(x) = p(x-c)$ where c is designed so as to minimize the Importance Sampling variance. This biasing strategy is today the standard by which all other Importance Sampling methodologies are measured.

As an example, let us consider a 5 dimensional signal in additive noise classification problem. We evaluate two cases, one with Gaussian additive noise and the other with Laplacian additive noise. The true (unknown) error rates for these two examples are 4.5×10^{-5} and 2.5×10^{-5} respectively. This in turn requires 1.2×10^{6} and 2.2×10^{6} simulation trials respectively to achieve the confidence interval of $\left[\frac{2}{5}P_F, \frac{8}{5}P_F\right]$ with probability 0.95. However by

incorporating Importance Sampling, it can be shown that no more than 300 simulation trials are required for the same level of accuracy. To demonstrate this, in Figure 6 we plot the Monte Carlo and Importance Sampling probability of error estimates as a function of the number of simulation trials T. One may readily observe the strong "stability" of the Importance Sampling estimators when compared to the standard Monte Carlo estimate.

4.4 Application to ATR

Since the ATR environment is well known to be more complicated than simply signal plus noise, the so-called linear shift biasing strategy described previously must be abandoned in lieu of alternate approaches. Cotrell et al[13] and Sadowsky and Buckew[14] have proposed an exponentially tilted version of the original density as a biasing density. This is formulated mathematically as

$$P_0^*(x) = \frac{\exp(sx)P_0(x)}{m(s)}$$

where $s \in \Re$ and is the real version of the characteristic function. An extended version of this Importance Sampling approach is given by $p*(x) = p_0^s(x)p_1^{1-s}(x)|\mathbf{K}$ where $\mathbf{K}$ is the appropriate normalizing constant. It has been shown that these Importance Sampling implementations result in many desirable asymptotic properties such as exponential gains over Monte Carlo simulations. In terms of implementation, in the binary classification problem, the optimal value of s results in the biasing density being half way between the original competing densities p_0 and p_1. It should be noted that this property is also shared by the optimal linear shift biasing density in the additive signal case.

Based upon these observations, new biasing strategies which are readily implemented can be formed as a linear combination of the original densities, that is

$$p*(x) = \frac{1}{2}(p_0(x) + p_1(x))$$

In the case of testing "grass" from "trees," this implies that a good Importance Sampling data set can be constructed by randomly selecting pixels from either "tree" data or "grass" data with probability 0.5. Clearly the error rate of any detector will be significantly increased when faced with this data set. However the Importance Sampling weighting function described previously will compensate for this by scaling each observed error by a data dependent constant typically less than one. As such, this process should offer a significant reduction in simulation variance.

To evaluate this simulation methodology, we consider evaluating ATR classification algorithms on *synthetic* i.i.d. grass and tree data. Only 100 importance sampling trials were required for each estimate of the probability of error, in order to estimate probabilities as low as 10^{-5}. This compares with over a million samples required for Monte-Carlo estimates of the same order of magnitude.

4.5 Challenges for the ATR Problem

While traditional implementations of Importance Sampling offer significant improvements in Monte Carlo simulations, they unfortunately have required a complete statistical characterization of the data (see the weighting function given above). Of course, this information is not available in a realistic simulation however, it might be the case when

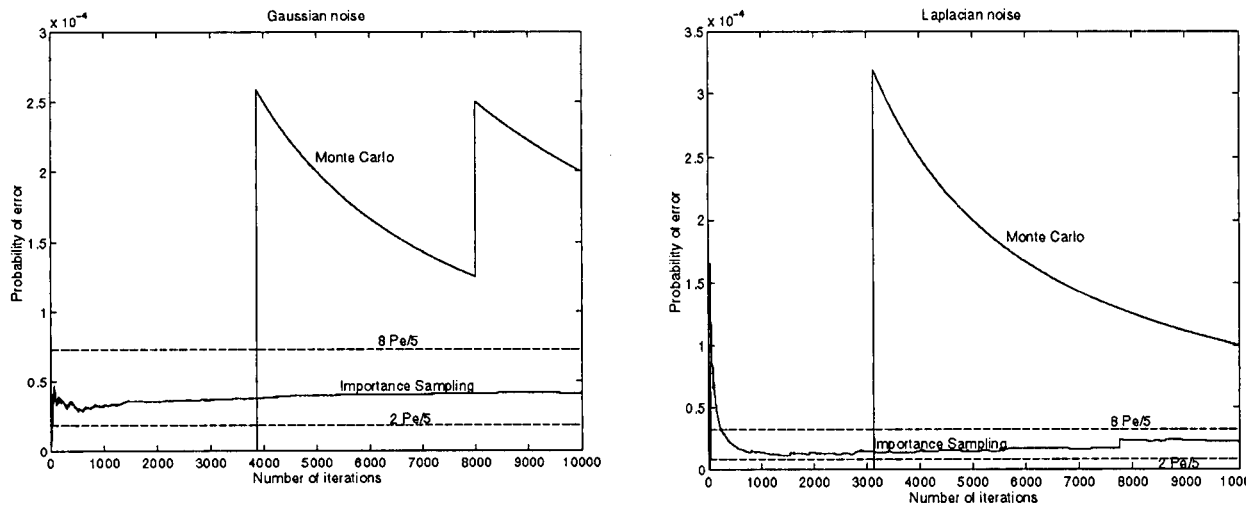


FIGURE 6. Importance Sampling saving compared with Monte Carlo methods

testing on synthetic data. Nevertheless, standard implementations may be approximated by incorporating density estimates based upon training data. With this modified version, one might experience some "bias" in the estimate, however, it is believed that the resulting simulation methodology will result in a significantly smaller average squared error than traditional Monte Carlo simulations. SAIC is currently pursuing this direction of research and will report on its findings once completed.

5.0 APPLICATIONS

5.1 Likelihood Ratio Test

The likelihood ratio test described in Section 2.3 has been shown to be the optimal test under the minimum Bayes error, Neyman-Pearson and minimax criteria. It requires the probability densities of each hypothesis to be known in order to be formulated, a requirement which cannot generally be met in the ATR problem. Nevertheless the test illustrates all of the key aspects of detection theory, even if it cannot be formulated precisely. Stein's lemma provides a relationship between the false alarm probability α, the miss probability β and the sample size n for Neyman-Pearson type detectors, but no such relationship exists for the determination of the operating threshold T. Such a requirement can be avoided by applying a detector to the data and determining the operating threshold from the output conditional densities.

We make use of the fact that the Kullback-Leibler distance is the expected value of the log-likelihood ratio and rewrite the likelihood ratio test in terms of Kullback-Leibler distance, using a powerful technique from large deviation theory known as the method of types, developed by Csiszar and Korner [Csiszar] who used it extensively in their work:

$$\log l(\underline{X}) = \log \frac{p_2(\underline{X})}{p_1(\underline{X})} = nD_{KL}(P\underline{X} \| P_1) - nD_{KL}(P\underline{x} \| P_2) \qquad (33)$$

The likelihood ratio test in Kullback-Leibler form is easily implemented for discrete valued data sets and, in accordance with Stein's lemma, the error probabilities associated with this detector fall to zero asymptotically, provided that the probability densities are known accurately. This scheme is easily extended to the multiple hypotheses case, although all M classes must be modeled. Thus there is one requirement for a low error probability classifier constructed as a likelihood ratio test: accurate probability densities for each hypothesis. This requirement, however, is a challenge that has eluded ATR designers for 20 years. The ATR problem domain as the data dependencies inherent in target identification and are difficult to model accurately and the sample sizes are limited by the number of pixels on target that the sensor resolution allows. Hence adaptive high dimensional models formulated for small data sets are required in this domain.

5.1.1 Example Analysis

Using data files from the SAR data grass and tree regions, each of the two classes contained four files of return power represented as 256 by 256 pixels. Each pixel was approximately one foot by one foot in dimension. To test the K-L LRT on our experimental data, we assigned two of the 8 chips - 1 grass and 1 tree-designated as training chips.

From this 256x256 chip, an estimate of the distribution was formulated using a histogram estimate. This estimate represents P_2 and P_1 in the log likelihood derivation. P_x represents the test data. Hence, a smaller observational window was constructed adaptively over the test images to construct P_x. Observational frame sizes of 32x32 and 20x20 were used. These two window sizes were chosen because most targets would fall within these resolution frames. Results of this filtering process using the hold out method is shown in Figure 7. As expected the better performance represents the 32x32 window (more information used). If the data were real target classes the true performance could be expected to fall between these two ROC curves. The resubstitution estimate (not phoned) performed with complete separability between classes with no errors (on 30,000 samples). Importance sampling would be necessary to provide an unbiased estimate. These low errors are encouraging, considering the challenging classes tested. The relationship between this increasing performance due to increasing spatial

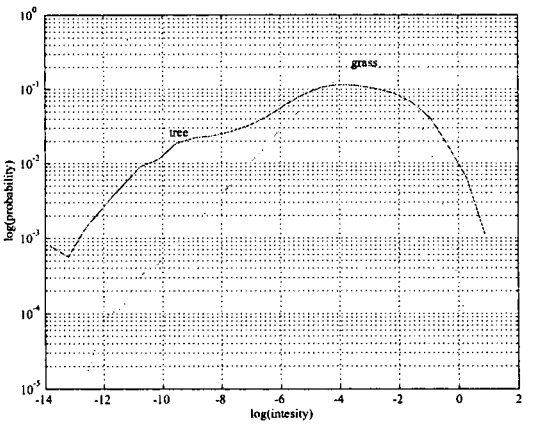


FIGURE 7. Performance for 2 observational frame sizes

window size coincides with the asymptotic equivalence relationship that will be described in section 5.4.

5.2 Parzen Kernel LRT

Pattern recognition problems are particularly difficult to treat mathematically due to the complexity and diversity of the data as well as the small sample sizes of the data relative to the number of features or primitives used in the algorithm design. The underlying distributions can be dissimilar for each class of interest and are often nongaussian. In the previous section the one dimensional distribution estimate performed very well, but we know this is suboptimal because the data is inherently dependent. In most cases the samples (or features) selected for algorithm design are not independent, but correlated.

In this LRT, the Parzen kernel is used to estimate the undersizing distributions:

where,

$$l(X)=\frac{P_2(\mathbf{X})}{P_1(\mathbf{X})}=\frac{\frac{1}{N_2}\sum_{j}^{N_1}K_2(\mathbf{X}-\mathbf{X}_j^{(2)})}{\frac{1}{N_1}\sum_{j}^{N_1}K_1(\mathbf{X}-\mathbf{X}_j^{(1)})}\gtrless T \qquad (34)$$

where,

$$K_i(\mathbf{X}-\mathbf{X}_j^{(i)})=a\cdot\frac{1}{r^n\,|\,\Sigma_i\,|}\cdot e^{(-b(\mathbf{X}-\mathbf{X}_j^{(1)})'(\Sigma_i\cdot r^2)^{-1}(\mathbf{X}-\mathbf{X}_j^{(1)}))} \qquad (35)$$

The parameters of the estimator, a and b, are normalizing constants that depend only on the dimensionality of the samples, n. The N values are the sample sizes of each class i. The estimated covariance matrix for each i class is obtained from the appropriate data set. The inverse and determinant of the sample covariance matrices are computed and used as parameters for the estimator. The threshold and the kernel size r are parameters that receive special treatment. The threshold is estimated from the data by finding the value that gives the minimum error. This value is then modified in certain ways to minimize the bias and variance due to the small sample sizes. The behavior of the Parzen Estimator as a function of the kernel size parameter has been found to have an important role in the small sample size bias and variance effects. At present, no satisfactory method has been found to find the best r for any given data sets. Currently the best solution is to compute the upper and lower bound error estimates for a sequence of r values and plot the results. If the bias removal techniques are working well, the upper and lower curves of the plot should be fairly smooth and both should be relatively flat over a broad range of kernel size values. If the sample sizes are adequate for good error estimation, both error curves should be fairly smooth and the separation between the upper and lower bound should provide a reasonable range of estimation error.

The k symbols in the numerator and denominator of the estimator equation are the kernel functions that are summed to estimate the probability density values for X. Note that the kernels are summed over all of values in the class i, $X_j^{(i)}$ when the lower bound is estimated. When the upper bound error is estimated, the kernel with the same $X_j^{(i)}$ as X is excluded from the sum. There is extensive literature on the Parzen kernel method of density estimation.[15] The Parzen kernel technique can be constructed conveniently to use the resubstitution and the leave one out methods for the lower and upper bounds.

5.2.1 5x5 Dimensional Error Results and Analysis

Four hundred samples of 5x5 observational frames were formed for both the grass and tree data in the following manner. For each of the four files for each class, one hundred 5 by 5 subentries were obtained from the upper left corner of each matrix, i.e., the upper most one hundred 5 by 5 submatrices of each matrix. The 25 column vectors were then obtained by row transforming the submatrices.

The data was then transformed using the power transform. The Quadratic Classifier was run on the untransformed and transformed cats sets, resulting in the error bounds of 0.30 to 0.36 and 0.18 to 0.27, respectively.

The two figures below show the output of the Parzen Nonparametric Error Estimator for the untransformed data and the transformed data

The upper and lower error bounds are 0.23 to 0.15. Because of the larger number dimensional estimate, fewer samples (100 samples) were available for the error estimate. Despite the small number of samples, the upper bound estimate is consistent with the Quadratic Classifier results.

5.3 Asymptotically Optimal Detector

The asymptotically optimal detection schemes formulated by Ziv [Ziv] and Gutman [Gutman] are optimized according to the following criteria:

- Error probabilities must asymptotically go to zero with increasing sample size
- Only training and test data sets are available

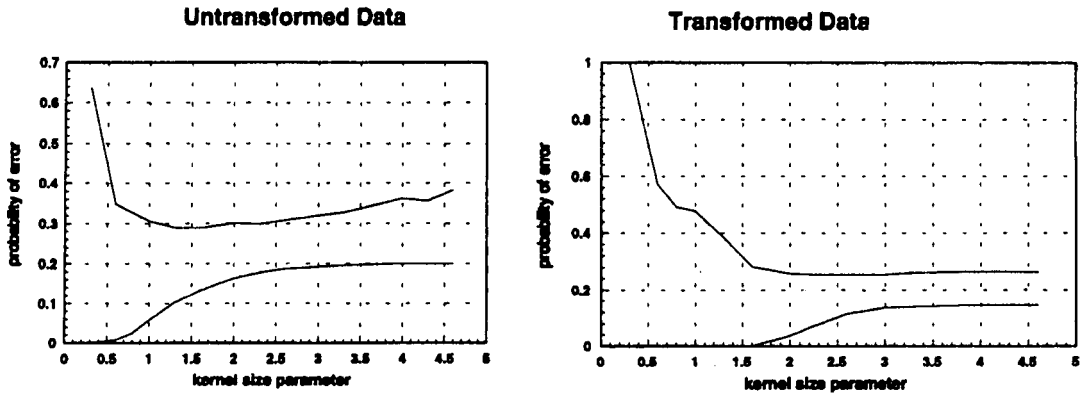


FIGURE 8. 25 Dimensional Error Estimates

Thus the major difference here is that unlike the likelihood ratio test, there is a reduction in the number of probability models required. It turns out from the derivation that for the binary decision case (target or no target) that only one training data set is required for a Neyman-Pearson type detector formulation. The classifier has a similar form to the likelihood ratio test in Kullback-Leibler form:

$$h(\underline{X},t_i,\lambda)=\frac{1}{n}\log\frac{\sup\limits_{P_1P_1\in P_K} P(\underline{X})P_i(t_i)}{\sup\limits_{P_i\in P_K} P_{i(\underline{X}),t_i})}-\lambda \gtrless 0$$

$$=D_{KL}\left(q_{\underline{x}}\,\|q_{\underline{x},t_i})\right)+\frac{n_{t_i}}{n_{\underline{x}}}D_{KL}\left(q_{t_i}\|q_{(\underline{x},t_i)}\right)-\lambda$$

This classifier is optimal in the sense that both error exponents are asymptotically maximized. The scheme is easily extended to the multiple-hypothesis case, as shown in [Gutman]. It may also be noted that as the training sample size tends to infinity, the Ziv-Gutman detector reduces to the maximum likelihood classifier, of which the LRT is a special (binary hypothesis) case.

The asymptotically optimal classifier, like the Kullback-Leibler form, includes all finite alphabet stationary ergodic Markov processes of a finite order. Hence, adaptive models formulated as Markov dependent data lends itself for these classifiers; however, models of this type have not been fully investigated for the ATR problem.

5.3.1 Example Analysis

A second-order Markov dependent Ziv-Gutman asymptotically optimal detector was constructed for the example detection problem of trees vs grass. The Markov state was derived from the pixel above and the pixel to the left of the current test pixel. Examining the sample data by eye indicates that this model will probably be quite effective for grass but trees, which have a clumpy nature, may not be modeled satisfactorily. The detector was trained only on grass using a 128 by 192 pixel window of grass data, and tested against windows of grass and trees, each of size 128x192 pixels. The test window size was varied from one pixel to 32x32 (1024) pixels for grass, and to 64x64 (4096) pixels for trees. The Kullback-Leibler distance between the grass and tree distributions was measured to be about 0.51, so a threshold of λ=0.19 was chosen. A three-level quantizer was used, for histograms with a total of 27 bins.

The results of this detector are given in figure 9. It can be seen that, as established by the derivation of the detector, the false alarm rate is tending exponentially to zero, and that the miss rate is tending towards zero as the sample size increases. The miss performance is poor (in fact initially it increases rapidly) due to the nature of the detector and the fact that trees are not well modeled by the second-order Markov assumption.

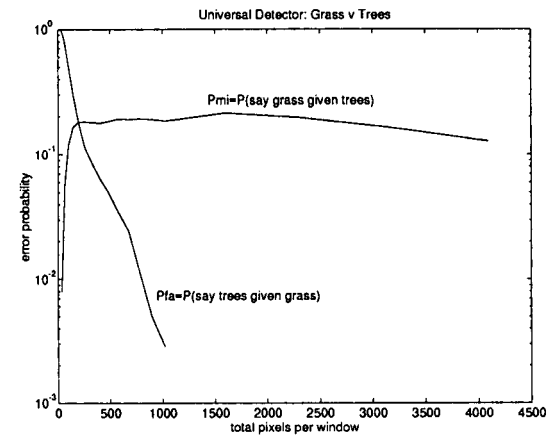


FIGURE 9. Asymptotic Optimal detector using a 2nd order Markov model

5.4 Asymptotic Equivalence and Conclusions

The minimum achievable probability of error for a given data set and sample size may be met by an optimal detection scheme for that data set. In practice, most detection schemes are suboptimal, and we now need to ask the question: how fast does the error probability go to zero for a given classifier and data set as the sample size increases, and what is the maximum rate at which the error probability can go to zero?

Suppose we have two competing ATR schemes, one of which has a probability of error of say 5% and the other 20% on a 2x2 window. If the former scheme has an error probability of 4% for say a 4x4 window, whereas the error probability of the latter falls to 1%, then clearly the latter scheme benefits from increasing the sample size and will then be a better detector. As the sample size increases, it may not be feasible to estimate the error performance (that is, the error rate is so low that no errors occur given the test data). The example given shows that it is not sufficient to merely measure the error performance for a smaller test window - the rate at which the error drops is also required. It has been shown that the error probability decreases exponentially; hence, the rate of decrease is expressed as an exponent of base 2.

We present the case for independent identically distributed data vectors which can be generalized to the dependent case. For a Bayesian detector, the maximum error exponent can be shown to be the Chernoff distance:[16]

$$\lim_{N\to\infty} \frac{1}{N} \log_2 \min P_e^N = -\, d_c(P_1,P_2)$$

$$d_c(P_1,P_2) = \max_r \left[-\log_2 \int_{R^N} p_1^r(x) p_2^{1-r}(x)\,dx \right] \tag{36}$$

This rate will be met by an optimal detection scheme such as the Likelihood Ratio Test or its asymptotic equivalent. The actual error exponent of a detector can be estimated by simulation and semilog regression, and can not asymptotically exceed the Chernoff distance between the competing distributions.

For Neyman-Pearson type detectors, for which one error rate (say the false alarm exponent) is fixed, the other error exponent (say the miss exponent) has been shown by Stein's lemma to approach the Kullback-Leibler distance as the false alarm rate tends to zero. If we define the false alarm probability to be α, which is to be bounded from above by ε and the minimum probability of miss to be β for some sample size N given the constraint on α, then:

$$\lim_{\varepsilon\to 0} \lim_{N\to\infty} \frac{1}{N} \log_2 \beta = -D_{KL}(P_1,P_2)$$

$$\beta = \min_{\alpha<\varepsilon} P_M \tag{37}$$

It can be seen in Figure 5 that the likelihood ratio detector error probability has approximately the same slope as the Bhattacharyya upper bound. For this example, the Bhattacharyya and Chernoff distances are very close ($S_{Bhauacharyya}$ = 0.5, $s_{Chemoff}$ = 0.45), hence the likelihood ratio detector has been shown to be optimal for this example in two ways, as the error probability,

- lies within the computed bounds on the minimum achievable error probability, and
- falls off at (or near) the fastest possible rate.

5.4.1 Example Analysis

To show this relationship, the maximum error exponent was plotted along with the LRT (Kullback-Leibler) using a resubstitution [Because of the number of computations involved only the resubstitution experiment was performed] estimate of the Bayes for various window sizes (Figure 10). These estimates were sampled to insure more independent samples. The results indicate an exponential decay rate for the Pe estimate as a function of the increasing observational frames. If the data were truly independent, then this would represent the optimal achievable performance.

However, because of the data dependencies, a higher dimensional model estimation technique is required. The Parzen kernel is an established approach, but the covariance is not sufficient to capture the higher dependencies. Similarly, diagonalization methods which uncorrelate the data suffer from the same

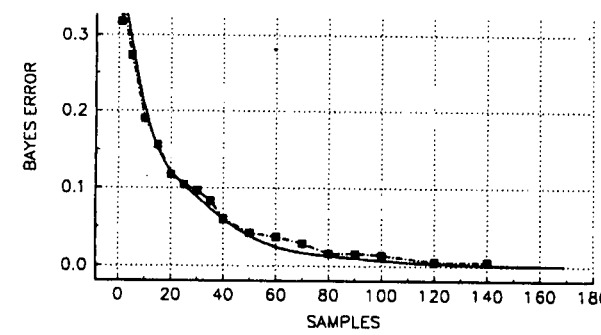


FIGURE 10. Simple Procedure for Estimating an N-pixel Detector Performance

assumption. Now of course, there are, finite Markov and finite context models that use inter-sample dependency to describe structures that have been used with considerable success in speech and text recognition, but applications for processing 2-D data have yet to be exploited.

Nevertheless, assuming a sufficient model, the same technique can be extrapolated to the individual errors P_D and P_{FA}, or even the conditional densities themselves $(\mathbf{P}\mathbf{L}|H_2)$ and $P(\mathbf{L}|H_1)$. This could provide a family of ROC curves that would represent an upper bound on any N-pixel classifier.

Summarizing, the following observations can be made:

- It is instructive to state performance in terms of the number of pixels per detector for fair comparison. The classifier that can utilize more samples of information should clearly perform better. Describing the performance measure in this fashion offers the ability to answer the critical question, "when is sensor resolution good enough?"
- Asymptotic equivalence indicates that evidence combined from multiple sensors and multiple looks should increase exponentially in performance. The challenge is how to combine this information with a classifier.

[1] J. Neyman and E.S. Pearson, "On the Problem of the Most Efficient tests of Statistical Hypotheses," Phil. trans. Roy. Soc. London, A 231, 289, (1933).

[2] Michéle Basseville, "Distance Measures for Signal Processing and Pattern Recognition," Signal Processing 18(1990) pp. 349-369.

]3] CH Chen, "On Information and Distances Measures, Error Bounds, and Feature Selection," Information Sciences 10, pp. 159-177 (1976)

[4] CH Chen, "On Information and Distances Measures, Error Bounds, and Feature Selection," Information Sciences 10, pp. 159-177 (1976)

[5] I Csiszár, "Information-type measures of difference of probability distributions and indirect observations," Studia Set. Math. Hungar. 2 pp 299-318 (1967)

[6] DE Boekee & JCA van der Lubbe, "Some Aspects of Error Bounds in Feature Selection," Pattern Recognition Vol II pp 353-360.

[7] DE Boekee & JC Ruitenbeek, "A Class of Lower Bounds on the Bayesian Probability of Error," Information Sciences 25 pp 21-35 (1981)

[8] Devijver, P., Kittler, J., Pattern Recognition: A Statistical Approach, Prentice-Hall International, London, 1982, pp. 345

[9] Devijever, P., Kittler, J., Pattern Recognition: A Statistical Approach, Prentice-Hall International, London, 1982, pp. 343-378

[10] K.S. Shanmugam and P. Balaban, " A Modified Monte-Carlo Simulation Technique for the Evaluation of Error Rate in Digital Communication Systems," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1916-1924, Nov. 1980

[11] G.C. Orsak and B. Aazhang, "On the Theory of Importance Sampling Applied to the Analysis of Detection Systems," *IEEE Trans. Commun.*, vol. COM-37, no. 4, pp. 332-339, April 1989.

[12] D. Lu and K. Yao, "Improved Importance Sampling Technique for Efficient Simulation of Digital Communication Systems," *IEEE J. on Selec. Areas in Comm.*, Jan 1988.

[13] M. Cotrell, J. Fort, and G. Malgouyres, "Large Deviations and Rare Events in the Study of Stochastic Algorithms," *IEEE Trans. Auto. Control*, vol. AC-28, no. 9, pp 907-920, Sept. 1983.

[14] S. Sadowsky and JA Bucklew, 'On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation," *IEEE Trans. Inform. Theory,* vol. IT-36, pp. 579-588, 1990.

[15] Devijver, P., Kittler, J., Pattern Recognition: A Statistical Approach, Prentice-Hall International, London, 1982, pp. 425-432.

[16] Thomas M. Cover & Joy A. Thomas, Elements of Information Theory, 1991 John Wiley & Sons, Inc., ISBN 0-471-06259-6.

THE MODULAR ALGORITHM CONCEPT EVALUATION TOOL (MACET) WORKSTATION

John Watson, Brad Williams, Sunjay Talele
Nichols Research Corporation
One Eleventh Avenue, Suite C-1
Shalimar, FL 32548

Sengvieng Amphay
Department of the Air Force
Wright Laboratory WL/MNGA
101 West Eglin Blvd.
EAFB, FL 32542

ABSTRACT

Target acquisition in a high clutter environment in all-weather at any time of day represents a much needed capability for the air-to-surface strike mission. A considerable amount of the research at the Armament Directorate at Wright Laboratory, Advanced Guidance Division WL/MNG, has been devoted to exploring various seeker technologies, including multi-spectral sensor fusion, that may yield a cost efficient system with these capabilities. Critical elements of any such seekers are the autonomous target acquisition and tracking algorithms. These algorithms will allow the weapon system to operate independently and accurately in realistic battlefield scenarios.

In order to assess the performance of the multi-spectral sensor fusion algorithms being produced as part of the seeker technology development programs, the Munition Processing Technology Branch of WL/MN is developing an algorithm testbed. This testbed consists of the Irma signature prediction model, data analysis workstations, such as the TABILS Analysis and Management System (TAMS), and the Modular Algorithm Concept Evaluation Tool (MACET) algorithm workstation. All three of these components are being enhanced to accommodate multi-spectral sensor fusion systems. MACET is being developed to provide a graphical interface driven simulation by which to quickly configure algorithm components and conduct performance evaluations. MACET is being developed incrementally with each release providing an additional channel of operation. To date MACET 1.0, a passive IR algorithm environment, has been delivered. The second release, MACET 1.1 will be presented in this paper using the MMW/IR data from the Advanced Autonomous Dual Mode Seeker (AADMS) captive flight demonstration. Once completed, the delivered software from past algorithm development efforts will be converted to the MACET library format, thereby providing an on-line database of the algorithm research conducted to date.

1.0 INTRODUCTION AND OVERVIEW

The Wright Laboratory Armament Directorate at Eglin AFB, FL has been actively involved in research for the purpose of understanding the phenomenology exploitable for smart weapon systems guidance for several decades. This research has evolved along with sensor technology from approaches using unresolved infrared signatures to high resolution, multi-sensor imagery. The knowledge-base which has been continually enhanced as a result of these efforts is critical to the development of robust acquisition and tracking algorithms. It is these algorithms which provide smart weapons the capability of autonomous, precision guidance.

The development of all-weather, time-of-day, terrain insensitive algorithms lends itself to a physics-based approach, in which reliable algorithm features can be derived from the target signature collected in a single spectral regime or over multiple bands [1]. Key to a physics-based algorithm development effort are tools which support the analysis of measured data, construction/refinement of models which realistically describe the physical processes associated with the target/background signatures, and simulations for the purpose of exercising evolving algorithm concepts. Such an algorithm development toolbox is required to support a systematic exploration of the scattering/emitting process inherent to target/background signatures, and to develop optimized algorithms which exploit this phenomenology.

The Munition Processing Technology Branch of WL/MNG is developing a multi-sensor algorithm testbed

to support precision guided weapons research. This testbed consists of the Irma signature model, data analysis workstations such as the TABILS Analysis and Management Systems (TAMS), and the Modular Algorithm Concept Evaluation Tool (MACET) algorithm workstation. All three of these components are being upgraded to accommodate multi-spectral sensor fusion systems. Both ladar and passive MMW channels [2,3] have been added to the baseline Irma IR signature prediction model and a four channel version is currently in prototype. Likewise, the MACET system has been upgraded to accommodate MMW/IR data fusion with later releases to be four channel capable.

MACET is being developed to provide a graphical interface driven simulation by which to rapidly prototype and evaluate acquisition and aimpoint selection algorithms. MACET is a user-friendly, graphics-based, software system developed to run on a Sun Workstation [4]. MACET was designed to support both unitary and multi-sensor fused configurations. An incremental developmental approach was adopted with interim deliveries of the product occurring every six months. The purpose of this paper is to present the MACET concept, describe its development plans, and demonstrate the progress realized to date.

2.0 MACET OVERVIEW

MACET is intended to be a user friendly software system to support the rapid prototyping and evaluation of air-to-surface acquisition and aimpoint selection algorithms. In order to facilitate this function, the following design requirements were specified for MACET: minimum training time, rapid prototyping capability, compatibility with existing data, capability to accept existing algorithm components, algorithm evaluation shell, multi-sensor algorithms, emulation of target acquisition and aimpoint selection algorithms, Sun Workstation platform, nominal development costs, generation of standard algorithm performance metrics, flexible graphical output, and data probe placement [5].

From these design requirements, it was determined that MACET would be constrained to execute on a Sun Workstation and must utilize existing software platform(s) in order to maintain nominal development costs. Other required capabilities include: a graphical user interface, the establishment of a standard data format, a computer aided algorithm design capability, and an ethernet interface with the VAX and PC systems.

The MACET concept is illustrated in Figure 2.1. A Sun SPARC Station 2 serves as the host platform for MACET. An ethernet is used to link the Sun with VAX systems located within the Radar Signal Processing Laboratory (RSPL) and Imaging Processing Laboratory (IPL) where existing data reduction and analysis utilities reside. Measured data from the TABILS database, laboratory tower and captive flight test exercises are used to exercise the prototype algorithm configurations. MMW and IR synthetic data from signature prediction codes, such as Irma, are also used for this purpose [6]. MACET will also contain a library of algorithm components generated during previous research efforts under 6.2, 6.3, and Small Business Innovative Research (SBIR) programs such as the Advanced Tactical LADAR Seeker (ATLAS), Low Cost Anti-Armor Submunitions (LOCAAS), Joint Adverse Weather Seeker (JAWS), Dual Mode Seeker (DMS), and others; thereby providing on-line access to over a decade of research in the area of autonomous acquisition algorithms for air-to-surface guidance. A graphical user interface is employed to minimize the system training time required of engineers and analysts to use the tool.

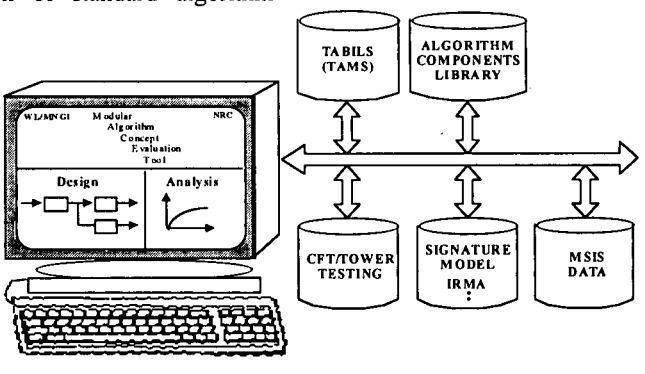


Figure 2.1 MACET Concept

The MACET system is developed upon the Paragon Image Logic and Khoros 1.0 software platforms. The Paragon Image Logic was modified to support the top-level functionality of the MACET architecture including the user interface. Signal and image processing routines from Khoros are used to provide lower level library and functional routines for the purpose of algorithm prototyping and output display. This configuration was selected for MACET development because Paragon was found to have the better graphical interface and offered better user support at that time. However, given the similarity of the Khoros and Paragon environments, subsequent releases of these two software packages have been closely monitored in order to take advantage of any increased functionality.

The MACET architecture is illustrated in Figure 2.2. As shown in this graphic, MACET consists of six major functional areas. These functional areas include: ground truth editor, data selection/format conversion, algorithm selection, performance evaluation, output selection, and utilities. The ground truth editor provides the capability to tag targets in the data scene before injection into the signal processing algorithm in order to facilitate performance scoring. The data selection/format conversion function allows the user to select a data set for algorithm testing and will convert all selected data to a format compatible with other MACET functions. The algorithm selection function facilitates algorithm prototyping using either existing algorithms or algorithm components provided in an on-line library or by providing the capability to define new elements. Once the algorithm has been defined and the data selected, the performance evaluation function can be activated. The purpose of this function is to compute defined algorithm performance metrics (such as probability of detection/probability of false alarm) or user defined metrics via probe placement. Upon completion of algorithm testing, the resulting metrics can be viewed

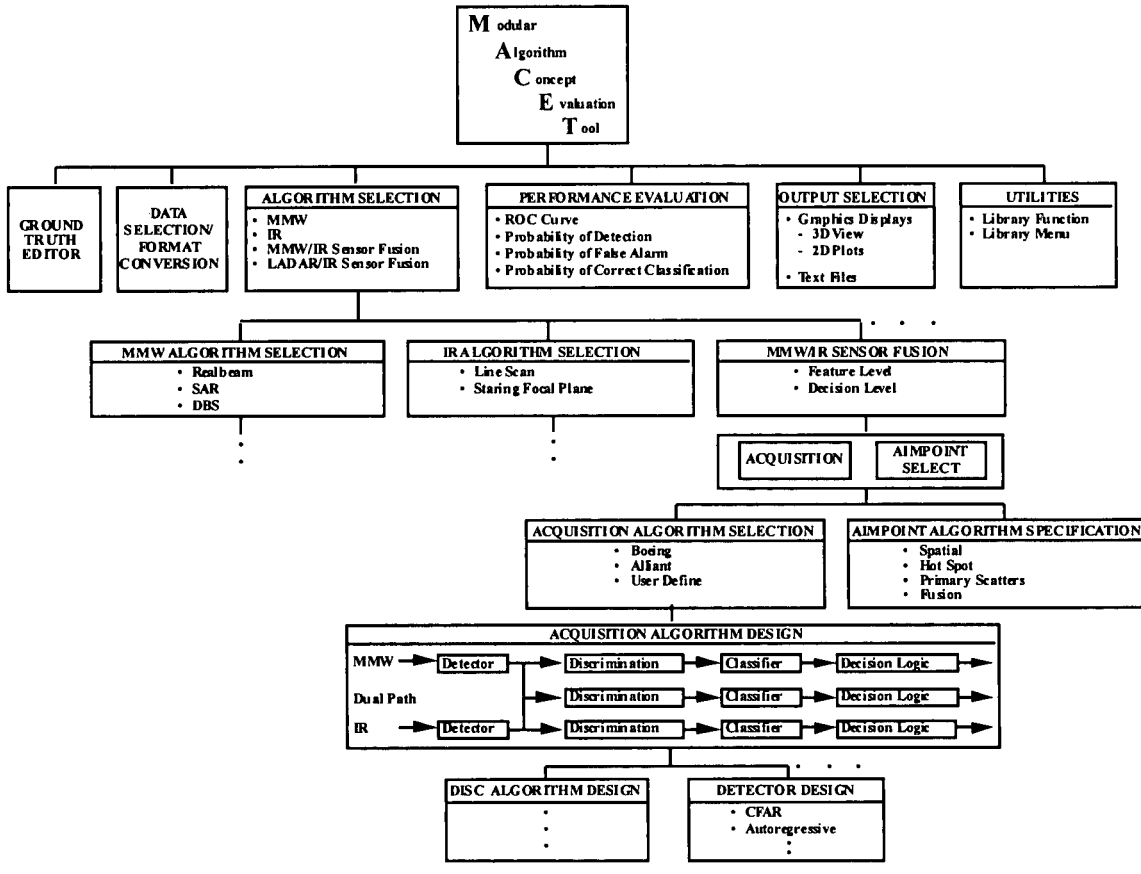


Figure 2.2 MACET Architecture

using the output selection function. Both graphical and text output are supported. Lastly, a utilities function is provided for workspace manipulation and accessing lower level MACET data processing routines.

Development of the MACET simulation is well underway with significant progress being realized in the areas of data conversion, ground truth editor, performance evaluation, and output selection. A schedule of MACET software releases is provided in Figure 2.3. As indicated by this chart, incremental versions of the MACET will be delivered as additional channels (sensor types) are added. This incremental development approach will result in a comprehensive active/passive IR/MMW version, capable of emulating either unitary or sensor-fused systems, to be delivered at the conclusion of the Data Analysis and Modeling (DAAM) contract in November 1995. This simulation will provide the flexible non-realtime emulation by which to rapidly prototype and evaluate unitary and sensor fused algorithms.

3.0 GROUND TRUTH EDITOR

A key element of the MACET architecture is the Ground Truth Editor. The purpose of the editor is to provide the user with the ability to designate targets and aimpoints within scenes under test including multi-spectral and multi-sensor imagery. This sensor independent truth data, along with computed scene statistics, is then attached to the image files under test for later use in algorithm scoring and performance curve generation.

3.1 *Design Philosophy*

The Ground Truth Editor's (GTE) design is based on the same fundamental goals as the remainder of the MACET system, i.e., provision of a powerful and flexible tool which remains simple and intuitive. To that end, the graphical user interface (GUI) of the GTE is similar to that of many contemporary computer aided drawing packages (see Figure 3.1). A palette of tools is displayed as a column of icons which graphically depict the function of each. Other functions are readily available via the menu system which remains consistent with the GTE's graphical user interface. The system runs under the X11 windowing system and was built using the industry standard Motif toolkit. This assures maximum portability across platforms and provides consistency with the MACET environment as well as most other X based applications. The MACET system is intended to be functionally intuitive so as to require minimal training investment prior to productive use. This philosophy is embraced in the user friendly design of the GTE.

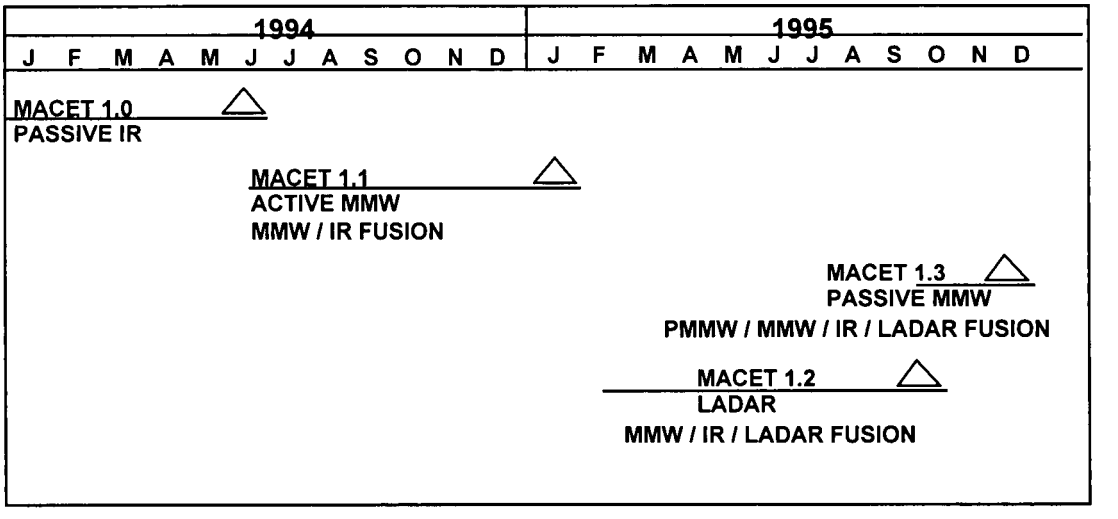


Figure 2.3. MACET Development Schedule

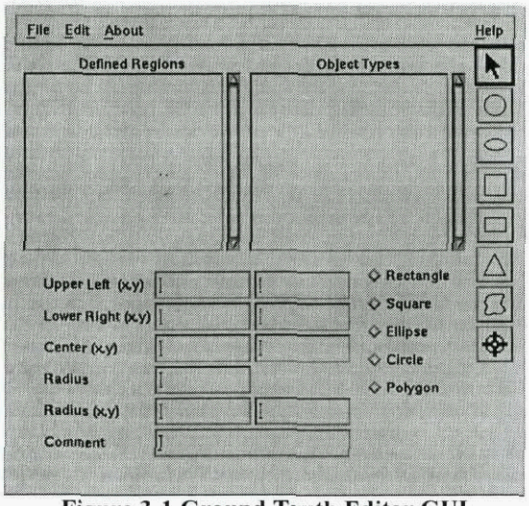


Figure 3-1 Ground Truth Editor GUI

3.2 File Format

The GTE/MACET design philosophy is further exemplified in the file formats adopted for ground truth and data. In the interest of maintaining compatibility with the widest possible cross section of data formats, the MACET data format was chosen to be compatible with the XVIFF format of Khoros. The XVIFF format was chosen not only because of the proliferation of the Khoros package throughout the image processing community, but because of the versatility of the format and readily available conversion programs. The MACET format (MVIFF) differs from XVIFF only when ground truth data is associated with the given "image", and then only in that a small portion at the end of the 512 byte comment field is used for ground truth related data. Khoros routines and any routines designed to operate on XVIFF data will function equally well with MVIFF data.

The ground truth file itself is an ASCII text file which could be edited by any conventional text editor if desired; however, the file is typically maintained solely by the GTE. An excerpt from an example ground truth file is shown in Figure 3.2. A ground truth file does not reference any specific images or data sets. It is in itself not constrained to data type or view related parameters such as pixel dimension, aspect ratio, or resolution and is thus completely sensor independent. Restricting the ground truth file to "absolute" truth information allows it to be referenced by unlimited numbers and types of data files without regard to sensor type. Included in a ground truth file is an absolute reference point specified by latitude and longitude. All target and feature locations specified within the file are given in meters offset from the reference with positively increasing x moving to the east and positively increasing y moving to the south. Each MVIFF data or image file specifies its vertical and horizontal orientation and the offset in meters of the center of upper left pixel from the reference specified in the ground truth file. The XVIFF format provides for specification of pixel dimensions in meters and these

```
%MGT1

# Example ground truth file

longitude       87.4
latitude        30.6
filename        "tcfl01"
object 9        "Red Square"
object 10       "Yellow Circle"
object 11       "Yellow Square"
object 12       "T62"
object 13       "BMP"
object 14       "Scud Launcher"
object 21       "Truck"
object 22       "Holes"
object 23       "Resolution Panels"

# Vehicles in the scene are listed below ...

region 23       rectangle [(110,120),(130,140), "Rectangle
comment"]
region 12       circle [(332,316),9,"Circle comment"]
region 13       circle [(431,132),6.5,"Circle comment"]
```

Figure 3.2 Example Ground Truth File

fields are utilized for that purpose in MVIFF as well. Other ground truth related information associated with an image or data set, including the ground truth file name, are inserted at the end of the 512 byte comment field. The beginning of this data is denoted by "%MVIFF1".

The autonomy of the ground truth file provides for efficient representation of ground truth. Since the information linking a given data set or image to its corresponding ground truth is minimal and is stored in pre-assigned fields within an XVIFF file, there is essentially no additional overhead required for incorporating ground truth. One ground truth file can be shared by many and diverse types of images as is illustrated in Figure 3.3. It is feasible that a single ground truth file describing a test range could be used for multiple test runs using various sensors. Corrections to the ground truth need only be made once rather than for each image.

3.3 Ground Truth Editing

In many situations the ground "truth" must be supplied by interpreting the data set itself, or from images made in conjunction with data collection. The single ground truth file system allows editing of ground truth from any registered image. A typical scenario might consist of digitized video co-registered with IR or MMW sensors. The video would be provided for extraction of ground truth. For this scenario, the digitized video imagery would first be loaded into the GTE. If surveyed scoring panels or land marks are available, these would be used to specify the ground reference for the ground truth file. The markings are selected in the image with the cursor and registered by making the "register" selection from the menu. Multiple points may be used for registration, in which case an average is calculated for positioning, and orientation is determined by best fit. Registration between images (relative truthing) is accomplished by simply selecting a point in each image and making the "register" selection. Multiple points are handled in the same manner as in absolute truthing. After an initial image is registered, additional images which are loaded and not yet associated with a ground truth file can be set to default to identical parameters. All that is needed to register a sequence of images is to load each and register relative to the preceding image using a common point in an overlapping region. If the relationship is defined and known between images such as is the case with co-registered data, that relationship can be automatically incorporated into an entire family of images so registering each individually is not necessary. If the relationship is initially unknown, but consistent, it can be established with registration techniques (e.g. registering corner reflectors visible in IR or visible data with MMW returns), and then incorporated in the remaining images. After data sets (MVIFF files) are registered, designation of targets and aim points can be accomplished using registered data and will appear appropriately scaled in all corresponding data.

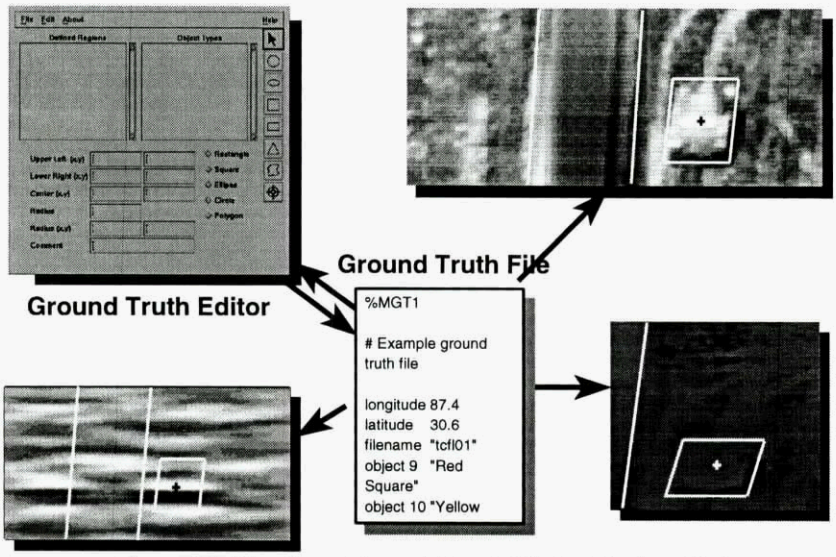


Figure 3.3 The Same Ground Truth File for Various Data

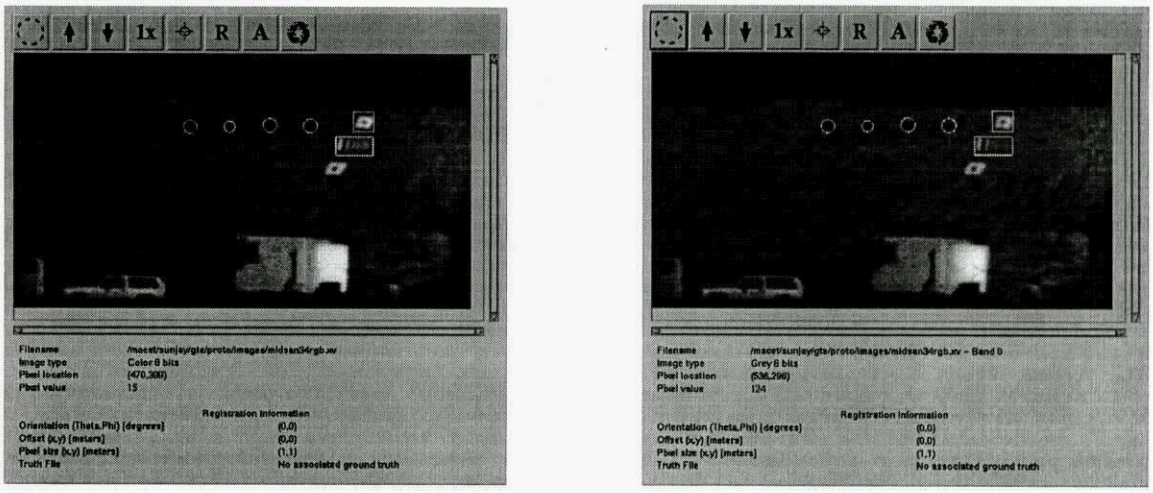


Figure 3.4 Co-registered Visible (a) and IR(b) Images of Minefield

3.4 Multi-Spectral Data Fusion Examples

Several examples of multi-spectral images that have been truthed using the MACET Ground Truth Editor are provided in this section. These include multi-band visible and IR images, as well as IR/MMW data. A discussion of LADAR imagery is also provided.

3.4.1 *Multi-band Visible / IR*

The Ground Truth Editor can manipulate multi-banded images, either individually, or by merging bands into single images. Figure 3.4 illustrates an RG image of a minefield, along with the corresponding IR band. The RGB image was obtained by merging three visible bands. This band merging is done by the ground truth editor at the request of the user. In this example, the images were created in a co-registered manner, so no registration information needed to be added by the user.

The corresponding ground truth data for the images, showing the object types and regions of interest, are displayed in the main window of the ground truth editor, as shown in Figure 3.5

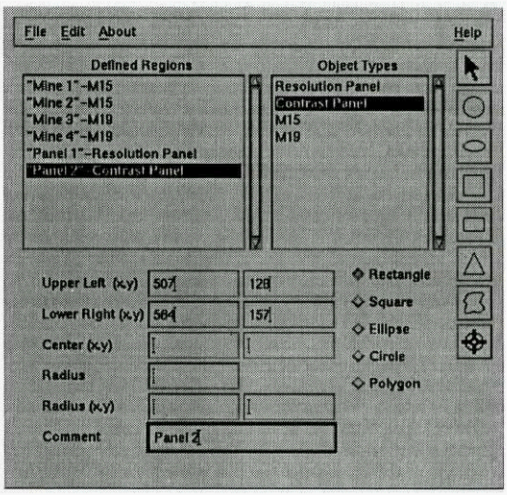


Figure 3.5 Ground Truth File for Visible/IR Image

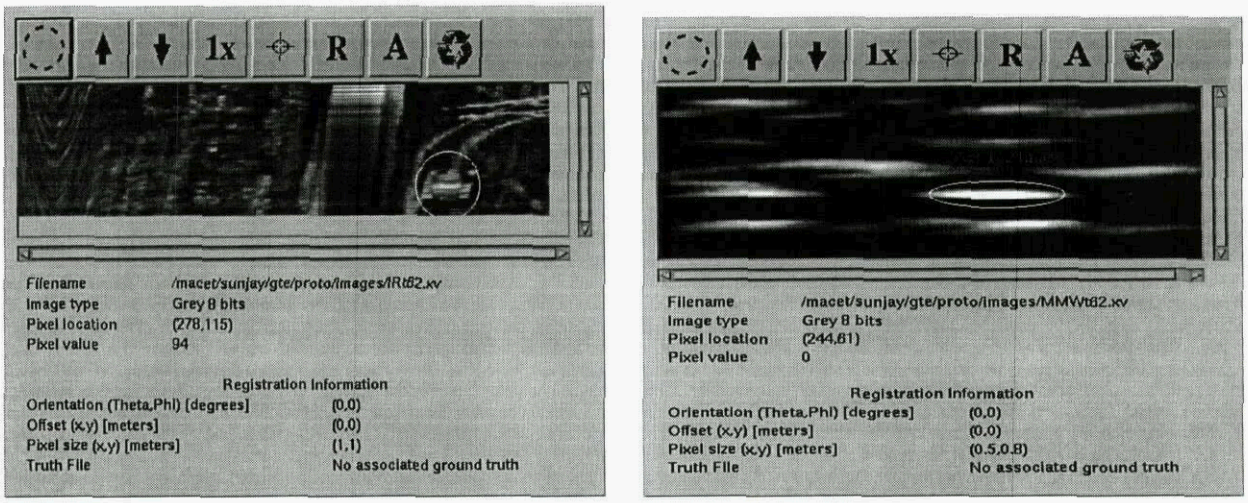


Figure 3.6 Co-registered IR (a) and MMW (b) Images of T-62 Tank

3.4.2 *IR / MMW Data*

Figure 3.6 shows an IR image of a T62 tank, along with the corresponding millimeter wave image [6]. The images represent different fields of view; therefore, all ground truth designation has to be appropriately rescaled. By default, the ground truth editor assumes that an image pixel represents a square meter. Given this is not the case in the millimeter wave image, the pixel size in the registration information has to be updated.

After registering the images, regions of interest can then be defined. Since a new truth file is being created, the user must first define object types for the ground truth. Since the only object of interest in the images is the tank, this can be accomplished by simply adding the T62 object type in the main truth window depicted in Figure 3.7. Next, a circular region enclosing the T62 in the IR image is defined, as shown in Figure 3.6(a). Note that the region is translated and scaled automatically in the millimeter wave image shown in Figure 3.6(b). Users may add comments to each region of interest, which are displayed as annotations in the image windows.

3.4.3 *LADAR Data*

Currently, the ground truth editor is only capable of handling two dimensional data. Adding functionality for three-dimensional data, such as the LADAR imagery depicted in Figure 3.8, would not be difficult given the flexible formats of both the images and the ground truth data. The MVIFF format has three-dimensional capabilities already built in, and the ground truth format could easily be extended (in a backward-compatible fashion) to provide this capability. These modifications will be included in the next MACET release.

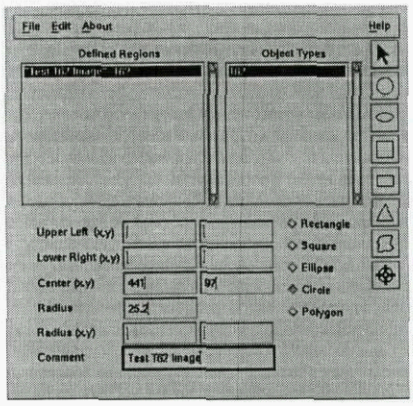


Figure 3.7 Ground Truth for T62 Image

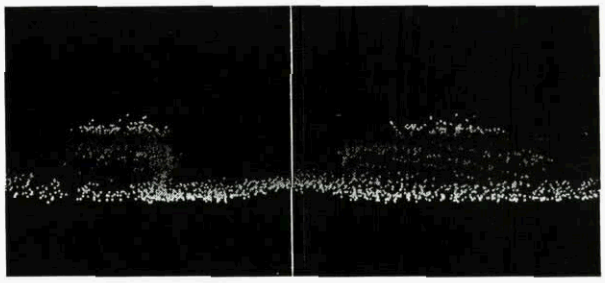


Figure 3.8 LADAR Image of Tank (front and side views)

4.0 PROTOTYPE ALGORITHMS

Prototype algorithms have been developed with each MACET release for the purpose of providing generic baseline templates. These are functional algorithms in themselves, but are not intended to replace more robust and task specific systems. The supplied algorithms provide a baseline structure for algorithm development and illustrate methods of incorporation using MACET. These algorithms are all based on classical statistical techniques, however, the MACET environment is not limited to these designs. Fuzzy set technique and neural network implementations are possible and will be straight forward to implement because of the modular nature of the tool. At the present stage of development, MACET includes three algorithm templates: infrared, active real-beam millimeter wave, and dual-mode IR/MMW. Each of these incorporates different examples of MACET capabilities so as not to make a given algorithm unnecessarily complex, but to demonstrate as completely as possible MACET functionality. Multiple orientations of a detected object are evaluated with the IR prototype algorithm, for example, but not in the IR/MMW dual-mode prototype. The dual-mode prototype, on the other hand, tests multiple objects extracted from a single image whereas only a single object is extracted with the IR example. The supplied prototype algorithms are described here in an effort to highlight MACET functionality.

4.1 Infrared

The infrared prototype algorithm is shown in Figure 4.1. It is composed of four basic components: Prefilter, Detect, Discriminate, and Classify. The components of the prefilter module and its sub-modules are shown in Figure 4.2. The prefilter module filters the image with a 1x3 sliding window mean filter implemented by selecting the optional mean image output of the "CFAR" routine. Further enhancement of target pixels is accomplished with a quadratic classifier module (PixelEnhance). This module implements a two class quadratic classifier over an image of texture feature vectors extracted with the "fractal_feature" module. Feature vectors representing every pixel in the image are transformed using predefined target and background transformation parameters into values of class discriminants. Differencing of the discriminant values at each pixel results in an image of log likelihood ratios which is passed on to the "Detect" module for further processing. The features extracted are comprised of statistics of neighborhoods of differing sizes surrounding each pixel. Means and standard deviations of three different sized windows centered at a given pixel define the six element feature vector corresponding to that pixel. These elements are computed by three instances of the "CFAR" routine and banded together in the "fractal_feature" module.

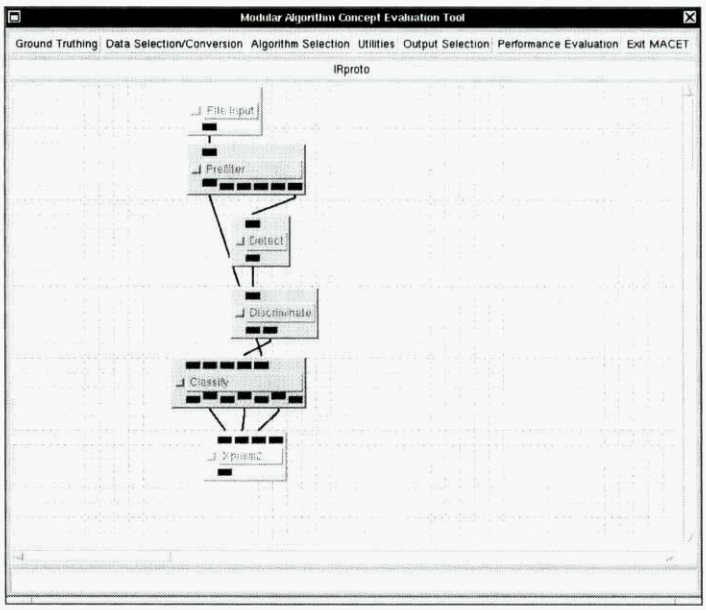


Figure 4.1 IR Prototype Algorithm

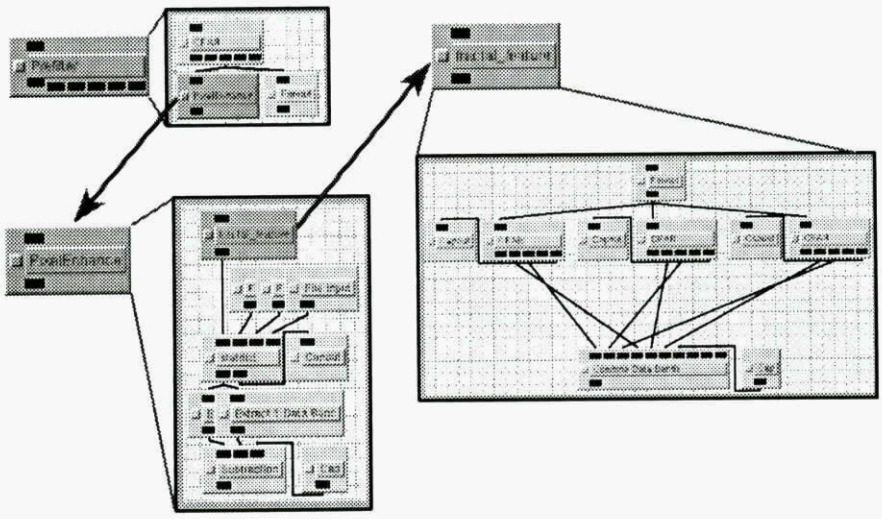
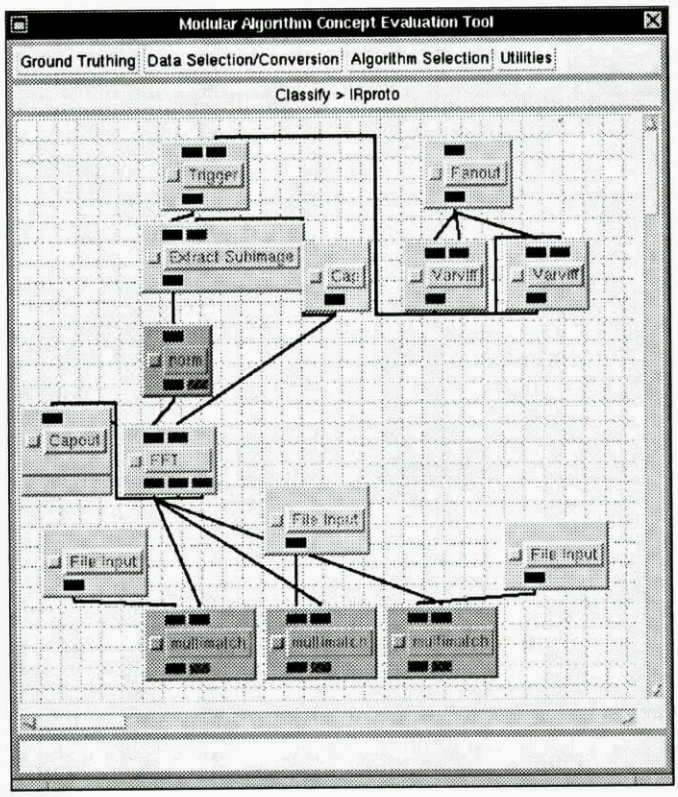


Figure 4.2 IR Prefilter Module

The enhanced image from the "Prefilter" module is presented to the "Detect" module which applies a sliding window mean filter (implemented again with the "CFAR" routine) of the approximate size of a target. This serves to deemphasize objects smaller than the expected target and accentuate targets. Output from this stage is processed in the "Discriminate" module which uses the routine "objseg" to segment groups of adjacent pixels which exceed a specified threshold into objects. These objects are further screened in "objseg" according to number of pixels, of which a minimum and maximum are specified. The output of "Discriminate" is a list of row/column coordinates of objects which meet the threshold and number of pixels criteria.



4.3 IR Classify Module

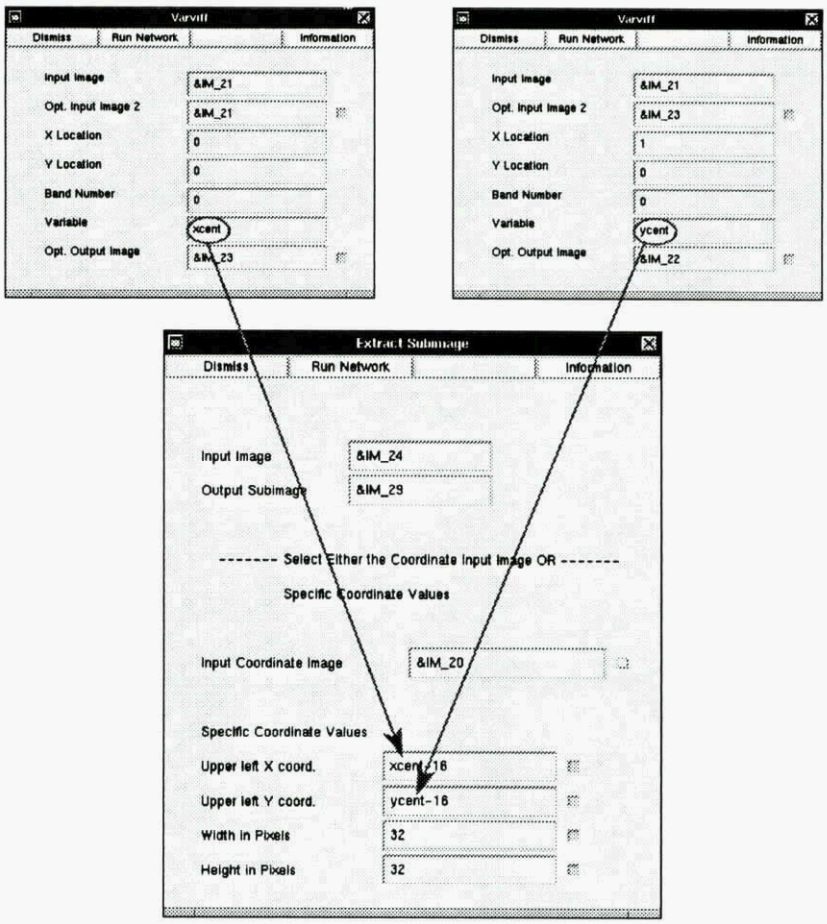


Figure 4.4 Variable Parameter Assignment

The "Classify" module is shown expanded in Figure 4.3. This module receives the object coordinate list from "Discriminate" and a filtered version of the original input image from "Prefilter". It extracts a region of interest centered at the object coordinates from the filtered image and compares this to pre-formed templates of candidate targets at various orientations. The region of interest extraction is accomplished using the "Extract Subimage" routine which requires image coordinates as parameters. MACET provides for output dependent parameters by allowing variables and expressions to specify parameters. Figure 4.4 shows the parameter entry forms for the routine "Extract Subimage" as well as the "Varviff which accomplish the glyphs variable assignments. The extracted region of interest is normalized to unity variance and zero mean prior to correlation with the target templates. This is accomplished in the "norm" module which is shown expanded in Figure 4.5. Correlation of the region of interest with potential targets is done in the transform domain. An FFT is therefore performed prior to presentation to the "multimatch" modules which accomplish the multiple orientation correlations. Three "multimatch" modules are necessary in order to perform the correlation operations with the three candidate target types. A target template which consists of twenty-four orientations of the particular target is presented to each module. These templates are the complex conjugates of the Fourier transforms of the given target data at the various orientations. The slant angle is constant. As seen in Figure 4.6, "multimatch" consists of two instances of the routine "multi" followed by "Image (2D) Stats", a statistics calculation routine. The routine "multi" allows functions which provide only for single band data

inputs to operate on data of multiple bands without the need of splitting out bands and looping or duplication. In this example, the first instance of "multi" invokes the multiply routine "vmul" to form the product of the transform of the region of interest image with each of the twenty-four bands or orientations contained in the template image. The second instance of "multi" calls "vfft" with the direction switch set so as to calculate the inverse transform of each orientation band. This produces a correlation image for each orientation all of which are banded together in a single structure. The statistics routine is designed to function with multi-band data, and is used here to determine the maximum value in each band. These are incorporated into a file which is made available external to the "multimatch" routine. An ascii file is also available which summarizes other statistical information.

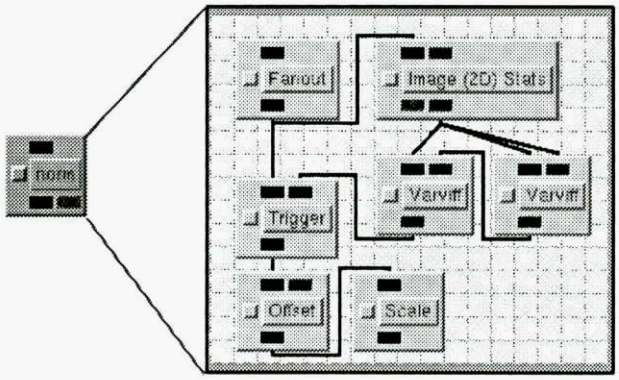


Figure 4.5 Normalization Module

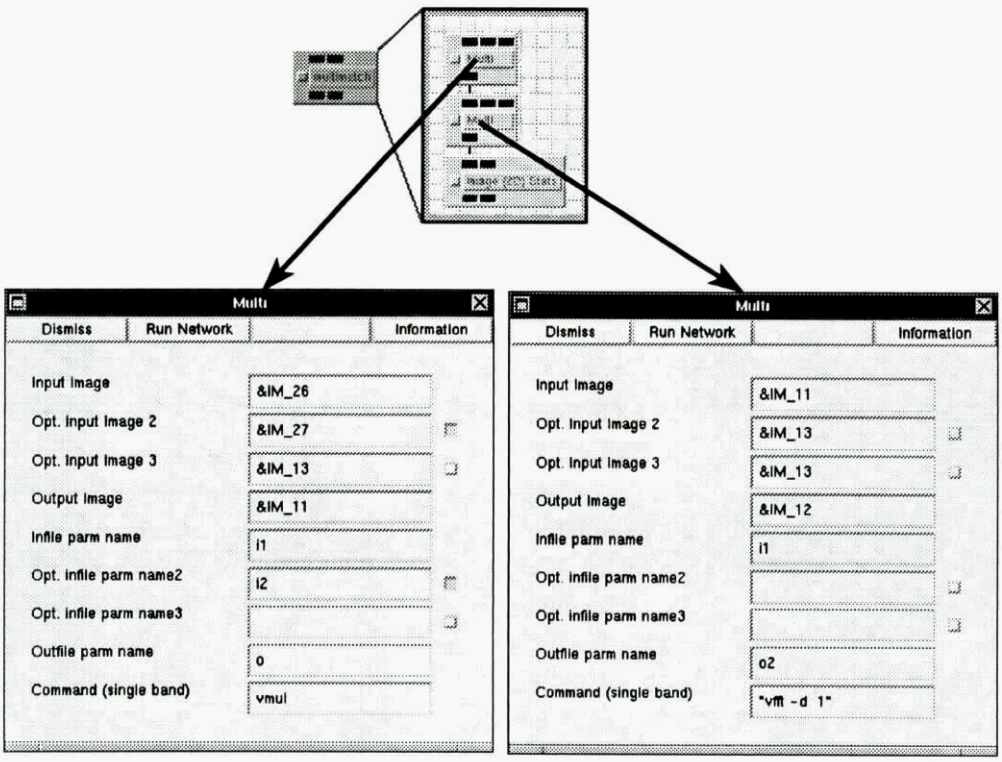


Figure 4.6 Multimatch Module

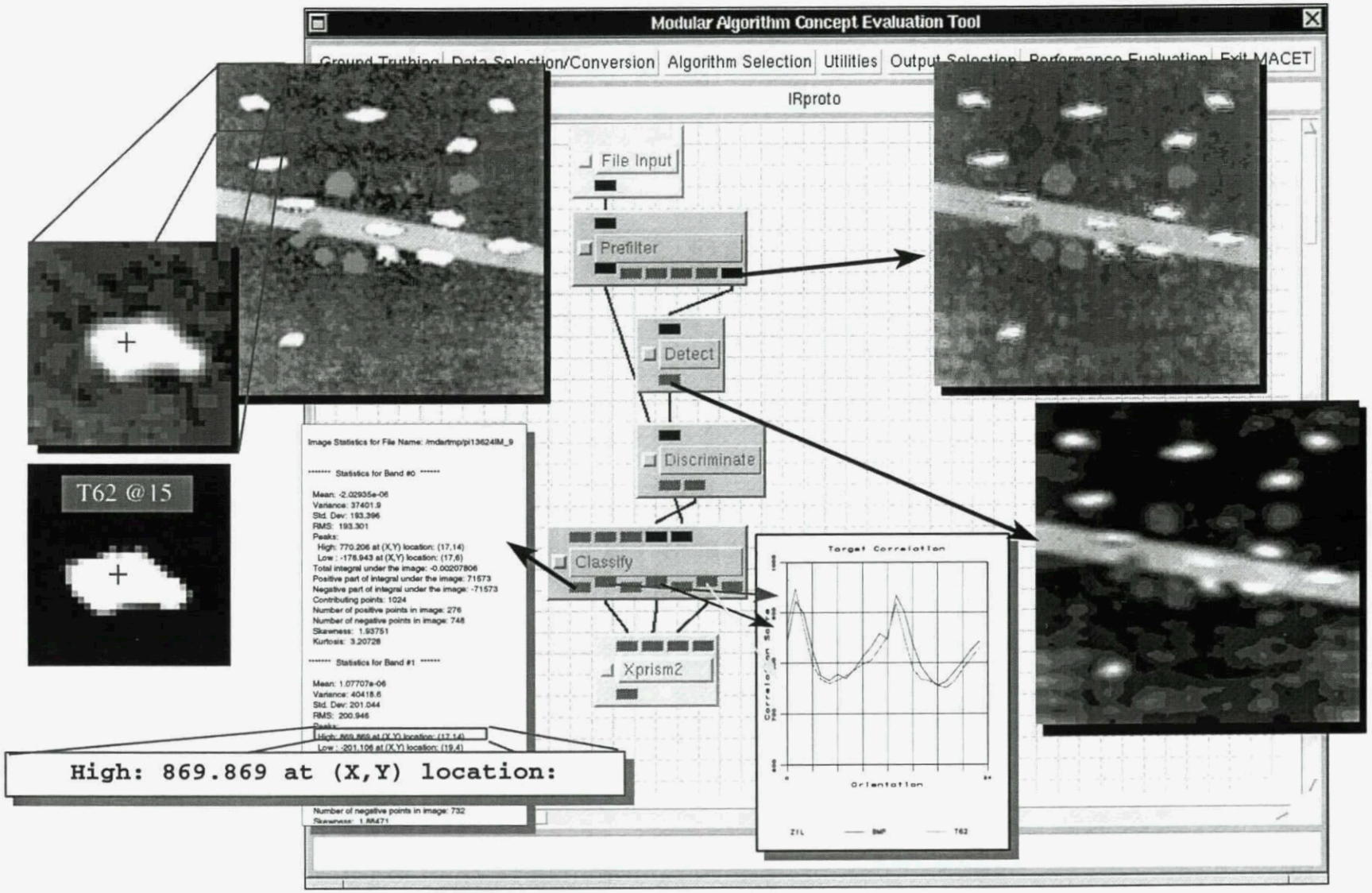


Figure 4.7 IR Prototype Algorithm and Intermediate Outputs

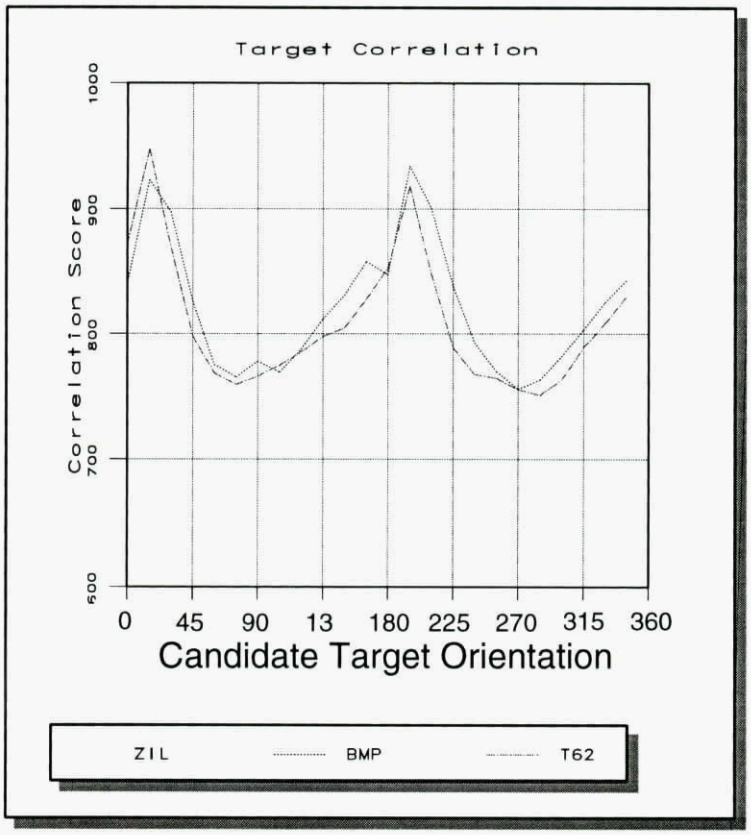


Figure 4.8 IR Target Correlation Plot

The maximum value outputs of the three instances of "multimatch" are forwarded out of the "Classify" module and presented to a plotting routine for display. The ascii files are also available for discerning additional information about the selected region of interest correlation file such as the position of the maximum. This corresponds to the point in the region of interest which aligns with the center of the target template, and could be used for aimpoint determination.

Intermediate output from different processing stages of the algorithm are shown in Figure 4.7. The input image is synthesized IR with three different target types (T62, BMP, and ZIL) at various positions and orientations. In order to simplify the prototype algorithm, only a single object is processed. The detected region of interest is shown along with the template which was chosen as the match and alignment points as determined from the ascii statistics output also shown in the figure. The output plots in Figure 4.8 show the peak correlation corresponds to a T62 oriented at approximately 15 degrees. The similarity between BMP and T62 is apparent from the close correspondence between plots. Although the ZIL is distinctly different a correlation according to orientation is obvious in all three targets.

4.2 Active Millimeter Wave

The active millimeter wave prototype algorithm in MACET is designed for use with real-beam radar data. The configuration shown in Figure 4.9 is intended for data captured in four contiguous dwells which are stored in separate files. The separate data are concatenated into a single file in the "DwellCat" module and passed to "DETECT" for the first stage of processing. This module correlates the data with a pattern which corresponds to the beam shape of the system in order to enhance the true returns while reducing noise. The processed data is then presented to the "Segment" module. In this stage, the data is correlated on a beam basis with a waveform template which represents the anticipated shape of the return from

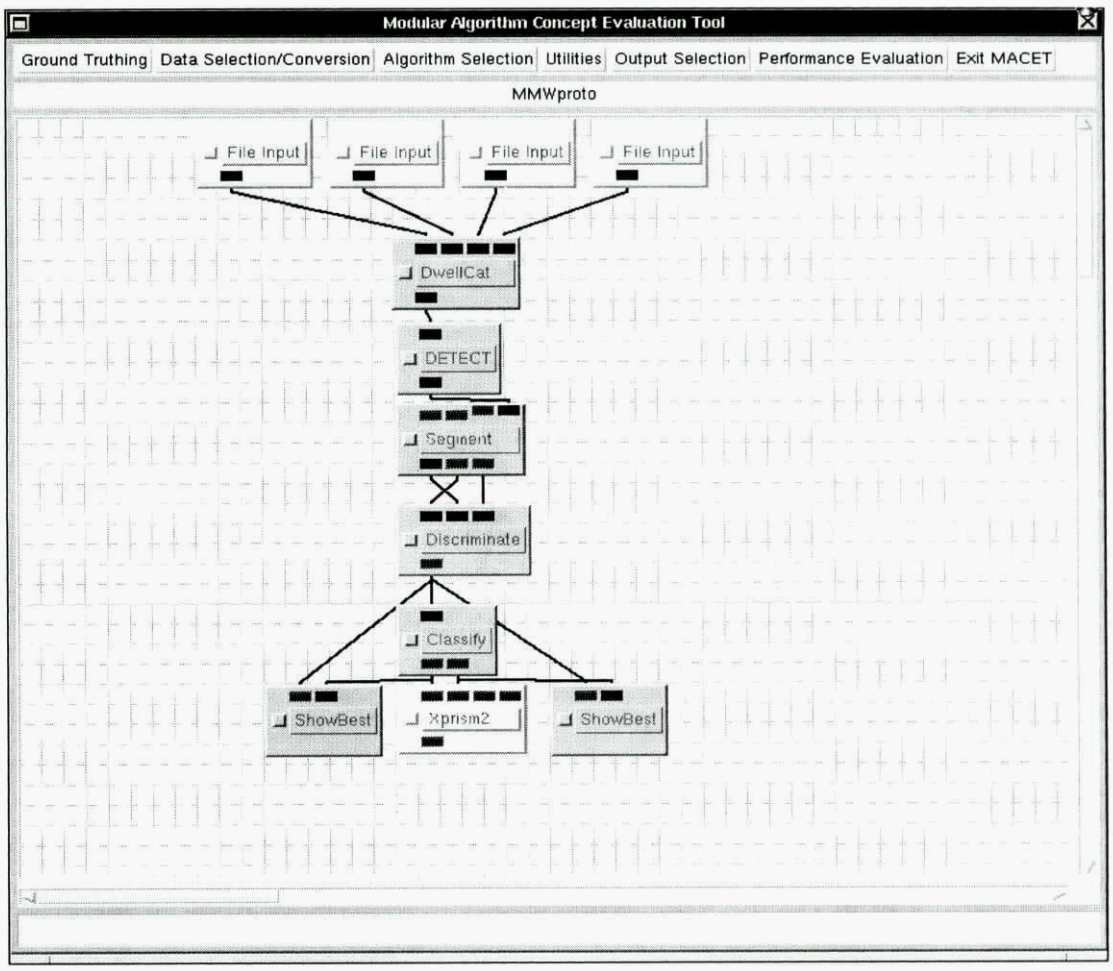


Figure 4.9 MMW Prototype Algorithm

a hard target. A threshold is applied to the output of the correlation operation, and neighboring data which exceed this threshold are formed into objects. The coordinates of these objects are passed to the next processing stage along with the processed data and an image of labeled objects. These data are received by the "Discriminate" module which uses the label image solely to determine the number of objects to process. The coordinates of the centers of the detected objects become the centers of regions of interest which are extracted and banded together to be processed by "Classify". The "Classify" module correlates these against predefined target templates in an attempt to classify the detected objects. The prototype algorithm provides for selection between two classes. "Classify" outputs the correlation score for each object in two files corresponding to the two target types. These files are plotted by "Xprism2", and made available to two "ShowBest" modules. The "ShowBest" modules find the object which corresponds to the highest correlation with a given target type, and displays the corresponding region of interest.

4.3 Dual-Mode

Elements of both IR and AMMW were modified and incorporated into a dual-mode millimeter wave and infrared prototype algorithm. The implementation of this algorithm is shown in Figure 4.10. Additional support logic beyond the single channel algorithms is necessary to accomplish appropriate scaling and registration between the millimeter wave and infrared data. The data set used is comprised of real-beam millimeter wave data sampled such that two IR rows correspond to a single range bin. The radar return beams are stored as rows, and hence must be rotated as well as scaled with respect

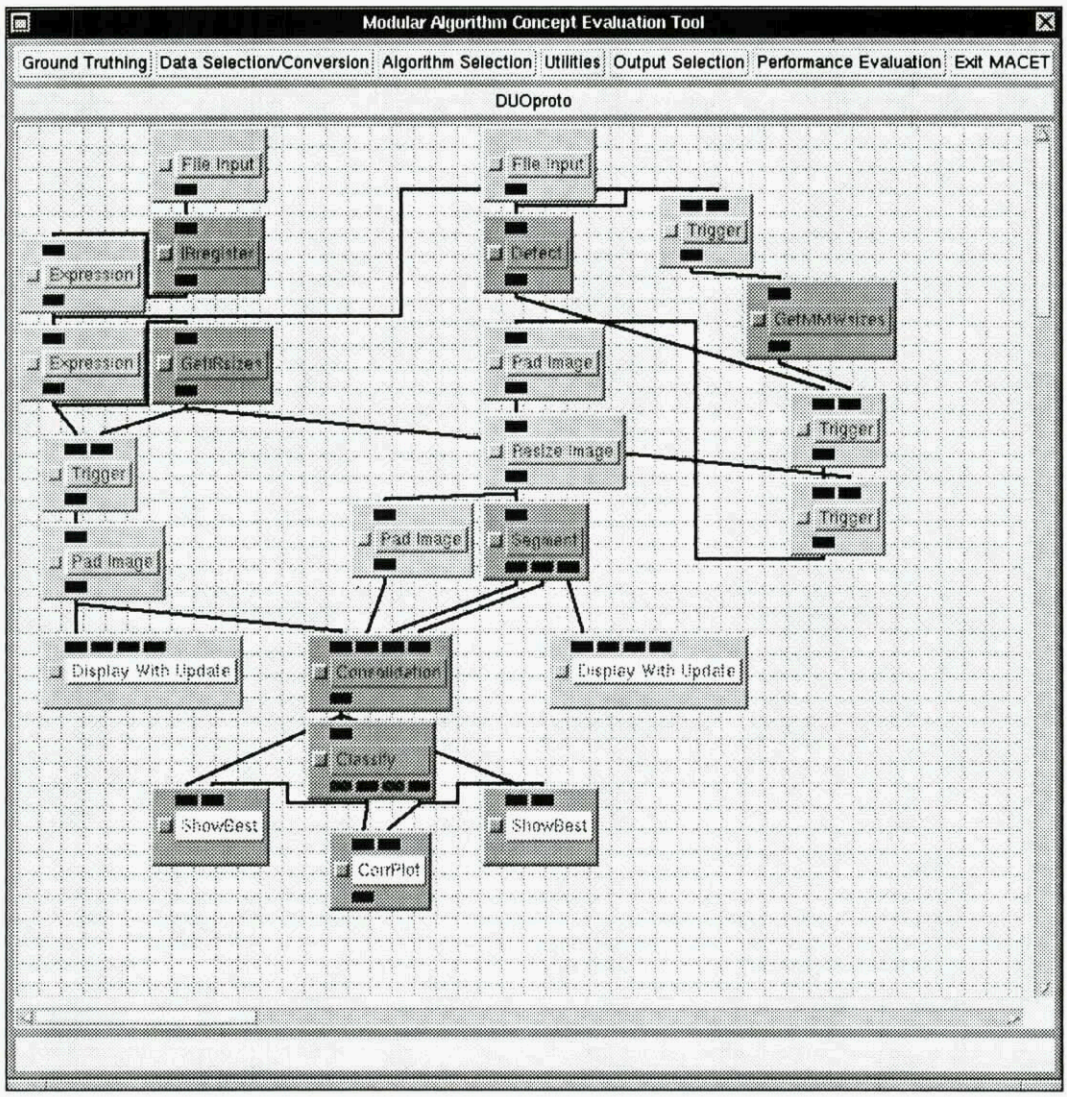


Figure 4.10 Dual Mode Prototype Algorithm

data is multiplied to the IR data in order to accomplish proper registration for processing. The geometric manipulation for registration is performed on the IR data in the "IRregister" module. The "GetIRsizes" and "GetMMWsizes" as well as the assorted image padding and resizing are used to maintain pixel to pixel correspondence between the two data formats as the data are processed. Triggering is necessary to ensure synchronization between the two channels.

The algorithm uses the MMW channel for detection of objects, and the IR channel for classification and aimpoint selection. The impetus behind this is the high contrast expected in the radar data between the large metallic targets and natural background clutter. Once the target objects are identified and a bounding region of interest defined, the higher resolution of the IR data is exploited for target identification and aimpoint determination.

Data processing in the MMW channel proceeds in the same manner as the single channel MMW algorithm up to the "Consolidation" module. This module performs the same function as the "Discrimination" module in the single channel algorithm, with an additional task of consolidating the IR regions of interest into a banded data set. It is the IR data which are passed to the "Classify" module for further processing. The "Classify" module is functionally the same as that in the

MMW single channel algorithm, and is shown expanded in Figure 4.11. The "Multi" glyph allows the FFT procedure to be performed on each band in the input multi-band data set. The resulting multi-band transform by the single-band template transforms of the candidate target types to form correlation images in the "multimatch" modules. The peak values are extracted in "multimatch" as well and passed on. Ascii file output is also available which summarizes other statistical information including the position of the maximum. This corresponds to the point in the region of interest which aligns with the center of the target template, and can be used for aimpoint determination.

The "ShowBest" modules display the extracted region of interest which scores highest for the corresponding target class. "CorrPlot" plots the correlation scores versus object number for each class. This plot is shown in Figure 4.12 for the example data set which contains two actual targets along with the regions of interest selected as "best" for the two candidate target classes. Crosshairs depict the point at which the template aligned with the regions for the best fit.

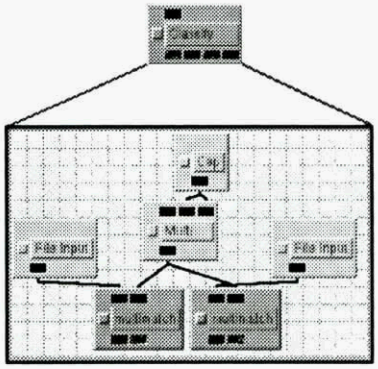


Figure 4.11 MMW Classify Module

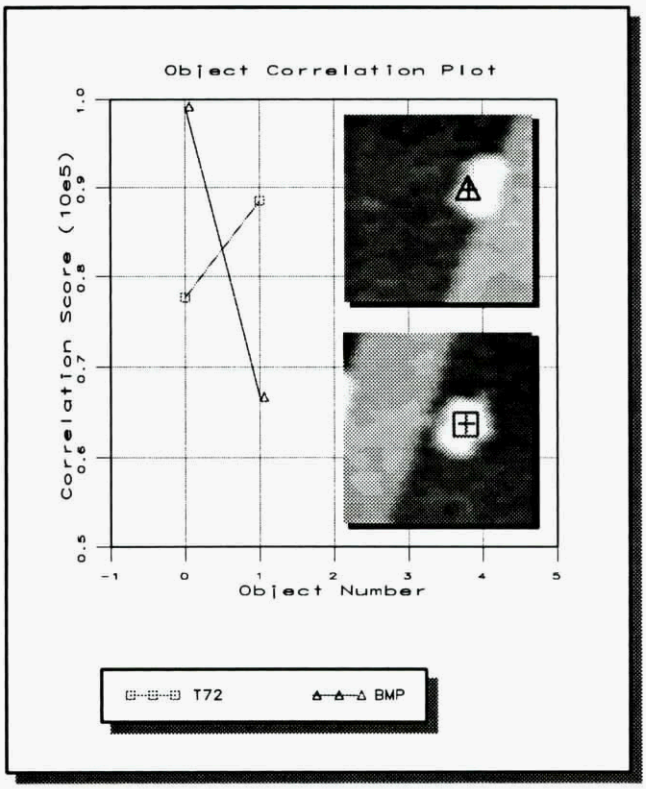


Figure 4.12 Target Correlation Plot and Aimpoint Identification

5.0 PERFORMANCE EVALUATION

Once the algorithm has been specified, and the test data set designated, the user can activate the MACET algorithm performance evaluation function. The purpose of this function is to compute certain defined algorithm metrics, and to facilitate the calculation of other user defined performance values. The metrics incorporated into the performance evaluation function are listed in Table 5.1. A description of each of these metrics is also provided in the table.

MACET uses the "Probdet" module to generate confusion matrices, which in turn, are used to calculate overall probabilities of detection and false alarm. These confusion matrices are used to correlate targets in truth data with detected locations in a target image file. Probdet may be used to accumulate decisions over a series of images. The process used by the Probdet module is shown in flow diagram form in figure 5.1.

The confusion matrices, generated by the Probdet module, are floating point images of dimension N+3 rows by N+2 columns, where N is the number of targets found in the truth data and image scenes. The values of the image in locations (1..N, 1..N) correlate target types from the truth file with detections in the target image file. Entry (1,1) contains the number of background pixels correctly identified.

Row 0, columns 2..N are the probability of detection (and correct classification) of the type associated with the corresponding row. Column 0, rows 2..N are the probability of false alarm (or incorrect classification) of the type associated with the corresponding column.

Entry (1,0) is the overall probability of detection inclusive of all targets ignoring classification. Similarly, entry (0,1) is the overall probability of false alarm. Entry (0,0) is the probability of target objects being detected at all, i.e. that a single pixel will be indicated as a target.

A final row is included whose entries are the actual numeric target types as found in the ground truth file, which are represented by their column of residence. A zero target type is used to indicate background.

Selecting the *Probability of Detection* or *Probability of False Alarm* modules for a selected algorithm and data set results in the iteration of that algorithm over the data while tabulating the algorithm's performance confusion matrix (true detection, false detection, missed detection, and missed false detection). These results are then sorted as a function of target-to-clutter ratio (generated in the ground truth editor) and read out to a file. This file can then be used to generate performance plots using the MACET *Data Display* module. All results will be for a single threshold setting.

Selection of the *ROC Curve* module will also result in iterative testing of the specified algorithm using the designated data set, but for multiple threshold settings. The step size and range of these variations are specified by the user upon selection of the *ROC Curve* option. The output of these tests are files containing Probability of Detection and Probability of False Alarm results as a function of threshold setting. An example ROC curve, using 20 threshold settings, is shown in figure 5.2.

Table 5.1. Available Algorithm Performance Metrics

METRIC	DESCRIPTION
ROC Curve	Probability of Detection versus Probability of False Alarm
Probability of Detection	Number of correct detection decisions divided by number of target trials.
Probability of False Alarm	Number of incorrect detection decisions (false alarms) divided by the number of background (non-target) trials.
Prob. of Correct Class	Number of correct classification decisions divided by the number of correct target trials.
Aimpoint Selection Error	Difference between computed aimpoint (centroid, hot spot, etc.) and the aimpoint designated by the user in the ground truth editor

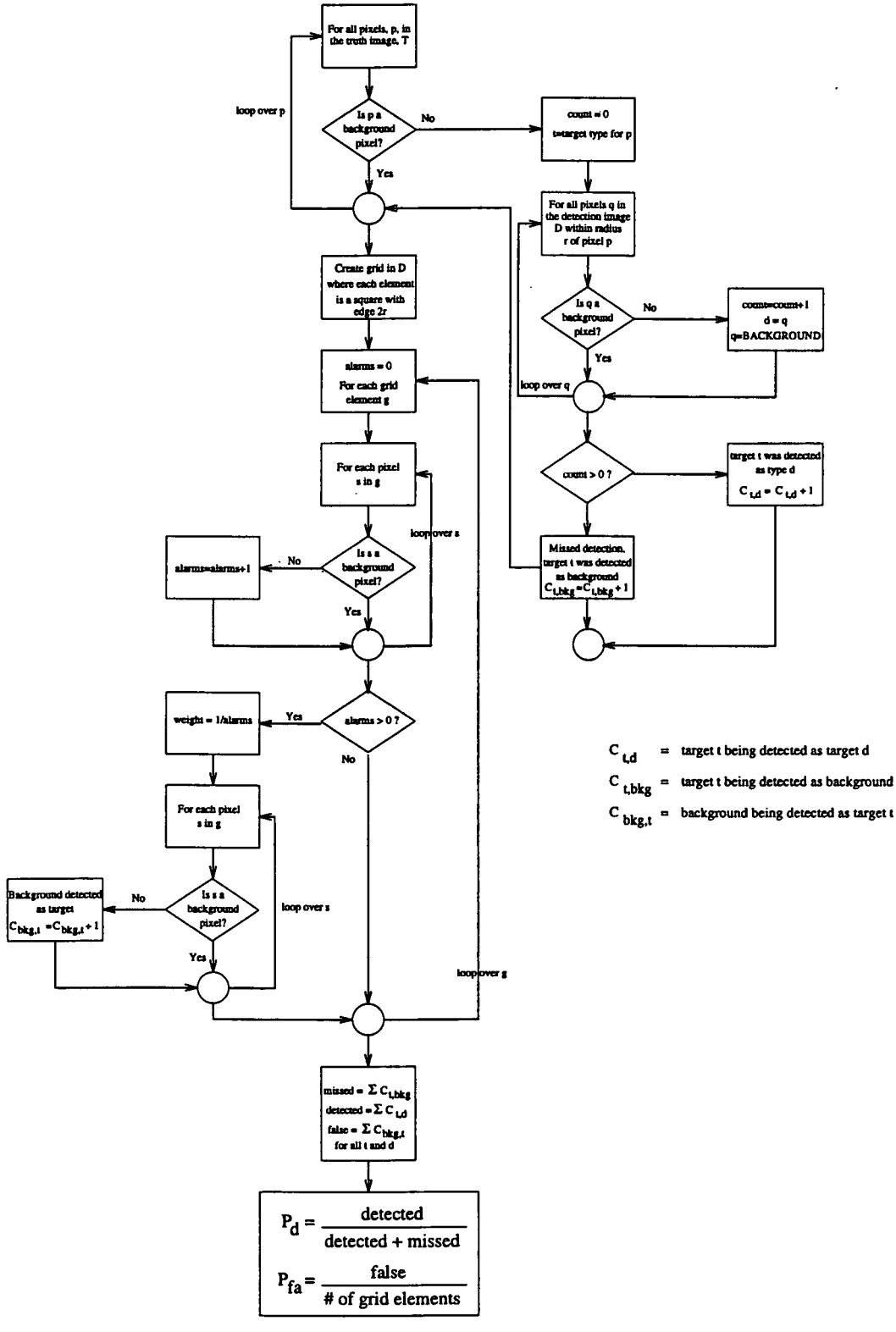


Figure 5.1 Flow Diagram of Confusion Matrix Generation

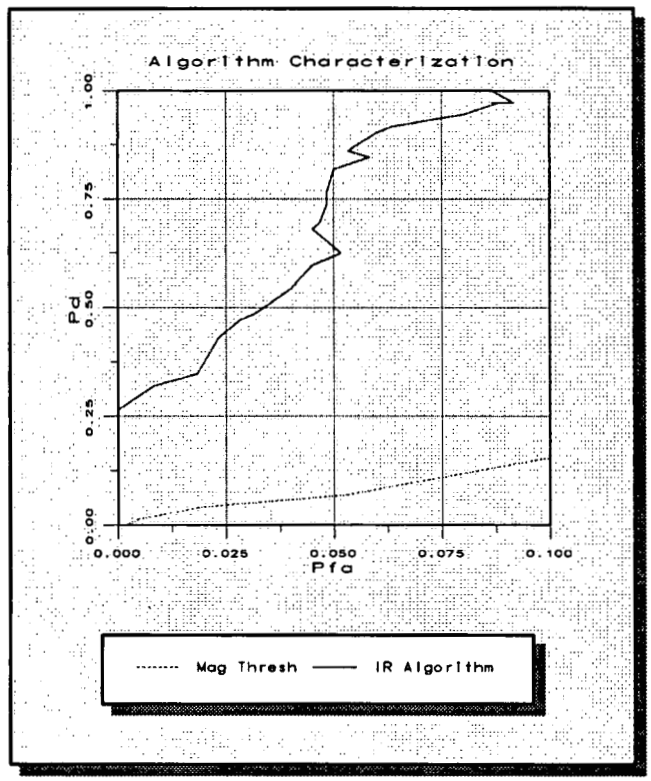


Figure 5.2 IR Prototype Algorithm ROC Curve

6.0 CONCLUSIONS AND RECOMMENDATIONS

MACET is a graphical interface driven system by which to rapidly prototype and evaluate acquisition and aimpoint selection algorithms for precision guided munitions. It is being developed as a component of WL/MN's multi-sensor algorithm testbed, and therefore it has been designed to accommodate multi-spectral sensor fusion systems.

MACET is being developed on the Sun Workstation, although it has been successfully rehosted to other workstations such as the DEC Alpha. It is built upon the Paragon Image Logic and Khoros 1.0 software systems. Because of the similarity of the two systems, upgrades of these products are being monitored in order to take advantage of any increase in functionality that they may offer.

The MACET architecture was developed to satisfy the following requirements: user friendly, minimum training time, rapid prototyping capability, compatibility with existing data, capability to accept existing algorithm components, algorithm evaluation shell, single and multi-sensor algorithms, emulation of target acquisition and aimpoint selection algorithms, Sun Workstation platform, nominal development costs, generation of standard algorithm performance metrics, flexible graphical output, and data probe placement. This architecture has been successfully implemented and all capabilities demonstrated for two incremental versions of the simulation: IR only (MACET 1.0) and IR/MMW (MACET 1.1). Continued MACET development will result in a four channel capable (ladar, passive MMW, IR, active MMW) version of the simulation. This version is scheduled for completion in November 1995. Current development efforts are directed toward MACET 1.2, which will introduce the ladar capability.

Each MACET build includes not only the enhanced environment functionality required to represent new sensor types, but also template algorithms. Such template algorithms provide the inexperienced algorithm evaluator with a baseline by which to initiate prototyping. They

also serve to validate the increased functionality of the MACET environment. Two such algorithm prototypes, an IR and fused IR/MMW, have been delivered with the MACET 1.0 and 1.1 builds [8,9].

Future versions of MACET will likely address the incorporation of legacy algorithms into the MACET algorithm library, execution speed enhancements, and extensions to unconventional classes of algorithms such as neural networks and fuzzy logic. These added capabilities are intended to establish MACET as an on-line repository of algorithm component research and to enhance its utility in providing the rapid turn-around required of captive flight test exercises.

Overall, MACET will give the Armament Directorate of Wright Laboratory a very powerful tool in the development and evaluation of algorithms for precision guided munitions. With its development, MACET will aid in the design of austere, all-weather seeker systems thereby enhancing the global reach of the U. S. Air Force.

REFERENCES

1. B. Sundstrom, J. S. Watson, "The Application of MMW/IR Sensor Fusion to Tactical U. S. Air Force Weapon Systems"; NATO Symposium on Multi-Sensors and Sensor Fusion; Brussels, Belgium, 8-11 November 1993.

2 J. S. Watson and D. S. Flynn, "IRMA Multi-Sensor Predictive Signature Model, "Symposium on Sensor Fusion, April 1992.

3. J. Watson, D. Flynn, M. Wellfare, M. Richards and L. Prestwood, "Irma Multisensor Predictive Signature Model," SPIE's International Symposium on Aerospace/Defense Sensing & Control and Dual-Use Photonics, April 1995.

4. J. S. Watson, B. D. Williams, K. D. Trott and N. A. D. Thompson, "MACET Sensor Fusion Algorithm Testbed," Symposium of Sensor Fusion, April 1993.

5. J. S. Watson, B. D. Williams, D. H. Harrison, "Modular Algorithm Concept Evaluation Tool (MACET) System Design Report", Nichols Research Corporation, Technical Report NRC-TR-92-131, Contract F08635-91-C-0110, 10 July 1992.

6 J. S. Watson and D. H. Harrison, "The Application of Model Based Vision to MMW/IR Sensor Fusion," Symposium on Sensor Fusion, April 1992.

7. M. Dunn, J. S. Watson, and L. Love, (Dual Model Seeker (DMS) Sensor Evaluation Program Captive Flight Data Reduction), Nichols Research Corporation, Technical Report NRC-TR-91-053, Contract F08635-91-C-0110, 24 June 1991.

8. K. Sullivan, S. Talele, J. Watson, B.Williams, "Modular Algorithm Concept Evaluation Tool (MACET) Version 1.0. System Description Document," Nichols Research Corporation, Technical Report NRC-TR-94-069, 30 June 1994.

9. K. Sullivan, S. Talele, J. Watson, B. Williams, and G. Woodall, "Modular Algorithm Concept Evaluation Tool (MACET) Version 1.1 System Description Document," Nichols Research Corporation, Technical Report NRC-TR-95-016, 31 January 1995.

STUDIES AND SIMULATIONS ON SENSOR FUSION AND CUEING FOR FIGHTER APPLICATION

by
Massimo Avalle

ALENIA AERONAUTICA
System Technology dept. T341
Corso Marche 41 - 10146 Turin, Italy

SUMMARY

A method to implement Sensor Fusion and Sensor Cueing on an advanced fighter aircraft is described in this paper.

Starting from a short introduction concerning the general aspects and theory of Sensor Fusion, the paper presents some choices adopted during the development of the Sensor Fusion process at ALENIA AERONAUTICA System Technology dept.

Sensor Cueing will be also introduced and some particular cases of interest for a fighter aircraft will be discussed.

The performances of the adopted solutions are then discussed on the basis of some experimental results obtained using a simulation tool.
An evaluation of the overall Sensor Fusion process performance and some considerations about possible alternatives will conclude the work.

LIST OF SYMBOLS

A/C	Aircraft
Az	Azimuth
DVD	Divisione Velivoli Difesa (Defence Aircraft Division)
El	Elevation
EW	Early Warning
FOR	Field Of Regard
FOV	Field Of View
IFF	Identification Friend or Foe
IRST	Infra Red Search and Track
RWR	Radar Warning Receiver
RWS	Range While Search
SF	Sensor Fusion
SM	Sensor Management
STT	Single Target Track
TWS	Track While Scan
VS	Velocity Search

1. INTRODUCTION

The studies on Sensor Fusion started in ALENIA AERONAUTICA in the context of a research program designed to study the capabilities of an advanced fighter A/C equipped with advanced functions like Sensor Fusion, Sensor Management, Situation Assessment and Attack Management operating in a high threat scenario.

During the first phase of the research, some items like accurate tracking, short response time and a high level of confidence in identifications were recognized as critical and so were used to tailor the algorithms for tracking, fusion and identification.

At the end of this phase the necessity to evaluate the just developed algorithms in a dynamic environment was recognized, in order to acquire more confidence in the phenomena and study the behaviour of the system in some simple operative situations.

For the above reasons the test of the entire system was executed using a simulator able to manage a simples manually piloted fighter A/C, equipped with the SF and SM algorithms to be verified, plus some autonomous targets.

During the SF studies and simulations, as well as the most complex case in which the System must be able to recognize as unique and track a single target detected by all the sensors, some specific and more frequent cases are also investigated. In particular, the cueing of a sensor to a target detected by another sensor was identified as an interesting problem in some operative situations.

The results of the studies and the simulations concerning the SF and the Sensor Cueing are showed in this paper.

2. SENSORS SIMULATION

Since the output data from sensors represents the main input for a SF process, then the more realistic are these inputs, the better will be the "reality" of the tests and the reliability of the results. The Sensors have been simulated taking into account the relevant characteristics of the commercial sensors, including the capability to be managed by both human operator (the Pilot) or an automatic function like the SM.

The modelling of the sensors may be split into two different blocks:

1. Antenna scan inside FOV.
2. Detection functions and measurement errors.

Both the detection functions and the errors modelled in this research program are the typical ones for each sensor and are not tailored to any existing sensor, with the exception of the RWR that has better accuracies than any existing sensor of this type.

The choice to model a RWR with high angular precisions was made to investigate the upgrading, in the overall Weapon System performances, by the introduction of another sensor able to contribute to the track of a target.

2.1 RADAR

In this simulation the Radar supports four modes: VS, RWS, TWS, STT.
In VS, it supplies only angular location and range rate.
In RWS, it provides location (also range) and range rate.
In TWS, the Radar executes track for a reduced number of targets supplying location, rates, accelerations and course.
In STT, it supplies the same parameters as TWS, more precise, but for only one target.

2.2 RADAR WARNING RECEIVER

In this simulation RWR measure the angular position of the electromagnetic sources existing in the scenario and tries to identify them by comparing the relevant received parameters with a stored emitters database.

2.3 INFRA RED SEARCH AND TRACK

In this simulation the IRST supports two modes: SEARCH and STT.
In SEARCH it is able to supply only Az and El.
In STT it supplies angles, angular rates, angular accelerations and a poor estimate of range.

2.4 EARLY WARNING AIRCRAFT

The EW supplies range, range rate and angles.

3. THE SENSOR FUSION PROCESS

As stated in [1][2][3] the SF is a process able to collect information from multiple sources, and to associate (correlate) them in order to achieve an estimation of the state and the identity of targets and threats.

The SF process can be seen as being composed to two different, sequential (in time) functions: Data Association and State Estimation.

The Data Association has the scope to determine which new sensor measurements and existing tracks have a common source; the State Estimation is performed after such assignment, in order to obtain the best estimate of the target/threat state.

Estimation of the targets/threats state are also used to predict the future state of each existing track at the time of the sensor measure.

Using a SF process it will be possible to provide the crew with the most accurate presentation of the threats and targets in the battle zone; the SF outputs, summarized in information concerning position and identification, are then transferred to a process able to assign a priority to each specific threat/target.

3.1 SENSOR FUSION DEVELOPMENT

The development of the SF process started from the general items explained above, and went on with tailoring such items to our specific application; the result was the process shown in Fig. 2.

The observations and/or the trackfiles from the four sensors are collected in a Tracks & Observations Receiver and distributed to the Association and State Estimation Functions.

The existing tracks will be extrapolated to the received track/observation time and then compared in the Gating and Correlation Blocks on the basis of their position and accuracy.

The correlated pairs are then filtered and combined before being used to upgrade the proper fused trackfile.

The main functions constituting the SF process are described in the following subsections.

3.2 ASSOCIATION

The aim of the Association Function is to search among the fused tracks Database looking for the best candidate (if it exists) to be updated by the new received sensors data.

The Association function is here split into two different and time subsequent functions: Gating and Correlation.

3.2.1 GATING

Gating is a technique to decide if a new track/observation received from a sensor can be associated to an existing fused track.

Such a decision is based on stochastic methods taking into account the detecting sensor and the fused track accuracies, in conjunction with the differences between the sensor track/observation and the extrapolation of the existing track.

The Gating Function then consists of constructing, for each existing track, a region (gate) in the measurement space within which the tracks/observations from sensors are expected to arrive; only those fused tracks receiving a sensor track/observation within their gate will be candidates for updating.

If a track/observation is applicable to more than one gate, the final decision about the best fused track candidate for updating will be taken by the Correlation function.

Track/observation not fitting into any gate will be considered as a candidate for new track initiation and controlled by the Track confirmation and deletion function on the basis of successive detections.

The gating implemented for the Association function will be rectangular and can be described by the following equations:

$$|Y_o - Y_p| < 3\sigma_r \quad (1)$$

$$\sigma_r = \sqrt{(\sigma_o^2 + \sigma_p^2)} \quad (2)$$

Y_o = observation

Y_p = prediction

3.2.2 CORRELATION

The Correlation Function has the purpose of making the final decision in the Data Association process by analyzing the Gating results and using them to find the best match between observations and fused tracks. Fused tracks will be then updated using the sensor data in the State Estimation Function.

The above sentence implies the computation of the normalized distance between the gate centre and the position of any track/observation satisfying the gate.

Assigned the value $\vec{y}$ to the new track/observation, the value $\vec{x}$ to an extrapolate trackfile (the centre of the gate), and defining $\vec{y}_n$ as their distance, the normalized distance d_{xy} between these two variables is defined by the following formulae:

$$d_{xy}^2 = \vec{y}_n^T S^{(-1)} \vec{y}_n \quad (3)$$

$$\vec{y}_n = \vec{y} - \vec{x} \quad (4)$$

S = covariance matrix of $\vec{y}_n$

The normalized distance d_{xy} has a chi-square distribution, deriving from linear operations on independent Gaussian random variables with zero means, thus it can be associated to the probability that a sensor track/observation could satisfy the gate.

The simplest case in which just one observation respects only one track gate will be simply solved by assigning the sensor track/observation to that fused track if the Equation (3) supplies an acceptable value.

If more than one sensor track/observation will satisfy the Equation (1), the best observation and track pairs are chosen using the maximum likelihood method. This method is able to decide among n fused tracks and m observations/tracks, which are the better candidate pairs for Association on the basis of their relative d_{xy} values.

3.3 STATE ESTIMATION

State Estimation should be a flexible process able to self-adapt to any sudden variation of inputs such as missing detections, asynchronous detections from sensors (targets are *not* detected at the same time by all the sensors) and target and/or fighter manoeuvres.

In order to implement such a process, the Kalman Filter was choosen due to its capability to automatically determine the gains to be given to the input measurements data in order to minimize the mean squared error on the basis of the target manoeuvre estimation and changing detection histories, such as missing detections.

The accuracy of estimations is also computed by the Kalman Filter via the covariance matrix; therefore, such a matrix can be used, in conjunction with the statistics of the measurement errors, for gating.

The Kalman filter used to implement the State Estimation is composed of three coupled range and angle (Az and El) filters.

The range filter uses range, range rate and range acceleration as states and its measurement matrix is composed of range and range rate measures.

The two angular filters work in polar coordinates and use as states the angles in conjunction with the velocity and acceleration components along the horizontal and vertical axes lying in the plane perpendicular to the line of sight toward the target (see Fig 1). The measurement matrix of such filters is composed of the angular only measurements.

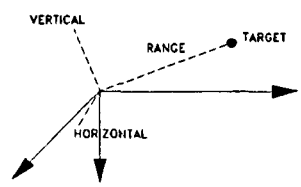


Fig 1 – Kalman Filter coordinate system

These three filters are coupled in the sense that the range filter needs the angular rates estimations from the angular filters, while the angular filters needs the range and range rate from the range filter to work properly.

Using this method, the SF process will not be able to execute passive tracking because angular filters need to be periodically refreshed with updated range and range rate estimates, therefore only targets detected by Radar can be fused.

This choice was made in order to avoid the addition of more computational charge for the computers, already high for the adoption of the Kalman Filters.

4. THE SENSOR CUEING PROCESS

Any Avionic System having an advanced function like the SF should include a Sensor Management process.

Such a system should be able to optimize the search coverage and scan rate for sensors, to guarantee the emission control and to facilitate the target identification by varying (managing) some sensor parameters like centre of scan, FOV, scan rate and mode.

SM is usually a very complex process composed of a variety of functions and usually needs to be accurately tailored to the specific application.

In Fig. 2 the SM process is shown as a functional block evidencing its capability to execute the Sensor Cueing (Ref [1]). Following are some considerations about this particular function, while the test results concerning some particular cases will be discussed in Section 5.

Sensor Cueing is a cooperative process between different sensors oriented towards the acquisition of data on a common target. This is achieved using observations or tracks from one sensor to point another sensor toward the target direction and by analyzing the second sensor detections in order to ensure that the target acquired is the same detected by the first. Such analysis should be, in its simplest case, a simple gate and correlation test.

In a particular application like a fighter A/C, Sensor Cueing could be useful in some situations like the following:

A. RADAR to IRST

In some situations it could be useful to track a target while reducing active emissions toward it but still maintaining a good level of accuracy.
Such goals can be achieved by combining the high angular accuracies of a passive sensor like IRST with sporadic Radar emissions, designed to acquire range and range rate information in order to maintain the SF tracking capability.
Obviously, working with aged range informations results in suboptimal performance for the State Estimation Function, but at medium range, the track accuracy maintains good performance.

B. IRST to EW A/C

EW A/C can acquire targets at long range and guide fighters toward them. Later, the fighters can autonomously acquire such target using their on-board sensors.
If active emissions must be avoided as much as possible, but at the same time a medium range accurate bearing is necessary to improve the fighter intercept path, the IRST can be used in the same way as described at point A until a more accurate track must be acquired using the on board Radar.

C. RADAR to RWR

A threat detected by RWR, may need more data in order to obtain more accurate information about position and identity. Such information can be obtained by cueing an active sensor like Radar toward the emissions detected by the RWR and using the Radar features to acquire more cinematic data and improve identification.

The tests relative to each of the above situations will be discussed in Sections 5.2 to 5.4.

5. SIMULATION RESULTS

Tests have been executed in order to evaluate, via simulation, the performances of the SF process implemented.

Simple fighter-target geometries in which the fighter follows an intercept path toward the target have been used to explore the possibility of tracking a target using all the on board sensors data and the influence of some relevant parameters on the target acquisition and the track precision.

The tracking system developed for the SF process uses, as described in Section 3, coupled range and angular filters. The range and range rate measurement can then be considered as relevant parameters for the system.

Consequently, the interval between two consecutive range and range rate measurements (i.e. the revisit time of Radar for a specific target) can influence the track accuracy.

Another relevant parameter to take into account when associating data from different sources is data precision. In fact inaccurate data can negatively influence Data Association and State Estimation.

In the following Subsections, the influence of the above parameters on the fused trackfiles will be shown using simulation results.

Only the relative Az data referred to the true scenario Az value (the true target Az) will be shown, because the simulation results for the El are similar to those for Az. Range and range rate are not significant to evaluate the SF process performances because such parameters are measured only by Radar.

Figures concerning the Cueing tests show the Radar detection times (not the detected values) and the outputs of both the filtering and extrapolation functions in order to better evaluate track divergence.

5.1 MULTISENSOR FUSION

The more relevant results of the simulation tests executed on the Multisensor Fusion system are shown in Fig 3 and Fig 4.

During these tests, the Correlation accuracy evaluated using the chi-square statistics applied to d_{xy} was always better than 50%, confirming the observation to gate assignment with a high confidence level.

Fig 3 shows the values detected by each sensor associated with the corresponding fused trackfile obtained from processing the input data.

Such figures can be used to demonstrate the poor influence of an inaccurate sensor like RWR (when compared with Radar and IRST) on the fused trackfiles, when the time interval between two consecutive RWR detections is short (less than four seconds in the example), then, with a RWR accuracy like the one considered here, it could be used to participate at the fused trackfile updating.

In Fig 3 it can also be observed that the fused trackfile has generally a better precision than the single sensor measure and is more stable.

Fig 4 is used to show the influence of the Radar scan interval (i.e. the target revisit time) on the fused trackfile accuracy.

In Fig 4-A, with the Radar scan interval of 2 seconds there are no appreciable divergences in the fused trackfile, nor when missed Radar detection occurs.

In Fig 4-B, enlarging the Radar scan interval to 6 seconds, considerable divergences can be noted, but the fused trackfile maintains a good accuracy, and the Association accuracy will always be better than 50%.

Enlarging the Radar scan interval to higher values, the RWR inaccuracy becomes significant and a loss of track could occur.

5.2 RADAR TO IRST CUEING

The results of the simulation tests concerning the Radar to IRST cueing are showed in Fig 5.

Tracks were initiated on the first Radar detection after the Cueing occurs, then the Cueing precision is evaluated by the system as a function of the Association accuracy.

Different Radar scan intervals of 6 and 10 seconds are considered in order to examine the influence of such parameters on the track stability.

In Fig 5-A the track maintains an acceptable stability and accuracy up to 8 Km of range, then instability becomes evident and a track loss occurs.
Enlarging the Radar scan interval it can be noted (Fig 5-B) that track divergence starts at a range of 14 Km.

This test can be useful to evaluate the performance of the Radar to IRST Cueing versus the Radar scan interval and range from target. Reduction of active emissions is feasible at medium ranges, but as the range decreases, the Radar scan interval has to be reduced in order to maintain the track stability.

5.3 IRST TO EW A/C CUEING

Fig 6 depicts the results of the simulation tests of the IRST to EW cueing at different IRST scan intervals.

Tracks were initiated at the first IRST detection after the cueing occurs and the Cueing precision is evaluated using the Association accuracy.

During this test the Az and El data from the EW A/C were used to set the IRST centre of scan while angle tracking was executed using Az and El data from IRST because the EW data are aged due to data link delays and usually not useful to update a trackfile.

EW A/C measurements of range and range rate are transposed to the fighter axis before supplying the angular filters with the new values.

Fig 6-A shows bad track accuracy due to the relevant differences between the filtered and extrapolated values. Reducing the IRST scan interval (Fig 6-B), the track accuracy improves and tracks can be maintained up to where it becomes necessary to use the onboard Radar to augment the track precision.

5.4 RADAR TO RWR CUEING

In this test, depicted in Fig 7, the detection of an air to air threat by the RWR was simulated and, consequently, the SM process decided to cue the Radar toward the threat direction in order to acquire more cinematic and identification data.

After receiving the RWR alert, the fighter starts a 2g sustained turn in order to position the threat to the front and facilitate Radar operations. The influence of such turns on the fused trackfile are evident in the first 15 sec of the test and are due to two main causes:

1. Filters are not yet in the steady state
2. RWR data are used to update the fused trackfiles in order to ensure the Cueing correctness.

In any case, the Association function was able to maintain the track although such angular errors. If necessary, after the Radar Cueing, the SM process can decide to update the trackfile using only the Radar data, or to change the Radar mode.

6. CONCLUSIONS

The performances of the SF process described in Section 3 and the feasibility of the Sensor Cueing in some particular situations have been shown by the simulation tests conducted.

It was also indicated that a tracking system using coupled range and angular filters could be useful in a SF process, but will have the constraints of quite short range and range rate measurements intervals, expecially when range decreases.

To avoid such problems a passive ranging function, able to track the range between two consecutive Radar detections or the passive angular track capability should be added, but the computational charge for computers will rapidly increase. In this case the solutions presented in this paper will have to be considered as a compromise between the necessity to have a State Estimation process able to self-adapt to varying scenarios and the constraint of avoiding heavy computational charges for the computers.

7. ACKNOWLEDGMENTS

I wish to thank my colleague, G. Barbera of ALENIA AERONAUTICA, System Technology dept. for his valuable work in developing the mathematical models and the Software code for the simulated fighter and targets used during the simulation tests, without which the simulation of the SF process and its evaluation would be less accurate.

8. REFERENCES

[1] Waltz-Llinas "Multisensor Data Fusion" - Artech House

[2] Y. Bar-Shalom "Multitarget - Multisensor tracking: advanced applications" - Artech House

[3] S. Blackman "Multiple target tracking with Radar applications" - Artech House

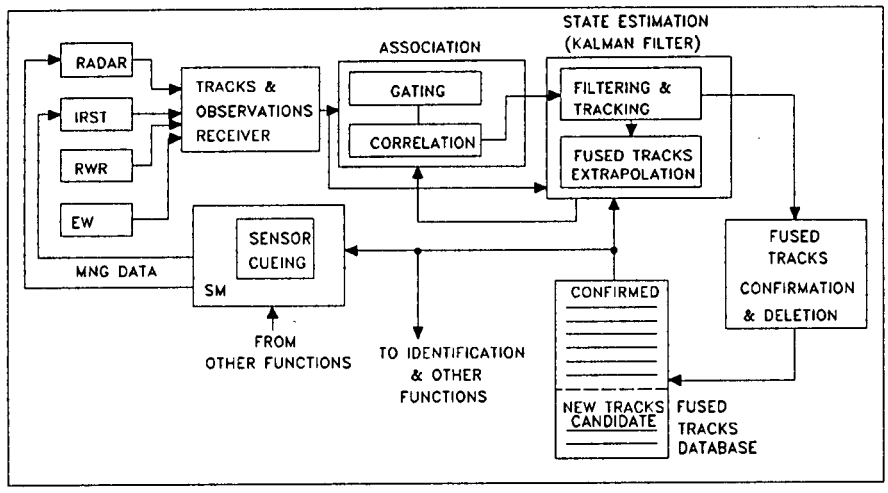


Fig 2 – The Sensor Fusion process

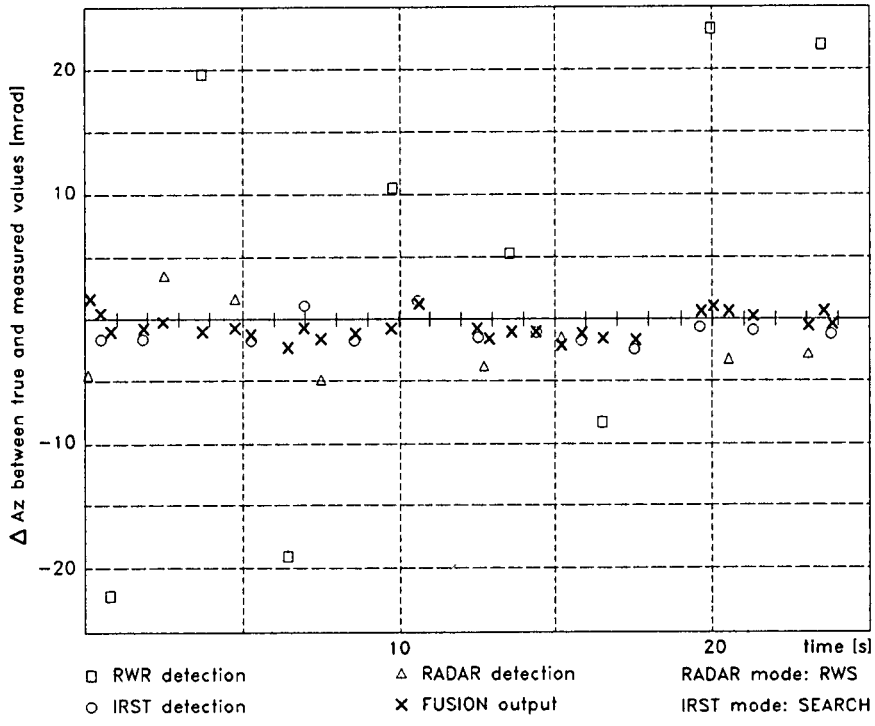


Fig 3 – RADAR IRST RWR Correlation: trackfile output and Sensor measures

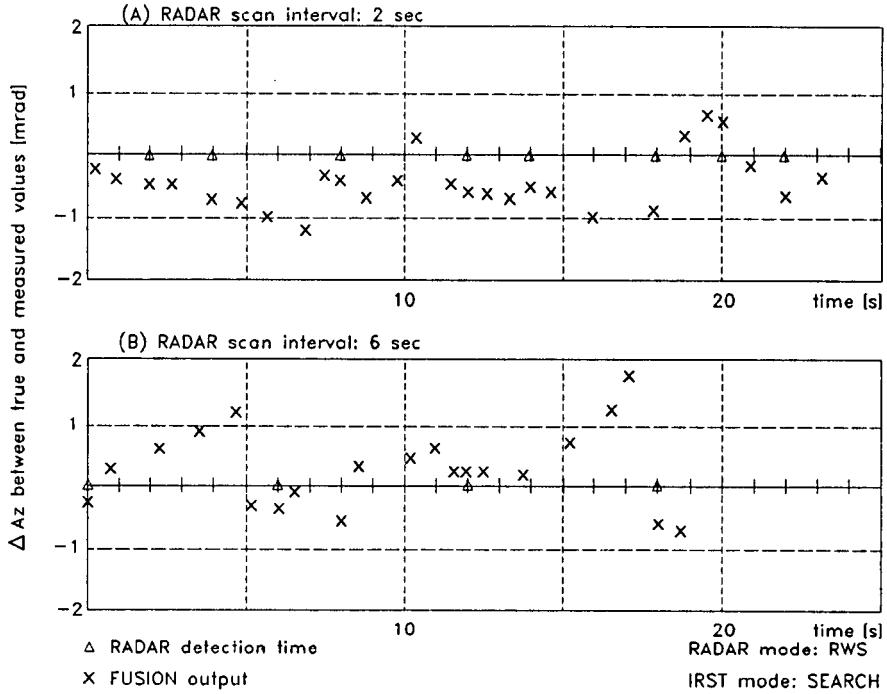


Fig 4 – RADAR IRST RWR Correlation: influence of RADAR scan interval

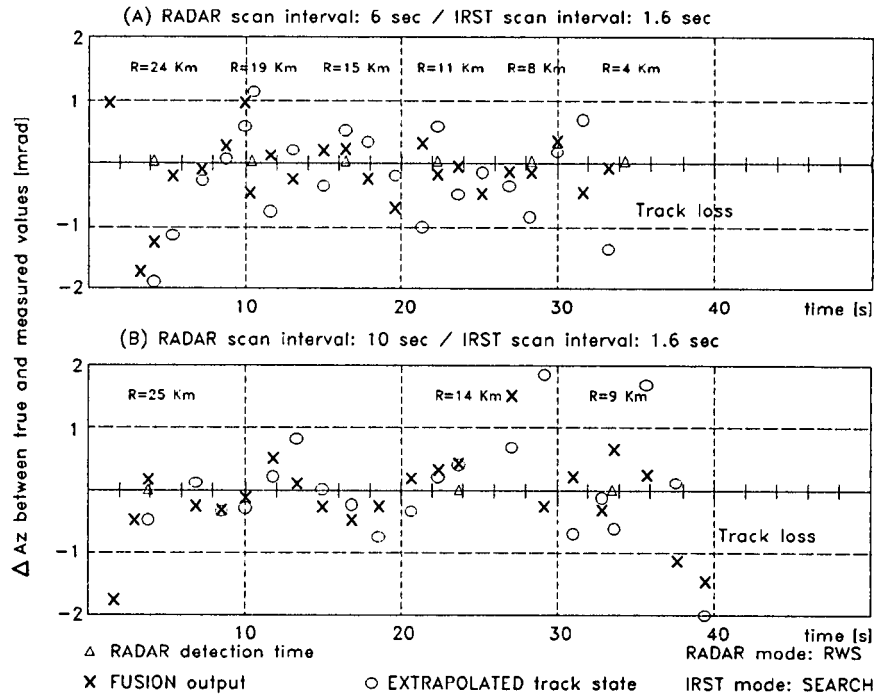


Fig 5 – RADAR to IRST Cueing

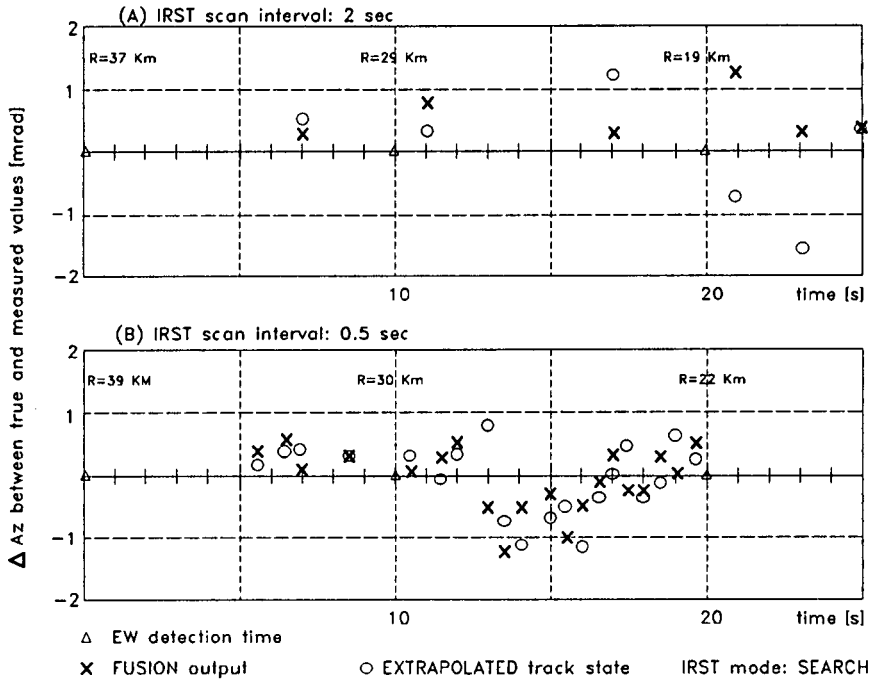


Fig 6 – IRST to EW Cueing

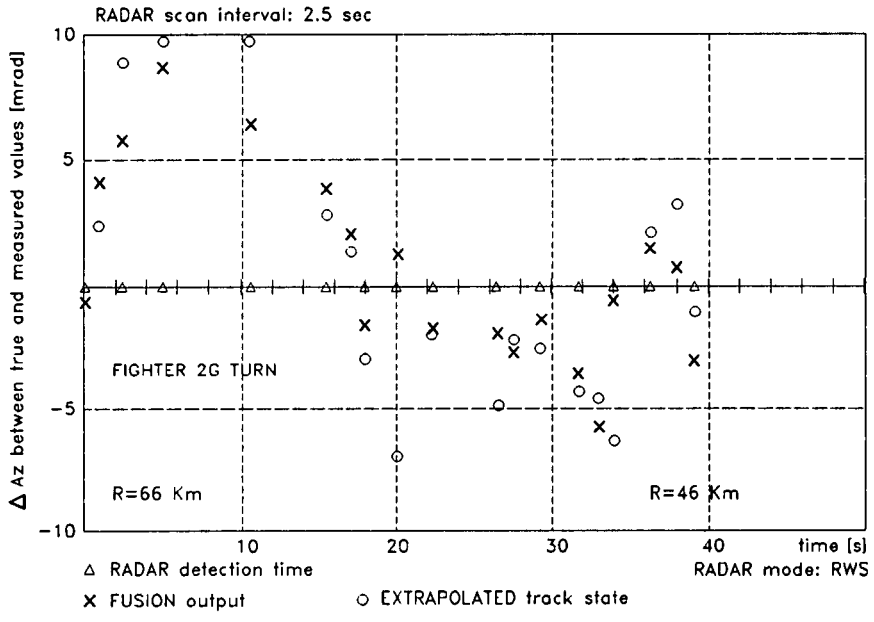


Fig 7 – RADAR to RWR Cueing

SECTION V

DATA FUSION FOR GUIDANCE AND CONTROL APPLICATIONS

INTRODUCTION

by

Dr. David F. Liang
Head, Space Systems and Technology
Defence Research Establishment Ottawa
Department of National Defence
Shirley Bay, Ottawa
Canada K1A OZ4.

The effectiveness of military operations is driven by increasingly sophisticated technology applications. The advent of more and more powerful processor and sophisticated sensors implies that the military has to cope with a powerful new generation of high speed, stealthy and accurate weapon systems. Particularly distressing is the fact that during the Gulf War, coalition air forces flew more than 5000 sorties over Iraq without a single confirmed kill of a mobile SCUD. To cope with increasing sophistication of threat and sensor systems, multi-sensor, multi-target tracking systems could be ubiquitously applied to smart weapons, target acquisition and tracking systems, battlefield surveillance and situation assessment systems, wide area surveillance systems, tactical and strategic defence systems.

The paper by S. S. Lim and D. F. Liang presents a Multiple Hypothesis Tracking algorithm that has been modified for efficient implementation in a multi-target air defence radar surveillance tracking assessment. The modified algorithm has been extensively tested against numerous sets of air defence radar measurements.

The paper by R. R. Suresh presents the *raisons d'être* and advanced testbed for using multi-mode sensor suites to support low cost applications in battlefield surveillance and weapon guidance. He also presented an Integrated Multiresolution Algorithm (IMA) paradigm, which in a single, seamless approach, provides detection, recognition, and identification as a by-product at successive levels of refinement.

C.A. Noonan dealt with sensor data fusion in the context of air-to-air situation awareness beyond visual range. He was able to show that for a future air superiority aircraft, equipped with non-commensurate sensors, algorithmic complexity in the fusion of the data has a lesser impact on system performance than the determination of the data relationships. For the communications network and tactical data exchange this means getting the data alignment right.

R. G. Zuidgeest presented the merits of artificial intelligence in command and control multi-sensor data fusion (MSDF). He presented the world model that includes MSDF and the four levels of hierarchical representations of knowledge sources: sensor, object, recognition, and relational level. He discussed a number of candidate AI techniques for the representation of the knowledge sources, and described a global distributed architecture for C2 networks where AI techniques can be applied.

AIR DEFENCE RADAR SURVEILLANCE SYSTEM TRACKING ASSESSMENT

S. S. Lim
112 Grassy Plains Drive
Kanata, Ontario
Canada
K2M 2M5

D. F. Liang and M. Blanchette
Defence Research Establishment Ottawa
Department of National Defence
Ottawa, Ontario
K1A 0Z4

SUMMARY

In this paper, an efficient Multiple Hypothesis Tracking (MHT) algorithm has been developed and implemented. This algorithm is a modified version of MHT combined with an N-scan back pruning approach. The principal objective of this modification is to effectively reduce the large number of hypotheses in the original MHT thereby yielding a fast tracking algorithm which requires moderate computing resources. The modified MHT algorithm has been extensively tested against numerous sets of real radar measurements. These real radar data contain complex tracking scenarios such as trajectories of several fighter aircraft going through high-*g* maneuvers, crossing tracks and close formation in a cluttered environment. This paper will demonstrate the performance of the modified MHT algorithm against a set of real radar data. Descriptions of the radar data and the extent of the clutter are provided. The effectiveness of the modified MHT algorithm in handling maneuvering targets will also be discussed.

1. INTRODUCTION

Future air defence surveillance systems must be able to cope with highly maneuverable and closely spaced targets. The detection and tracking of dim targets in high-*g* and dense environments will require low detection thresholds with resulting high clutter densities. Most target trackers aiming at radar applications have been tested and evaluated based on the measurement data produced from simulation subject to a priori mathematical model and simplifying assumptions. However, in real-time radar target tracking, the measurements are not normally governed by any assumed mathematical models and hence inevitably cause unforeseen difficulties.

In current literature, a number of well-recognized approaches have been proposed [1-4]. The nearest neighbor standard filter (NNSF) is the most straight forward, and produces a single unambiguous data association solution at each point of time, based on the previous association and the current sensor information. This however may not be the best choice, especially because it does not make full use of all prior sensor data. Once an incorrect association is made, it seems unlikely that the solution would ever recover.

The branching algorithm [5,6] is an alternative to the NNSF approach. In this algorithm, the correlation performance can be further improved by deferring some difficult assignment decisions until more data are collected. Correlation hypotheses are created and the unlikely branches are eliminated by computing the relative likelihood of each branch, under the assumption that each target is present (i.e., the detection probability is one) and without accounting for false alarm statistics. This is an improvement over simple sequential decision making such as in the NNSF. However, the main criticism of the branching algorithm arises from the fact that the association constraint (that an observation cannot simultaneously belong to different target tracks) is not used. This implies that a target can be associated with every measurement within its gate. Hence, measurements within several gates can lead to sets of data-association hypotheses that are not mutually exclusive. Smith and Buechler [6] have partially remedied this problem by the use of an ad hoc procedure to eliminate branches whose estimates are less than a special distance away.

Bar-Shalom proposed a track-oriented Joint Probabilistic Data Association (JPDA) filter [7-8]. The JPDA computes the probability of

association of the latest set of measurements to the existing targets. The joint probabilities are computed under some assumptions of false target distributions and all the hypotheses are combined into one in every scan. This method performs an averaging over observation-to-track data association hypothesis that have roughly comparable likelihood. While this algorithm maintains continuity superior to NNSF, this is done at the expense of accuracy, since all reports are used in track update computation even though only one report has really originated from the target. More over, this algorithm does not provide the track initiation procedure.

Reid [9] proposed a tracking algorithm, known as the multiple hypothesis tracking (MHT). The MHT approach will maintain several (perhaps many) possible data association solutions, and uses the history of sensor data to eliminate highly unlikely choices, eventually leaving only one best choice (hopefully). This should yield the best solution, however it does generally have periods of uncertainty. The main drawback is that in a dense target environment the number of hypotheses can increase exponentially with each scan, leading to severe computational burden. Alternatively, pruning the hypotheses becomes essential. Another disadvantage is that the data association decision is often deferred, and thus a single best estimate is not always available in a timely manner. Thus if one could not wait for the solution, then it would be necessary to take special measures. A sub-optimal track based variance of the MHT is the Track Splitting Filter (TSF). This filter deals with the problem of multiple measurements falling within a gate in a manner similar to the MHT, but it does not address the issue of multiple tracks competing for the same measurement. It is well suited for tracking known targets in closed formation.

Hence the principal consideration for the choice of a tracking algorithm for implementation in a real radar environment (rather than simulations), depends on the nature of the radar data and on the computational requirement for timely operation. In this paper a modified MHT (MMHT) algorithm has been developed for real time implementation and for the assessment of the performance against real radar data. The modified algorithm promises the feasibility of real-time tracking while requiring a moderate computer memory. The MMHT is a simplified version of the MHT but employs two modified hypotheses pruning techniques: *dynamic thresholding* and *N-scan back pruning*. The pruning schemes effectively delete unlikely hypotheses and maintain only a reasonably small number of hypotheses, thereby making the MMHT a real-time implementable algorithm. The performance of the MMHT is demonstrated against real radar data that contain complex track scenarios such as trajectories of several fighter aircraft going through high-g maneuvers, crossing tracks and close formation in a cluttered environment. The results of performance tests indicate that the algorithm is fast enough for real time implementations and capable of handling a large amount of track information. Further, it is robust in several distinct tracking situations.

The paper is organized as follows. In Section 2 the modified MHT algorithm is presented. To demonstrate the performance of the MMHT algorithm, numerical results against various radar data are displayed in Section 3. Concluding remarks are presented in Section 4.

2. THE MODIFIED MHT ALGORITHM

In this Section, a modified MHT implementation is presented. As mentioned above, the main drawbacks of the MHT implementation are the requirement for enormous computation time and huge computer memory. This is because the number of track hypotheses grow rapidly with time. Hence, to make the MHT scheme implementable for real time applications, modifications to provide effective management of the track hypotheses are essential. In this Section, the modifications are referred to as the modified MHT (MMHT) algorithm.

2.1 The Modified MHT Algorithm.

The practicality of the MHT is determined by the degree to which the unlikely hypotheses are eliminated, thereby keeping only a manageable number of tracks with high probabilities. For this sake, the MMHT algorithm employs two pruning techniques, i.e., dynamic thresholding and N-scan back

approaches, with appropriate modifications. The dynamic thresholding is an improved version of the thresholding scheme by dynamically updating the probability threshold for pruning. The N-scan back pruning is to select, every N scan, only the most likely hypothesis from each cluster and delete all other hypotheses within the cluster. This pruning results in significant improvement in both computation speed and memory requirement for the track management. The advantage of the pruning process is to reduce efficiently the unlikely hypotheses and to maintain only a reasonable number of hypotheses. Thus it prevents the number of track hypotheses from growing so rapidly that they can expand out of control. Further, the pruning scheme simplifies the task by only showing the user the track scenarios of interest. The modified MHT scheme is graphically presented in the flow diagram shown in Fig. 1.

The main part of the algorithm consists of clustering, hypothesis generation, pruning by dynamic thresholding and N scan back approach. The cluster (CLUST) subroutine associates measurements with the previous clusters. A cluster is a group of hypotheses containing associated tracks that do not interact with any other group of hypotheses within other clusters. The hypotheses within a cluster will not share measurements with the hypotheses of any other clusters. The basic goal of clustering is to divide the large tracking problem into a number of smaller ones that can be solved independently. The hypothesis generation (HGEN) subroutine creates new data association hypotheses for the set of validated measurements of each cluster. The probability of individual hypothesis is then computed in the PROB subroutine. Both the clustering and hypothesis generation procedures use the pruning subroutine to reduce the number of hypotheses which grow rapidly with time. The hypotheses satisfying certain qualifications will remain and all the unlikely hypotheses are eliminated in the PRUN1 subroutine. As part of the pruning process, similar hypotheses may be combined into one. The subroutine FILT computes the estimate of each hypothetical track based on the previous estimate and on a new measurement using a standard Kalman filter. Every N scan, only the most likely hypotheses are selected and the other hypotheses are all eliminated in the subsequent considerations, as explained above. This procedure is shown in the subroutine PRUN2. The major distinction between the MMHT and Reid's MHT is with the two pruning schemes, i.e., dynamic thresholding and N-scan back pruning (the subroutines PRUN1 and PRUN2 in Fig. 1).

2.2 Clustering

Clustering is a procedure for partitioning a large group of association hypotheses into several non-interactive sub-groups, each having a smaller number of hypotheses. The principal objective of clustering is to reduce a dimensional tracking problem into several smaller ones that can be solved independently. A byproduct of the clustering procedure is its suitability for efficient parallel processing of the tracking tasks.

Initially, one cluster is set up for each confirmed track. Each new measurement is associated with a cluster if it falls within the validated region of any track from that cluster. A new cluster is initiated any time a measurement is received, which does not fall within the gates of any track contained in an existing cluster. The cluster is initiated with the measurement using the alternatives (true or false alarm) associated with its source. In order that clusters remain distinct, the gates of tracks between the clusters must not be overlapping. Thus, when a measurement falls within the gates of two or more tracks from different clusters, the clusters are merged. The merged cluster is called a super cluster. If tracks within a cluster separate spatially and no longer have overlap in measurements, that cluster is subdivided accordingly into smaller clusters that can be managed independently.

2.3. Generation of The Association Hypotheses.

Let $L(k)=\{L_i(k), i=1,2,...,k\}$ be the set of association hypotheses up to time k. This set is obtained from $L(k-1)$ and the latest set of measurements $Y(k)=\{y_1(k), y_2(k),...\}$ as follows. New hypotheses are formed by associating the first measurement $y_1(k)$ with $L(k-1)$, then augmenting the resulting set by associating

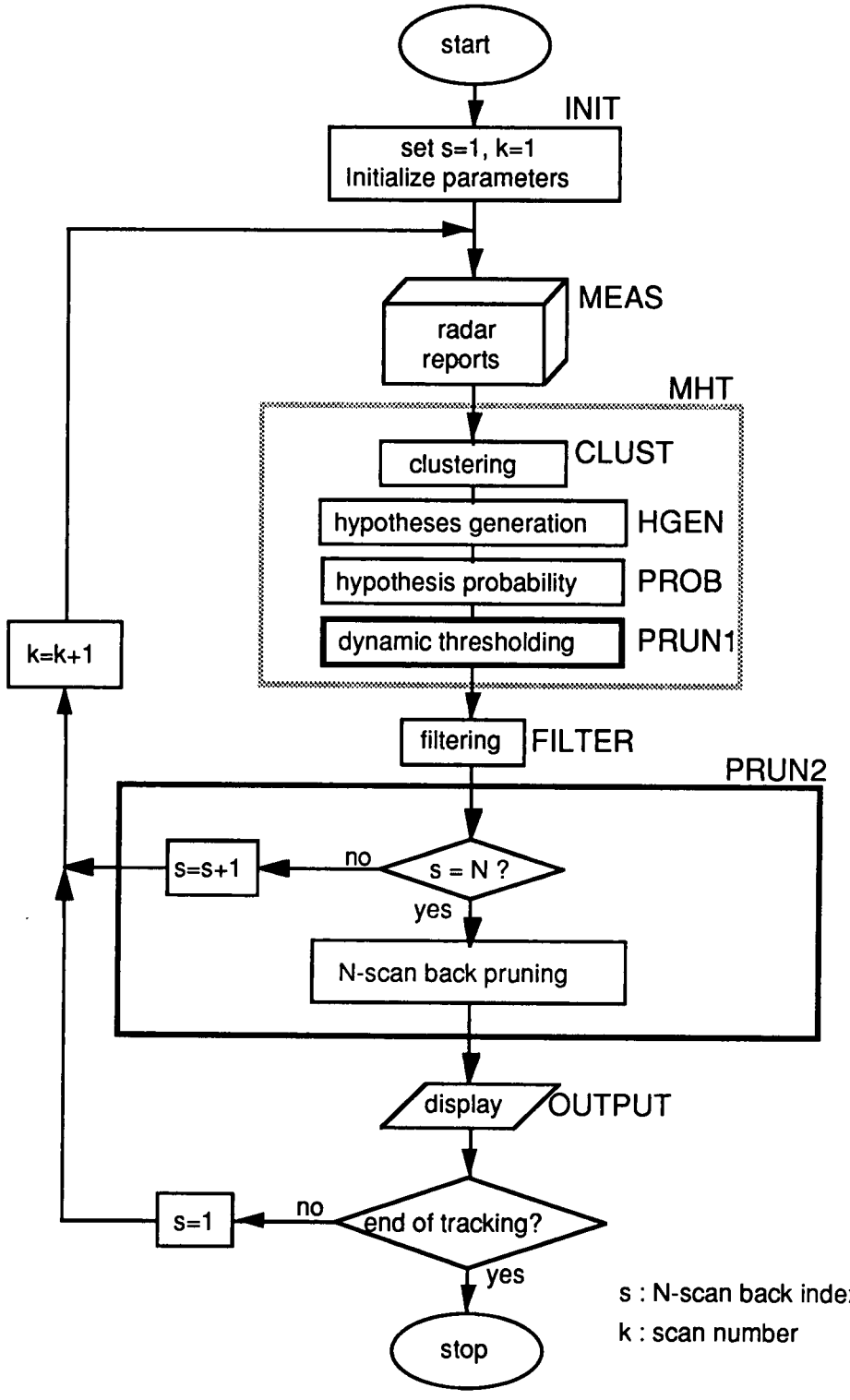


Fig. 1 The Modified MHT(MMHT) Algorithm.

$y_2(k)$, etc. The possible associations for the i-th measurement $y_i(k)$ are:

i) the continuation of a previous track;
ii) a new track;
or iii) a false alarm (or clutter).

However, the above assignments must be made subject to the constraint that each track can be associated with at most one current validated measurement which falls within the validated region of the track.

2.4 Computation of the Hypothesis Probability.

The sensor reports usually contain measurement noise generated from various sources, such as thermal false alarms, clutter and other targets. Suppose the number of such extraneous reports in any volume C obeys a Poisson distribution with mean bC, where b is the normalized extraneous report density. Let N_{DT}, N_{TGT}, N_{NT} and N_{FT} denote the number of measurements associated with the prior targets, previously established targets within the area of coverage of the sensor, new targets, and false targets, respectively. Then the probability of the hypothesis $L_i(k)$ given measurement $Y^k = \{Y(1), Y(2), \ldots, Y(k)\}$ is calculated[9] by

$$P^i(k) \equiv P(L_i(k)|Y^k)$$
$$= \frac{1}{C} P_D^{N_{DT}} (1-P_D)^{(N_{TGT}-N_{DT})} \beta_{FT}^{N_{FT}} \beta_{NT}^{N_{NT}}$$
$$\left[\prod_{m=1}^{N_{DT}} N(y_m - H\hat{x}, B)\right] P^l(k-1),$$

where $P_D, \beta_{FT}, \beta_{NT}$ are the probability of detection, the density of false targets, and the density of new targets, respectively. The c is a normalization constant, H is the measurement matrix and $P^l(k-1)$ is the probability of the hypothesis $L_l(k-1)$. $N(x,B)$ denotes the normal distribution given by $\exp\left(-0.5x' B^{-1} x / \sqrt{(2\pi)^n |B|}\right)$, $B = H\bar{P}H' + R$ where $\bar{P}$ is the covariance of the target estimate for the prior hypothesis $L_l(k-1)$ and R is the measurement noise covariance.

The probability can be recursively calculated in the following way. First, all the probabilities of the prior hypotheses are multiplied by $(1-P_D)^{N_{TGT}}$. Then the probabilities of new hypotheses are updated by multiplying either

1) β_{FT}, if the measurement is associated with a false target;
2) β_{NT}, if the measurement is assigned to a new target;
or 3) $P_D / (1-P_D) N(\tilde{y}(k), B_i)$ if the measurement y_j belongs to the i-th track, where $\tilde{y}(k) = y_j(k) - H(k)\hat{x}_i(k|k-1)$.

Finally the hypothesis probabilities are normalized.

2.5 Pruning Unlikely Hypotheses.

A track history $L(k)$ at scan k is defined by selecting, at each scan $j \le k$, a single sensor report $y_{i_j}(j), 0 \le i_j \le N_j$, where $i_j = 0$ refers to the hypothesis that none of the sensor reports within the validation gate is originated from the target. Hence the track history $L(k)$ is just the hypothesis that the entire sequence of measurements within $L(k)$ is correct, i.e., each sensor report $y_{i_j}(j)$ was originated from the target when $i_j \ne 0$, while no sensor report was received when $i_j = 0, 1 \le j \le k$. The track history $L(k)$ at scan k is obtained from the track history $L(k-1)$ at scan $k-1$, and incorporating it into the measurement set specified by $Y(k) = \{y_i(k), i = 1, 2, \ldots, N_k\}$. In notional terms, $L(k) = \{L(k-1), \quad y_i, i = 0, 1, 2, \ldots, N_k\}$. Clearly one history $L_l(k-1)$ at $k-1$ gives rise to $(1+N_k)$ histories $L_l(k)$ at scan k. Then, the total number of such hypotheses $L(k)$ at scan k is given by $(1+N_k)M(k-1)$, where $M(k-1)$ is the number of hypotheses at $k-1$. Hence the number of hypotheses will rapidly grow with time. Therefore, there is a clear need to limit the number of hypotheses. In order to eliminate the unlikely hypotheses, a few approaches have been proposed in the literature [1,2,9]:

i) The *Thresholding* approach is aimed at removing hypotheses with probabilities that fall below a predetermined threshold. Since the probability of each hypothesis decreases as the number of measurements increases in the association event, the disadvantage of pruning by this approach is that some of the probable hypotheses may be eliminated. This

problem can be overcome by computing the probability of the most likely hypothesis based on the threshold, and then by adjusting the threshold to a lower level for further processing. This procedure will be denoted as "*Dynamic thresholding*" and utilized in the PRUN1 of the MMHT shown in Fig. 1.

ii) The *Fixed Number* technique is aimed at allowing a predetermined number, say m, of the hypotheses to be maintained by ranking the hypotheses and choosing only the m most likely ones, as measured by the probabilities or score functions [2,6]. The limitation of this idea is that dissimilar hypotheses having almost the same probabilities may be eliminated, while retaining very similar hypotheses. Combining similar hypotheses prior to pruning partially addresses this problem. However, combining similar hypotheses is a complex procedure.

iii) The *Ranking Approach* is aimed at ranking and summing the probabilities of the more likely hypotheses. When this sum exceeds a threshold, the remaining hypotheses are then all eliminated. This pruning method can be an effective choice, however, the problem lies in terms of the computational cost for sorting a large number of hypotheses based on their probabilities for typical tracking situations.

Unfortunately none of the above methods was found to be effective in actual trials by many investigators. For efficient pruning, in the MMHT algorithm (Fig. 1), two pruning methods (dynamic thresholding and N-scan back pruning) are employed with appropriate modifications. The dynamic thresholding (the subroutine PRUN1 in Fig. 1) is an improved version of the thresholding scheme by dynamically updating the probability threshold. The dynamic threshold may be computed by multiplying a factor, for example 0.65, to the probability of the most likely hypothesis. Thus, the dynamic thresholding can reflect the entire hypotheses probabilities and does not eliminate the hypotheses with similar probabilities

In the modified MHT algorithm the major pruning is achieved by the N-scan-back approach (the subroutine PRUN2 in Fig. 1) in which only N-scan data (measurements from scan $k-N$ to k) are considered in the association and track management. The other old track information as well as measurements corresponding to scans from 1 to $k-N-1$ are ignored in the subsequent considerations. However, N-scan back approach does not yield sufficient pruning and hence a single hypothesis selection criteria is added. Therefore, the resulting output of the subroutine PRUN2 is the most likely hypothesis from each cluster and all other hypotheses are deleted from the corresponding clusters. This produces a significant improvement in pruning the unlikely hypotheses and ultimately resulting in a reasonably small number of hypotheses that can be maintained throughout the entire tracking operation. As well, these pruning processes make the MMHT an efficient algorithm that is suitable for applications in a real time environment while requiring only moderate computer resources .

3. NUMERICAL RESULTS

To illustrate the performance of the MMHT algorithm, the numerical results are presented in this Section. In Section 3.1, the recorded radar measurements are described and displayed in Figs. 2-5 and Figs. 7-8. The performance of the MMHT algorithm against the radar measurements is presented in Section 3.2.

3.1 The Radar Data

The plots of the radar data obtained from RATT (Raid Tracking Trials) are presented in Figs. 2 to 5. Five CF-18 fighter aircraft flying prescribed routes in tight formations served as raid targets for TEST1 and TEST2, and likewise, six aircraft flew for TEST3 and TEST4. The layouts of the two formations are shown in Fig. 6. The RATT routes are pentagonal in shape and the data that do not belong to the route are due to other targets in the vicinity as well as clutter. The prescribed altitude was 20,000*ft* and the speed was 500*knots*. The turn rates were about 1*g* for normal formations but at the end of TEST2 and TEST4 the aircraft executed 6*g* turns for 1 *nmi*.

A different set of plots consisting of RADAR5 and RADAR6 data are shown in Fig. 7 and Fig. 8, respectively. The examination of these plots reveals that the radar reports contain mainly

false returns due to heavy clutter. Especially RADAR6 data show only three distinct tracks while the rest of the data are clutter measurements originated from some other objects such as migrating birds. In fact, the number of such clutter data is about 200 per scan on the average. The extent of the clutter is highlighted in the zoomed-in view shown in Fig. 9.

3.2 Tracking Results.

To assess the performance of the MMHT tracker, evaluation tests were performed against the above mentioned radar data. For the N-scan back pruning, the parameter $N = 3$ was chosen. The results of the evaluation test against the radar data in Figs. 2-5 and Figs. 7-9 are presented in Figs. 10-13, respectively. From these Figures it is observed that as a whole, the MMHT estimated tracks closely follow the target trajectories, including the high-g maneuvers, close formation, crossing tracks, and scan misses in heavy cluttered environment. Further, it is noted that high-g maneuvers do not produce any remarkable degradation in the overall tracking performance. From the results of Figs. 10-13, it is clear that the MMHT is, indeed, robust to the various tracking conditions and shows consistent performance against many difficult track scenarios.

4. CONCLUSIONS

In this paper, an efficient implementation of the modified MHT algorithm has been developed for real time performance evaluations. The MMHT utilizes two pruning schemes to reduce the number of track hypotheses which grow rapidly with time in the original MHT algorithm. The dynamic thresholding and N scan back pruning are effectively implemented in the MMHT to produce a computationally efficient and real time implementable algorithm. The performance of the MMHT has been evaluated against eight sets of recorded radar measurements. From the test results it can be concluded that the MMHT algorithm yields reliable estimates even for difficult target trajectories such as close formation, crossing tracks, high-g maneuvers and scan losses under heavy clutters. This implies that the algorithm is robust in various distinct tracking situations. Further, the MMHT is fast enough for real time implementations while requiring only moderate computer memory.

Table 1. The CPU time and Memory used in the test.

Radar data & results	Mission time (sec)	Time per scan (sec)	No. of scans	No. of reports	CPU time(sec)	Mem. (Mb)
Figs.2 & 10	1800.0	5.0	360	5355	208.6	16.0
Figs.3 & 11	1815.0	5.0	363	4974	216.4	16.5
Figs.4 & 12	1810.0	5.0	362	6505	269.4	17.6
Figs.5 & 13	1805.0	5.0	361	5224	307.3	25.9
Figs.7 & 14	1128.0	12.0	94	7844	305.2	23.7
Figs.8 & 15	588.0	12.0	49	9214	530.2	15.7

In Table 1, the CPU time and maximum memory used in the evaluation test are summarized. The mission time is the total time of the mission which is given by the time per scan multiplied by the total number of scans. The CPU time and memory were measured on a Sparcstation 330 with 48Mb memory. From Table 1, it is clear that the algorithm is faster than real time while requiring only a moderate computer memory. Hence, the MMHT algorithm can be easily implemented for real time applications even on a personal computer system.

REFERENCES

[1]Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.

[2]S. S. Blackman, *Multiple-Target tracking with Radar Applications*, Artec House, 1986.

[3]Y. Bar-Shalom (Editor), *Multitarget-Multisensor Tracking: Advanced Applications*, Vol. I, Artec House, 1990.

[4]Y. Bar-Shalom (Editor), *Multitarget-Multisensor Tracking: Applications and Advances*, Vol. II, Artec House, 1992.

[5]P. Smith and G. Buechler, "A Branching Algorithm for Discriminating and Tracking Multiple Objects", IEEE Trans. on Automatic Control, vol. AC-20, Feb. 1975, pp. 101-104.

[6]A. Kountzeris, A Track Splitting Algorithm for Multiple Target Tracking, University of Sussex, School of Engineering and Applied Sciences, Report CE/S/43, Oct. 1989.

[7]Y. Bar-Shalom, "Extension of the Probabilistic Data Association Filter to Multitarget Environments", Proc. of the 5th Symposium on Nonlinear estimation, San Diego, CA, Sept. 1974.

[8]T.E. Fortman, Y. Barshalom, M. Scheffe, and S. Gelfand, " Sonar Tracking of Multiple Targets using Joint Probabilistic data Association", IEEE J. of Oceanic Engineering, Vol. OE-8, Jul. 1983, pp. 173-184.

[9]D. B. Reid, "An Algorithm for Tracking Multiple Targets", IEEE Trans. on Automatic Control, vol. AC-24, No. 6, Dec. 1979, pp. 843-854.

Fig. 2. TEST1 MEASUREMENTS

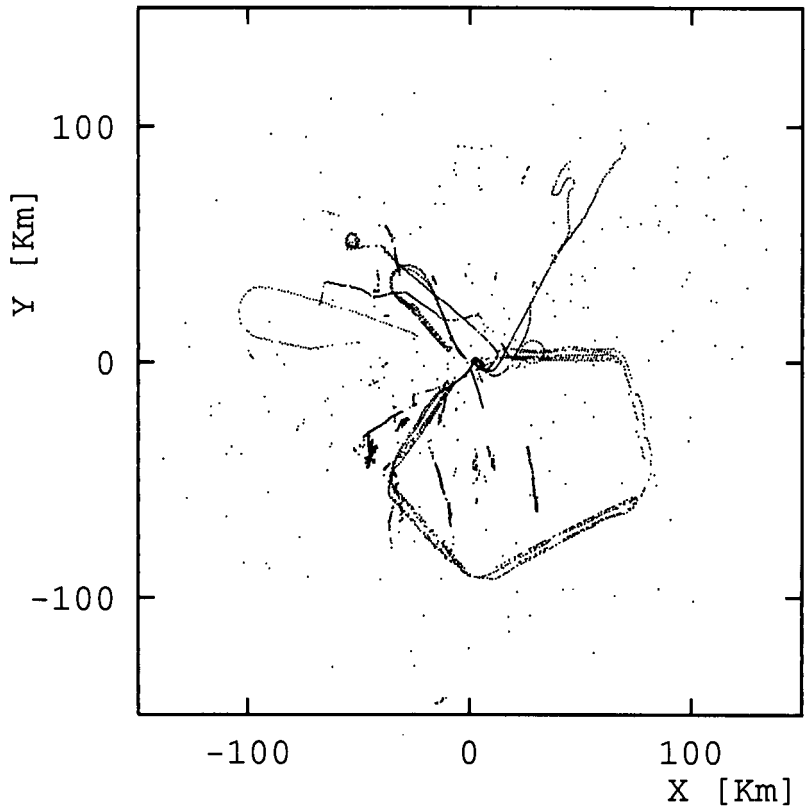


Fig. 3. TEST2 MEASUREMENTS

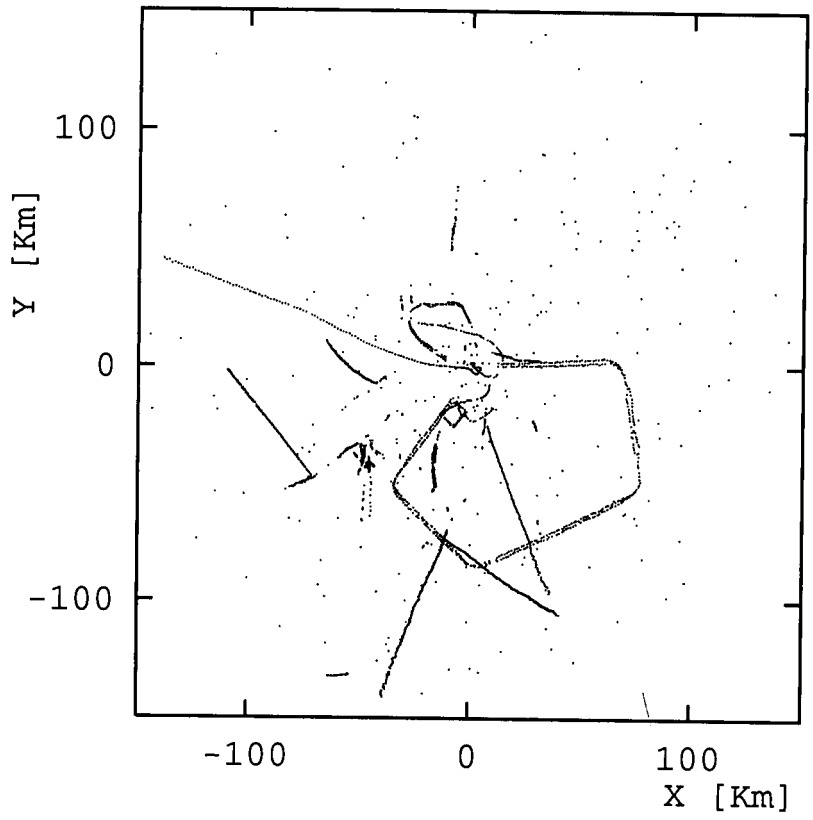


Fig. 4. TEST3 MEASUREMENTS

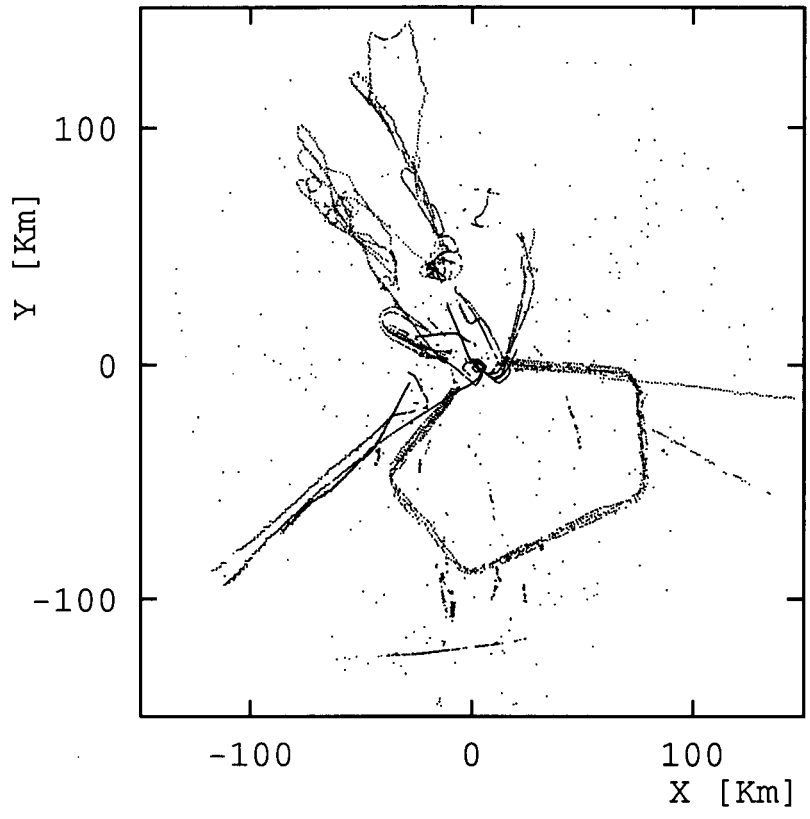
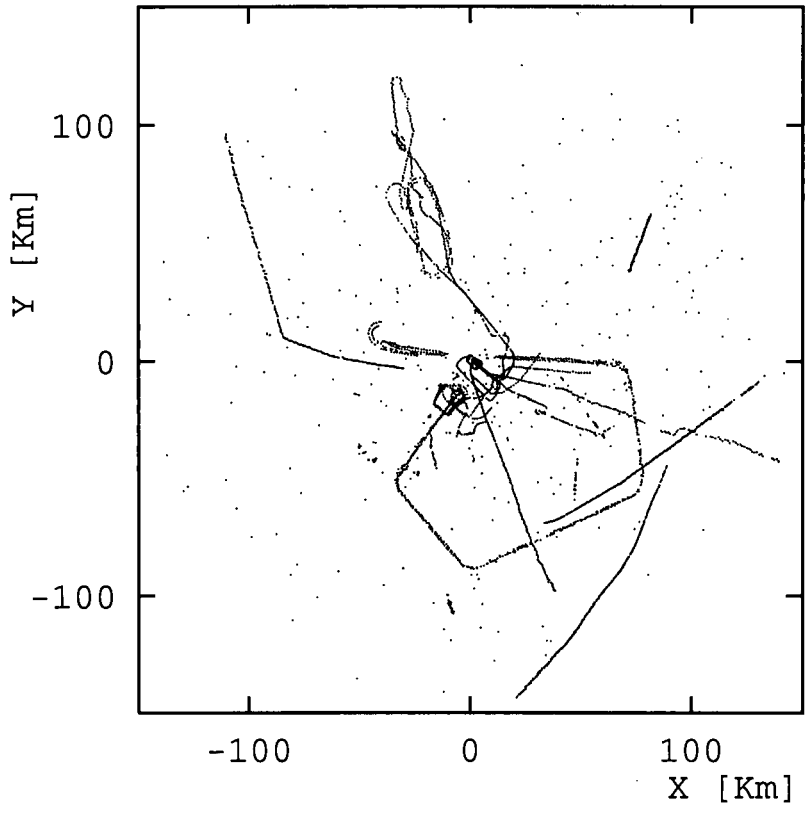


Fig. 5. TEST4 MEASUREMENTS



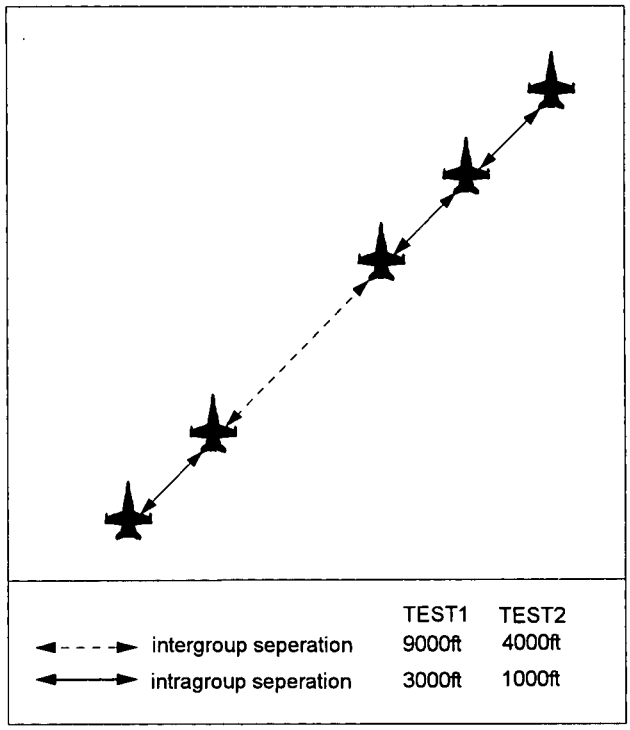


Fig. 6.a Formation layout for TEST1 and TEST2.

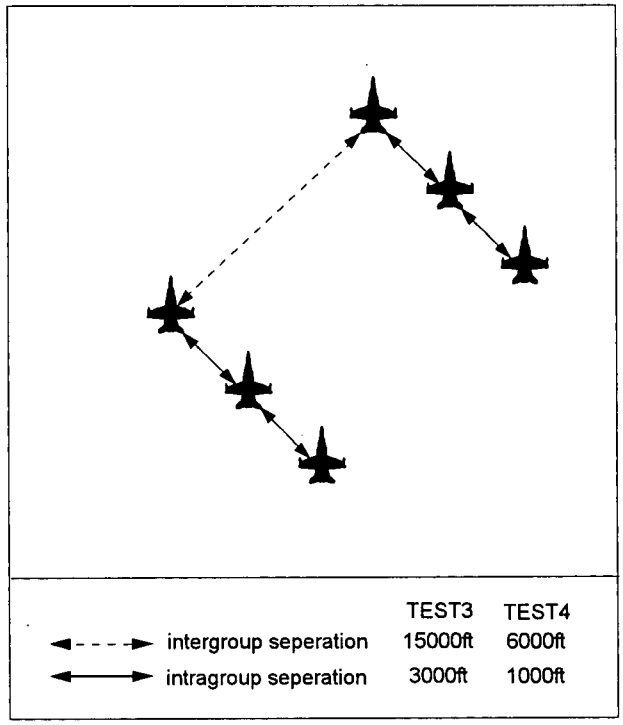


Fig. 6.b Formation layout for TEST3 and TEST4.

Fig. 7. RADAR5 MEASUREMENTS

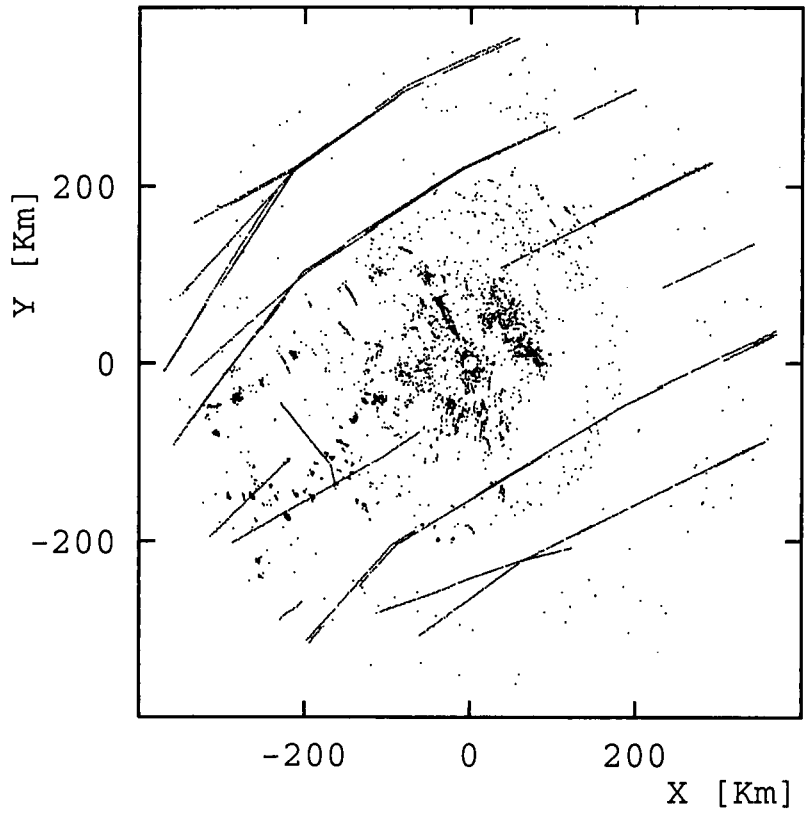


Fig. 8. RADAR6 MEASUREMENTS

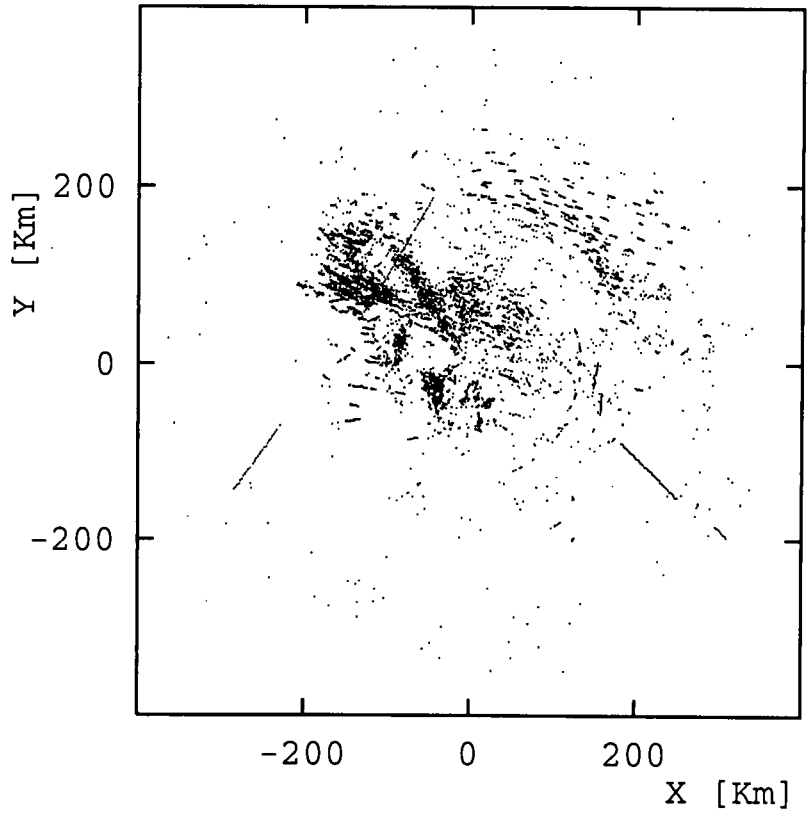


Fig. 9. RADAR6 MEASUREMENTS(ZOOMED-IN-VIEW)

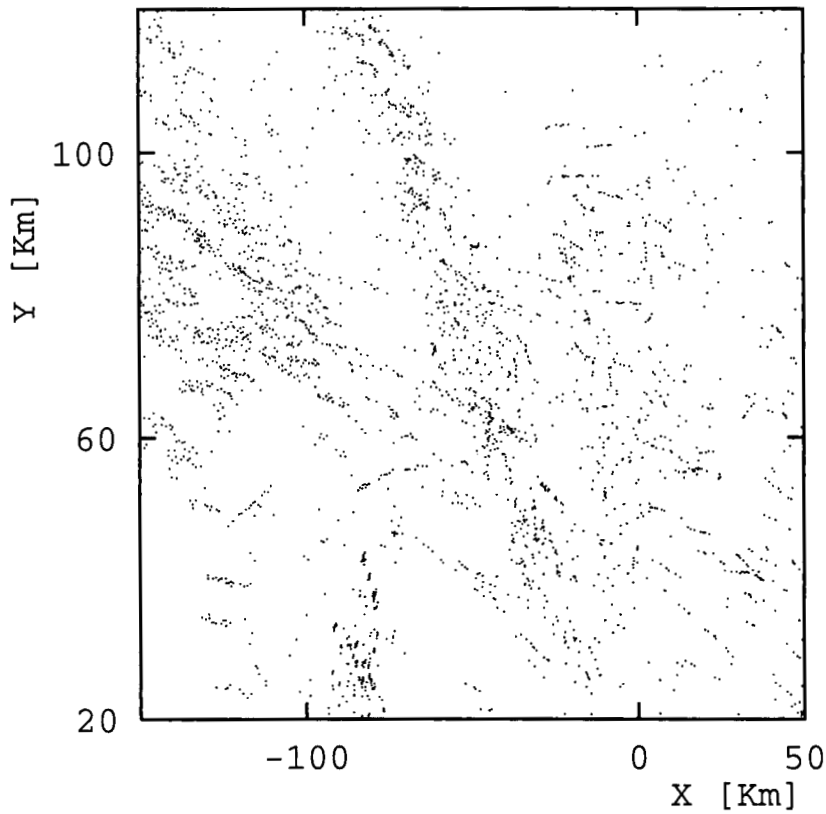


Fig. 10. MMHT TRACKING RESULTS OF TEST1 MEASUREMENTS

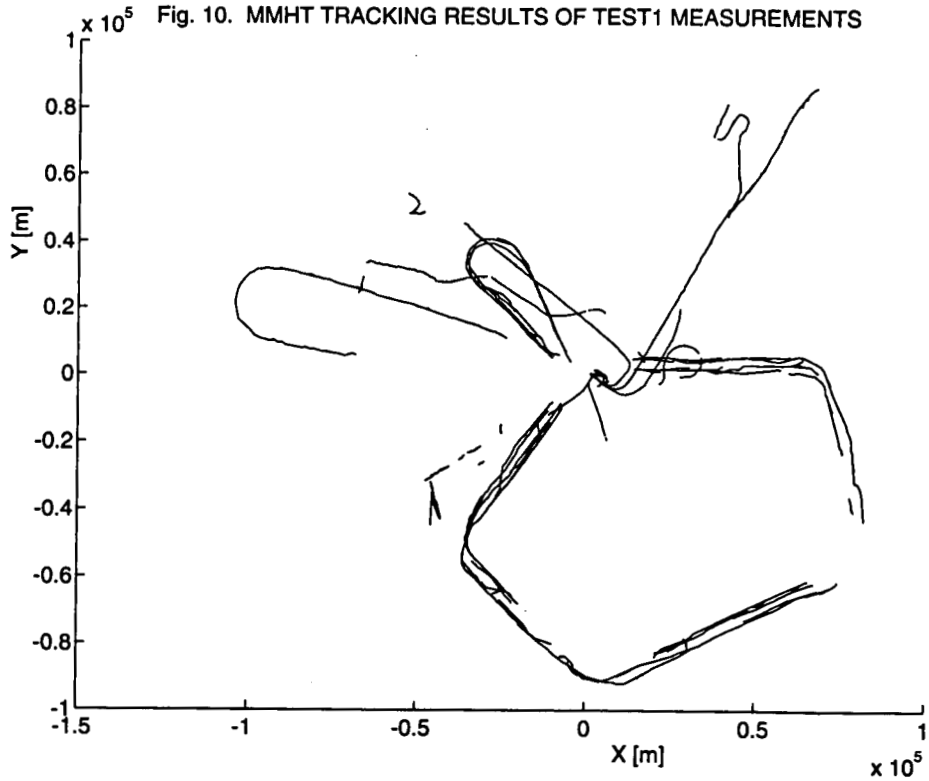


Fig. 11. MMHT TRACKING RESULTS OF TEST2 MEASUREMENTS

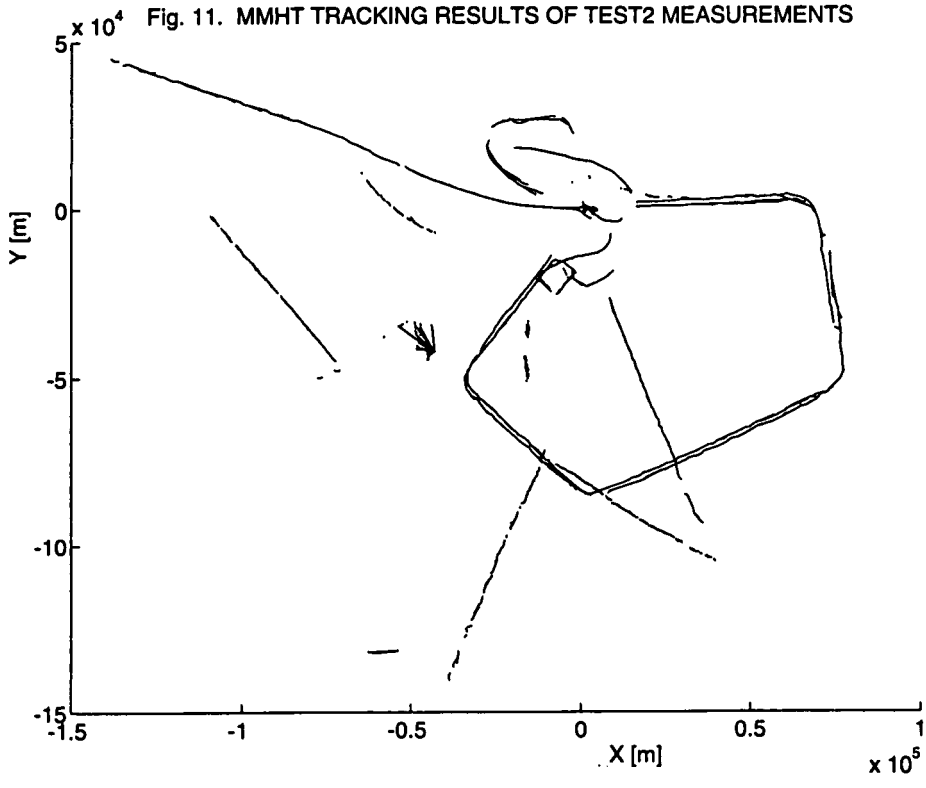
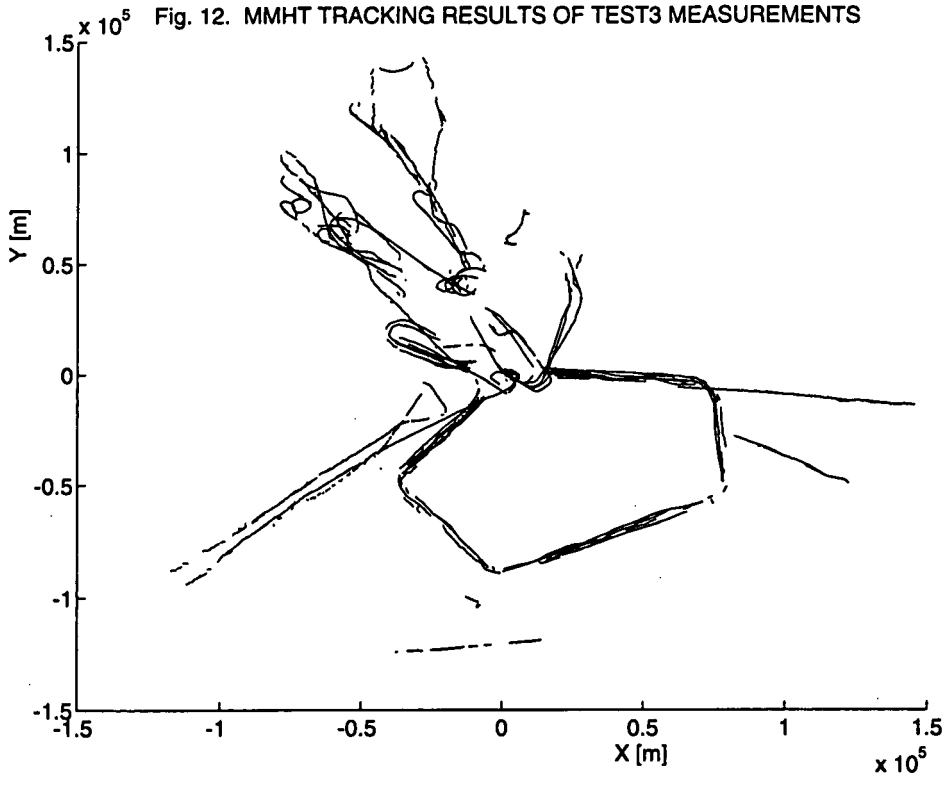


Fig. 12. MMHT TRACKING RESULTS OF TEST3 MEASUREMENTS



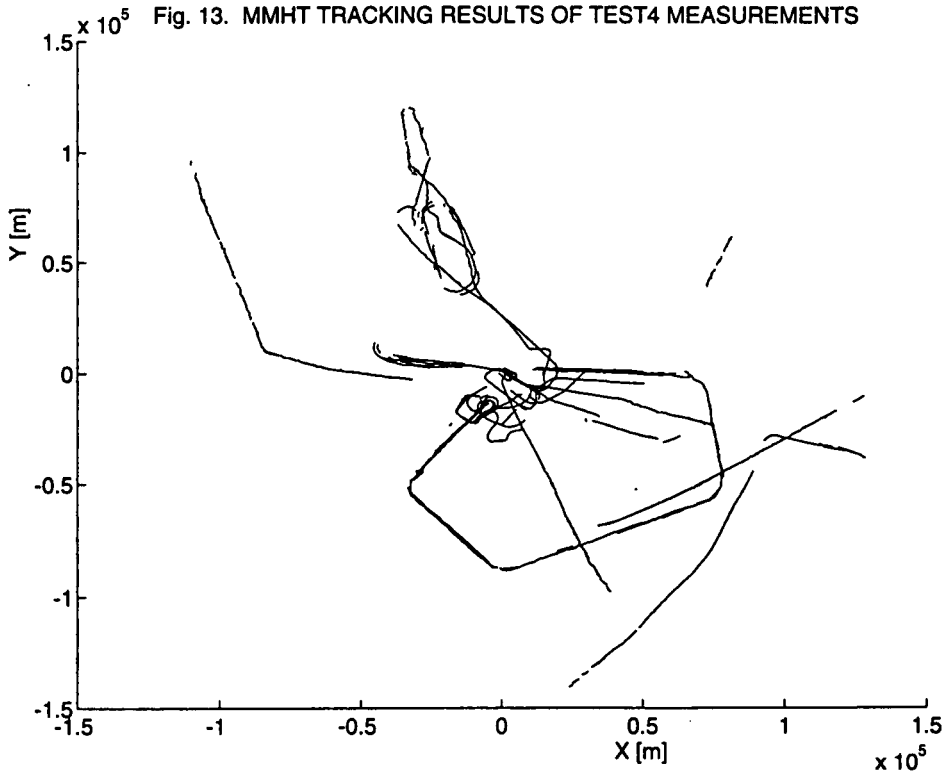


Fig. 13. MMHT TRACKING RESULTS OF TEST4 MEASUREMENTS

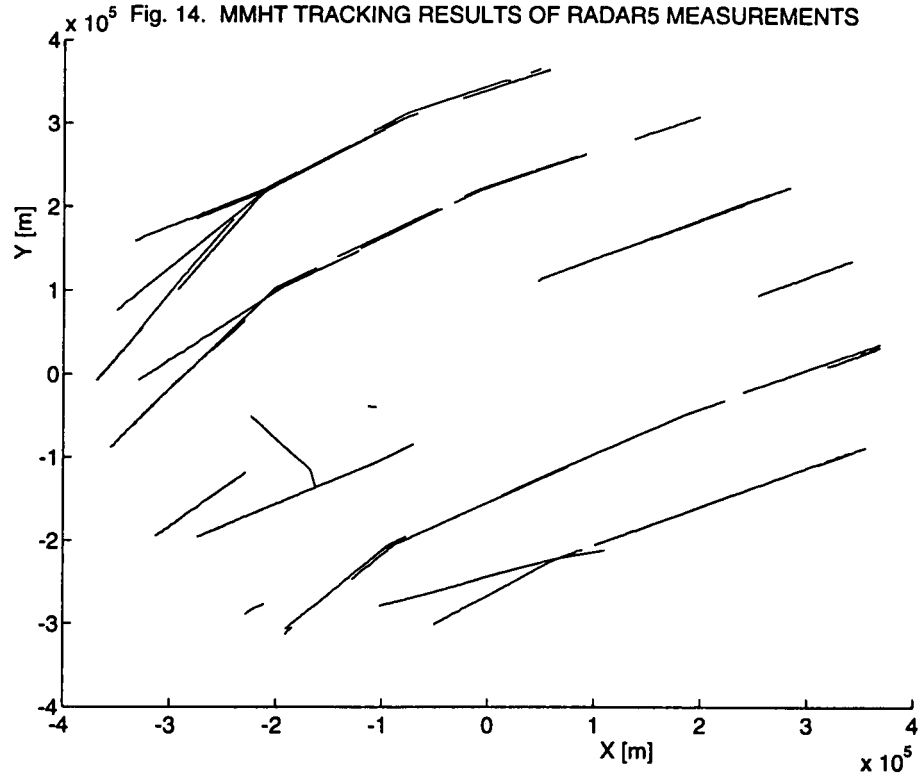
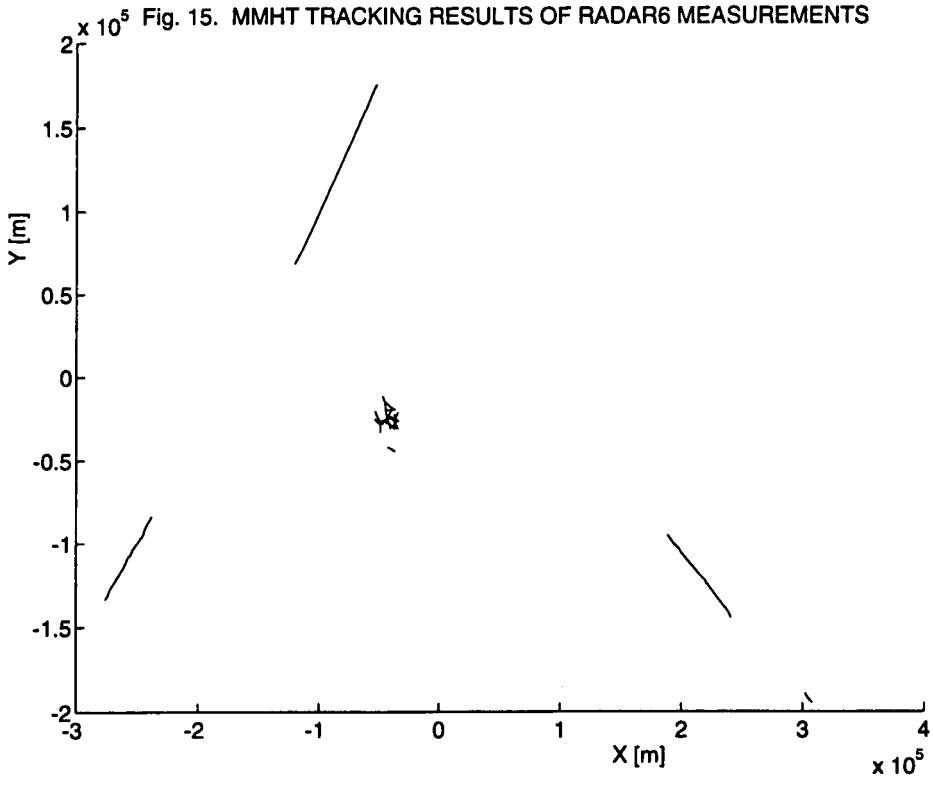


Fig. 14. MMHT TRACKING RESULTS OF RADAR5 MEASUREMENTS

Fig. 15. MMHT TRACKING RESULTS OF RADAR6 MEASUREMENTS



Low Cost Multi-Sensor Suites For Surveillance and Weapon Guidance

B. R. Suresh
Alliant Techsystems Inc.
600 Second Street NE
Hopkins, Minnesota 55343, U.S.A.

SUMMARY

With the changing geo-political landscape, there is an ever increasing need for low-cost sensors for application to surveillance systems and weapon guidance. Typically, these applications require MMW radars, IR imaging sensors, ladars and acoustic sensors as well as signal processing algorithms and high throughput miniature processors. The emphasis should be on developing low-cost individual sensors. The system applications then typically involve a suite of low-cost multiple sensors. The fusion of information from these sensors provides superior performance and an overall cost-effective product.

1. INTRODUCTION

"...Like looking for a needle in a haystack" is how General Norman Schwartzkopf described the problem of finding mobile Scud missiles in January 1991. In the aftermath of Desert Storm, battlefield surveillance and precision weapon targeting in adverse weather have emerged as significant problems to be encountered by NATO forces in future conflicts. Targets of interest include fixed high value targets, time-critical fixed, and mobile high value targets and vehicular targets. There is evidence that these problems can be solved by using multi-spectral sensors. We examine the problem from the perspective of an unattended ground sensor for surveillance and an autonomous smart weapon. A key element in developing such sensors is keeping them cost effective. An assessment of and prognosis for future multi-sensor systems will be provided.

2. MULTI-SENSOR TRADES AND SYSTEM ARCHITECTURES

Over the years, significant dollars have been invested in the development of low-cost sensor technologies. Promising low-cost sensors include MMW radar, passive imaging IR, Ladar, and Acoustics. Table 1 lists the operating characteristics of these sensors [1]. Table 2 delineates the target discriminants derivable from these sensors. It is apparent that these sensors have differing strengths and weaknesses. The design of a multi-sensor system requires the matching of complementary characteristics provided by these sensors. The benefits of such a multi-sensor system include adverse weather operation, countermeasure resistance, and high performance at low cost.

There is a popular misconception that multi-sensor systems are unaffordable; that dual mode costs twice as much as single mode, and can never be cost effective as single mode. It is important to recognize that dual mode is not, for example, the "best" MMW attached to the "best" IR; doing so indeed would defeat the very spirit of seeking a dual mode solution.

As shown in Figure 1, the key element in the engineering of a dual mode system is the degree to which the individual sensors are balanced to exploit their synergy and performance. This is in contrast to driving one sensor to be a "Cadillac" and expecting it to deliver most of the performance; since this ends up driving the sensor into the domain of diminishing returns, cost effectiveness will suffer. By keeping each sensor low cost (and thus, low performance) and using their synergy to regain the high performance, cost effective multi-sensor system designs can be obtained.

The ATR (Automatic Target Recognition) function in most RSTA systems and weapon seekers can be performed by a judicious combination of MMW radar, FLIR, and ladar sensors. Figure 2 depicts the roles of different sensors in the ATR function. The left hand column in the figure depicts the hierarchy of ATR functions—ROI search, detection, recognition, and identification. The roles of the different sensors in performing these functions is also indicated there. Due to the laws of physics, each sensor is inherently adapted to performing certain functions, versus others. For example, MMW radar due to its longer wavelengths, is well suited for rapid wide-area search and ROI selection. In contrast, imaging ladars with their high resolution 3-D signatures are very useful in target identification. Passive FLIRs have strengths in functions that lie between the functions where radars and ladars are strong. The key to a low cost sensor system lies in a judicious engineering of a multi-sensor suite where each sensor is performing in a physical domain suited to itself, and therefore can be mechanized at low cost. On the other hand, one can force a ladar to perform all the functions, including ROI search. No doubt, such a ladar can be built; but mechanizing such a device which challenges the laws of physics costs money—lots of it!

Figure 3 shows the sensor fusion architecture alternatives for smart weapons as an example. A variety of cost/performance trades are available for a variety of applications. The applications span the spectrum from very simple sensor fuzed munitions to more potent terminally guided missiles. The bottom line is to recognize that a multi-mode sensor is not a monolith; a family of multi-mode solutions exists to address varying cost/performance needs. Figure 4 illustrates a family of smart weapons that can benefit from a family of simple to sophisticated multi-mode sensors.

Table 1. Sensor Characteristics

	Active	Passive	Day/ Night Use	Adverse Weather Use	Viewing Geometry Signature Robustness	Temporal Signature Robustness	Rapid Scan	Broad Area Coverage	Cost	Non-Line-of-Sight
Passive IR	No	Yes	Yes	No	Yes	No	Yes	No	Low	No
MMW Radar	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Low	No
Ladar	Yes	No	Yes	No	Yes	Yes	No	No	High	No
Acoustics	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low	Yes

Dual mode system engineering involves partitioning the problem into domains best suited for each sensor — i.e., divide and conquer

T16326_13

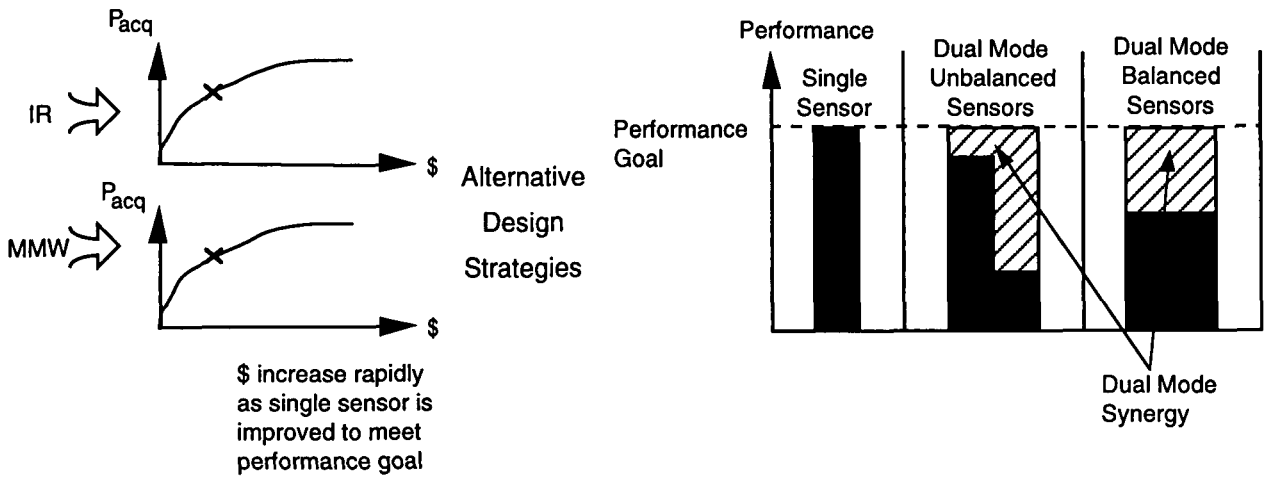
Table 2. Sensor Discriminants

	Down-Range Reso-lution	Cross-Range Reso-lution	Eleva-tion Reso-lution	Dop-pler	Trans-verse Veloc-ity	Vibra-tion	Polar-ization	Reflec-tance	Radi-ance	1-D Shape	2-D Shape	3-D Shape	Acoustic Emis-sions
Passive IR	No	Yes	Yes	No	Yes	No	No	No	Yes	Yes	Yes	No	No
MMW Radar	Yes	Yes*	Yes*	Yes	No	Yes	Yes	Yes	No	Yes	Yes*	Yes*	No
Ladar	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Acoustics	No	No	No	Yes	No	Yes	No	No	No	No	No	No	Yes

* With monopulse imaging

Synergy involves capitalizing on complementary discriminants to create new, robust discriminants

T16326_14



- "Balanced" Dual Mode ⟶ Lowest Sensor Cost
- Synergy Algorithms ⟶ Desired High Performance

T16326_01

Figure 1. Balanced Dual Mode Sensors with Synergy Exploitation Lead to Low Cost Solutions

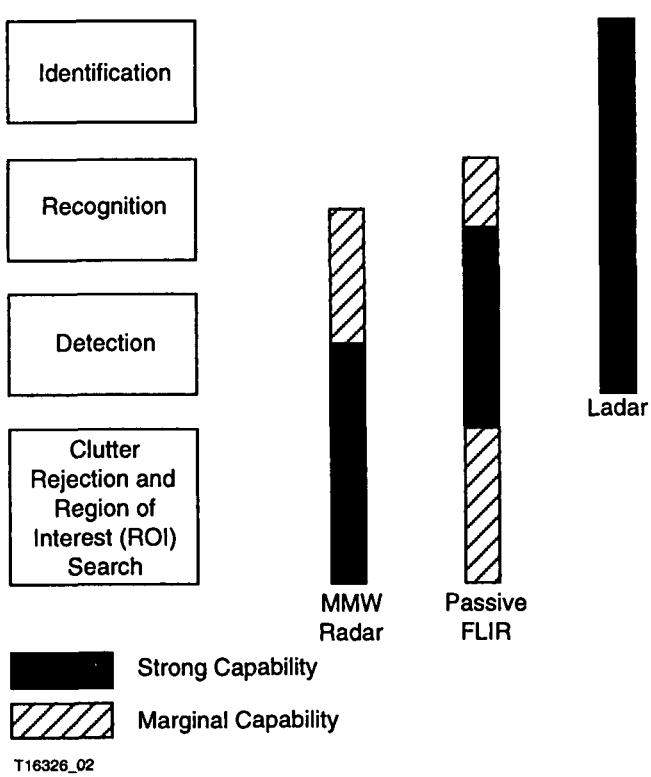


Figure 2. Multi-Sensor Roles in ATR (Automatic Target Recognition)

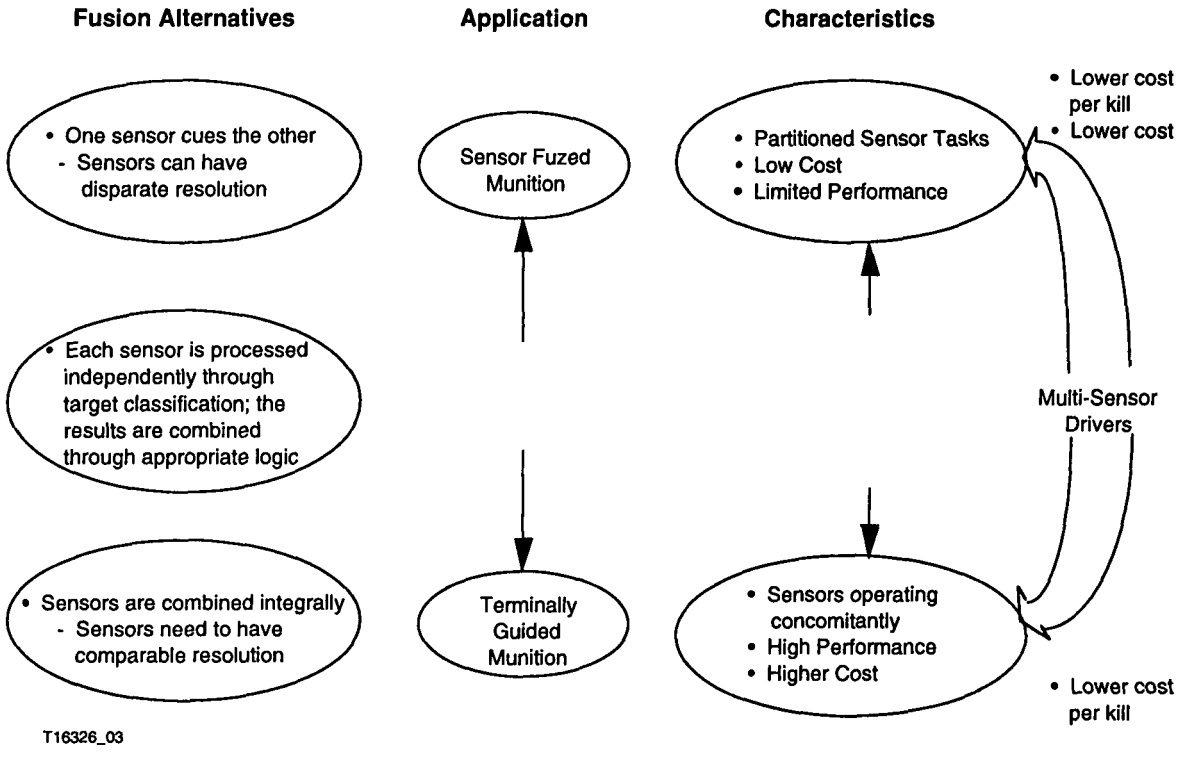


Figure 3. Sensor Fusion Architectures and Trades

3. ADVANCED SENSOR TESTBED

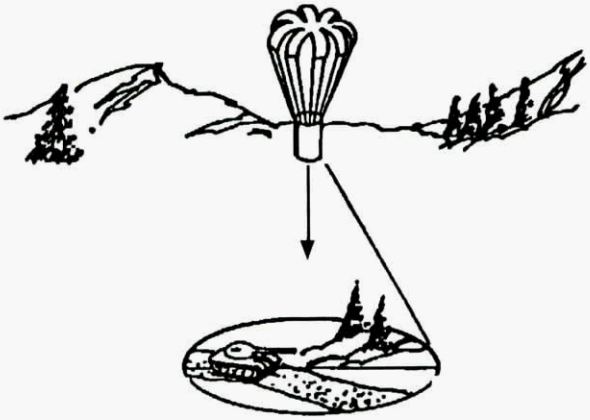
A key tool required in the development of multi-sensor systems is an Advanced Sensor Testbed to support sensor system R&D. Figure 5 shows a schematic of such a testbed, which is designed to be a flexible and rapidly reconfigurable tool to support a variety of system development programs. This testbed was developed by Alliant Techsystems as part of its ongoing R&D program. We have also developed a variety of sensors which "plug-in" to the testbed to permit evaluation of alternate sensor concepts. Sensors include MMW radars (W-band, Ka-band, and V-band), uncooled imaging IR sensors, imaging ladars, and acoustic arrays. The testbed is capable of being tower tested or flight tested.

Figures 6 and 7 depict the advanced sensor testbed including the instrumentation. This configuration includes a dual mode sensor suite with a MMW radar and an uncooled IR sensor. Figure 8 shows some of the target signatures collected with the advanced sensor testbed. These signatures can be used for phenomenology analysis, sensor evaluation, and algorithm development.

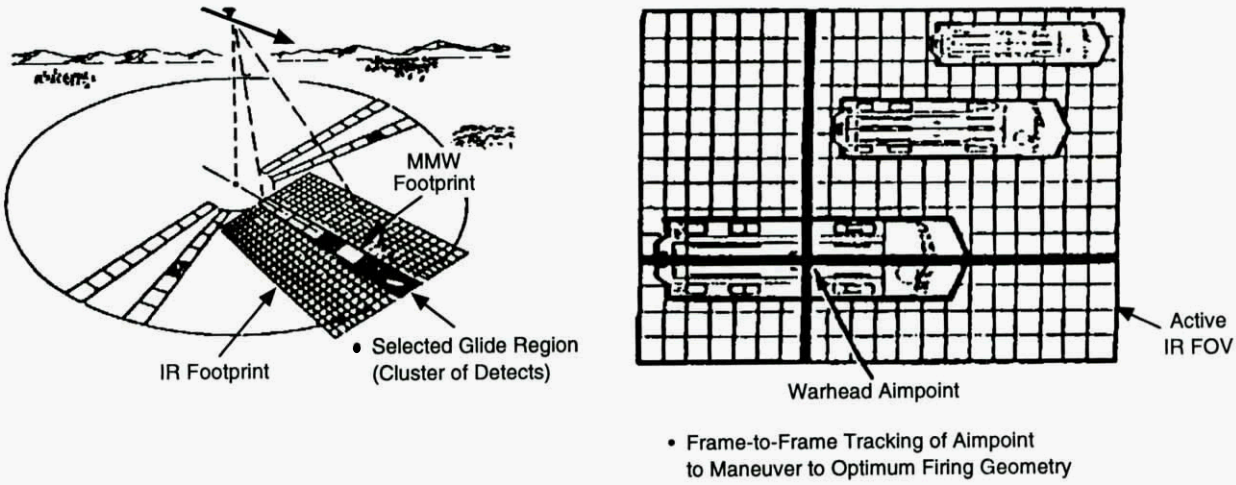
Imaging IR sensors are now becoming increasingly affordable with the advent of uncooled IR sensors. A promising approach to uncooled IR sensors is based on the unique microbolometer technology [2]. Unlike many cryogenically cooled IR FPA sensors, the microbolometer uncooled IR sensor has the following benefits:

- The microbolometer is based on monolithic silicon technology, and is extremely low cost
- Amenable to long duration, unattended operation
- Requires no cooler or chopper, leading to low power operation
- Mechanically rugged with no microphonics
- Long-wave (8-12 μm) operation with broadband response.

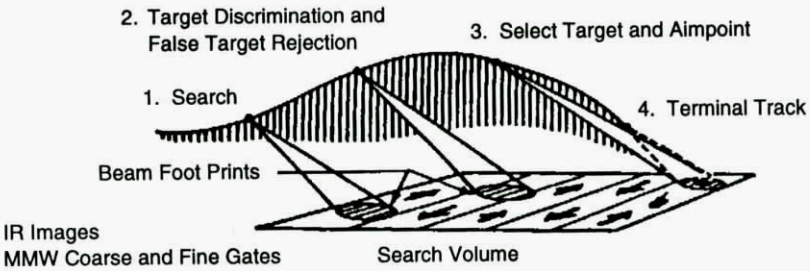
Figure 9 depicts an uncooled IR sensor (developed by Alliant Techsystems) along with its salient parameters. FPAs with 240 x 336 pixels have been fabricated with excellent uniformity characteristics. Note the sharp, high contrast imagery produced by the sensor.



Sensor Fuzed Munition



Maneuvering Sensor Fuzed Munition



T16326_04

Terminally Guided Munition

Figure 4. Multi-Sensor System Application to Smart Weapon Family

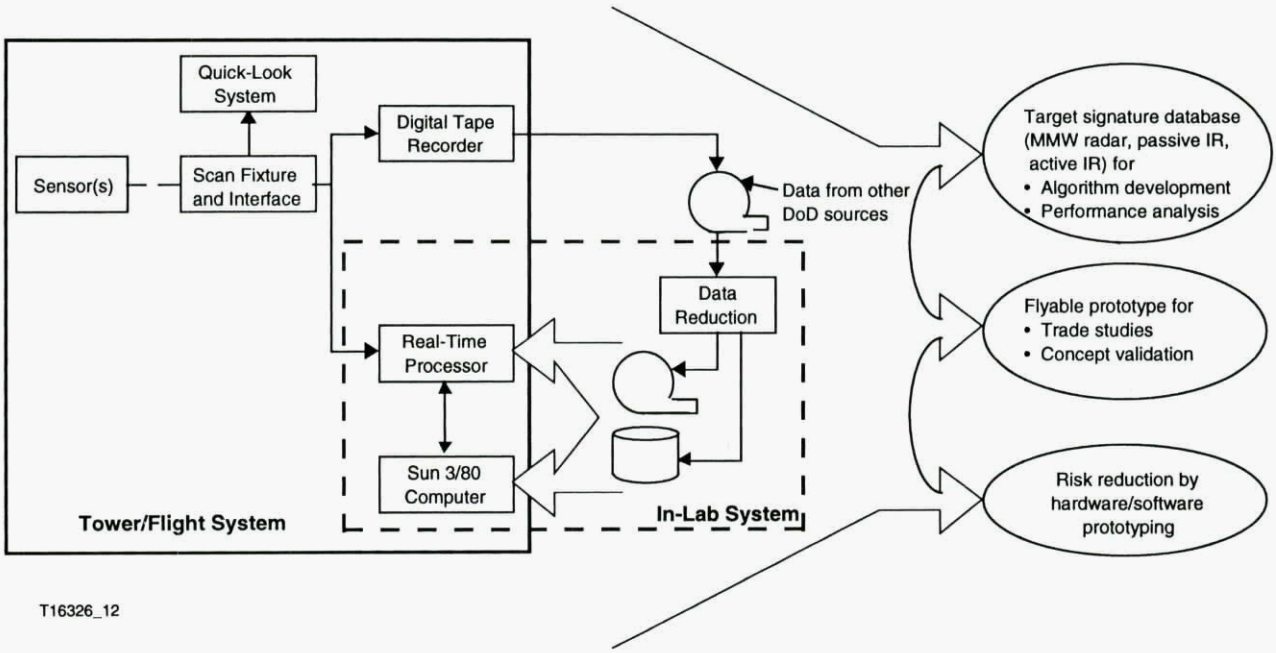


Figure 5. Advanced Sensor Testbed. A rapidly reconfigurable testbed for use on Multiple Sensor Development programs

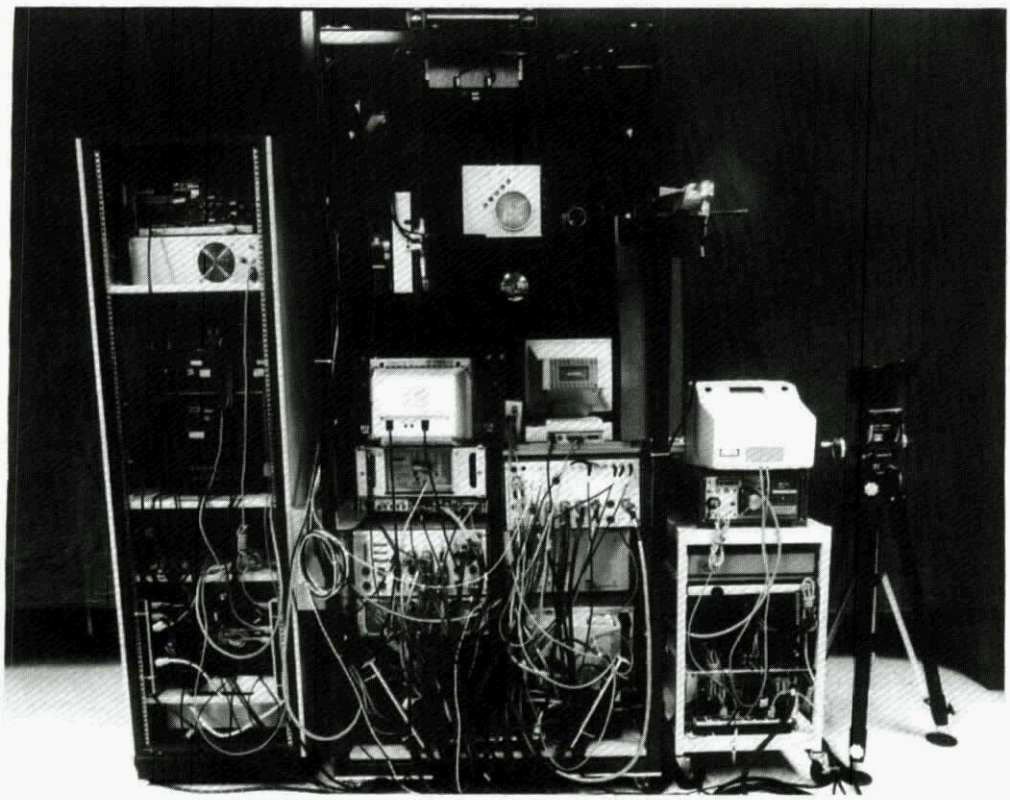


Figure 6. Sensor Testbed Front View. Sensor apertures are in upper center of photo.

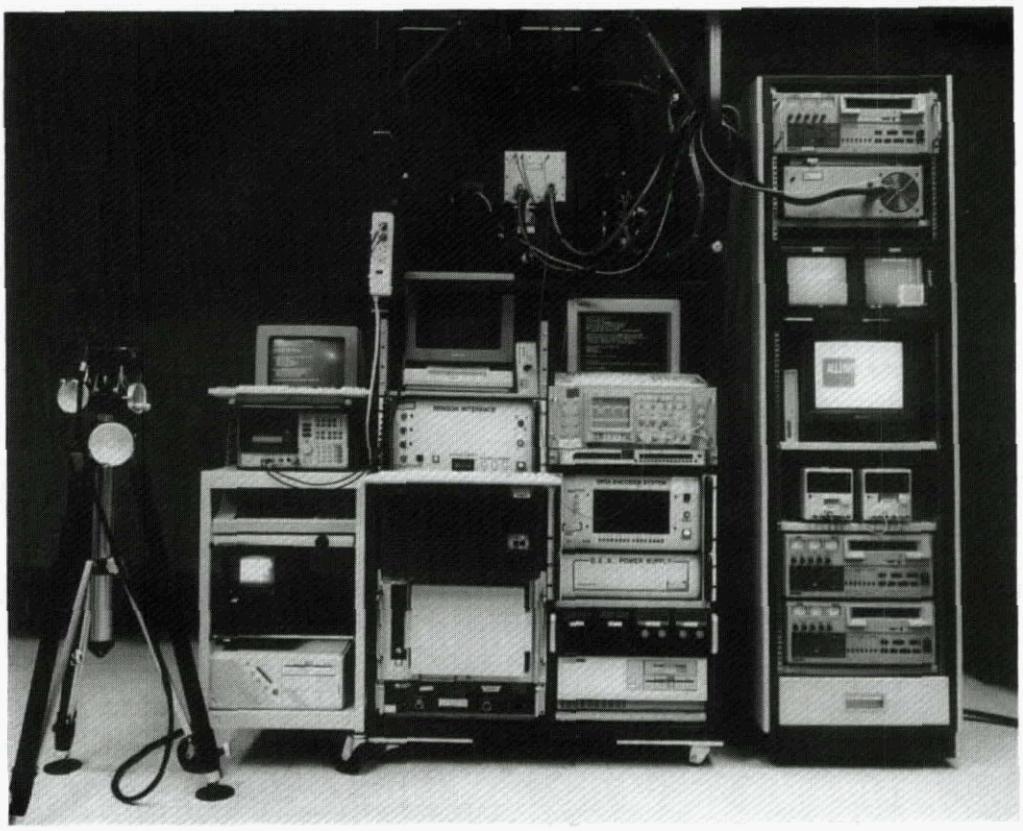
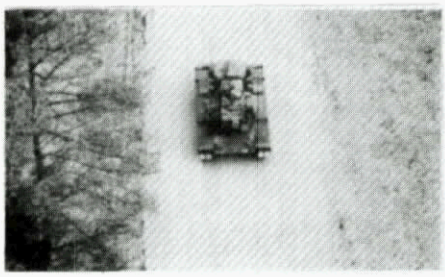


Figure 7. Operator Side of the Sensor Testbed. All controls and monitors are available to the operator on this side.

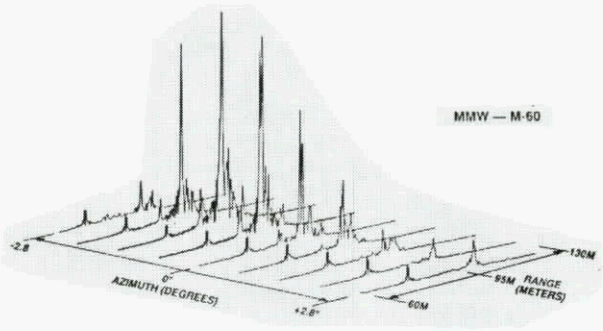
Algorithm and processor development generally follows a 3-step approach as shown in Figure 10. Algorithms are initially developed on general purpose computers such as VAX and SUN stations. In the second step, algorithms are ported to a HOL-programmable, real-time processor testbed. This allows field evaluation of algorithms for confidence building. Finally, in the third step, a packaged processor is built. Such a packaged processor is exemplified by Aladdin. Aladdin is a compact (4-inch diameter) high throughput processor intended for multi-sensor signal and image processing [3]. Aladdin has a modular architecture and can be scaled and reconfigured for different applications. Aladdin is programmable in Ada or C. Since the processors in steps 2 and 3 of Figure 10 are functionally equivalent, it greatly facilitates software migration.

A key to the success of a multi-sensor system is the maximizing of the synergy between the sensors. This can be done by developing fusion algorithms that maximize this synergy. It is this fusion synergy which is instrumental in yielding high performance out of two relatively low-cost sensors. Multi-sensor fusion can be performed by combining information from two sensors at various informational hierarchies. The approach which yields the highest synergy, called concomitant processing, is depicted in Figure 11. Concomitant processing maximizes the synergy by integrally fusing the information from two sensors at all possible hierarchical levels.

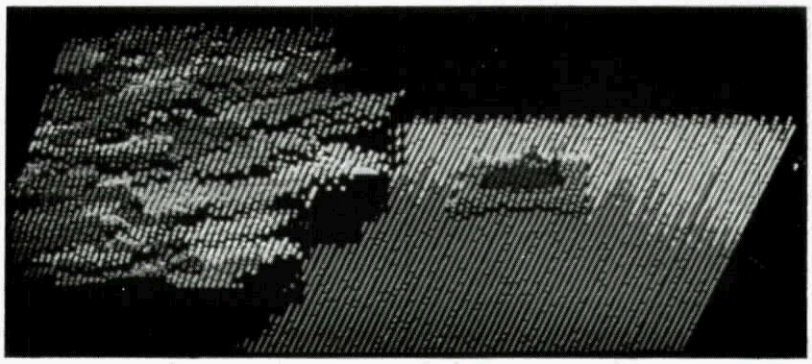
Multi-sensor systems can be leveraged for surveillance system and smart weapon seeker applications. These system applications are described next.



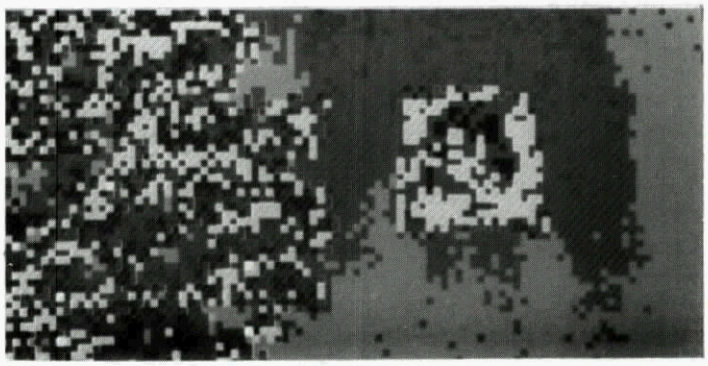
Ground Truth Tank Target



94 GHz MMW Signature



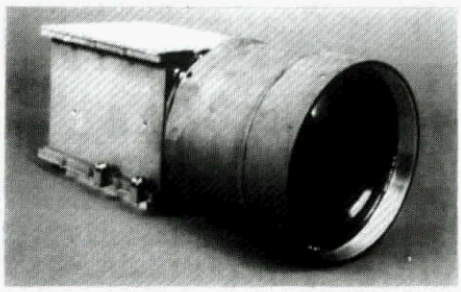
False-Color Ladar Range Image



False-Color Ladar Intensity Image

Figure 8. Target Signatures Collected with Advanced Sensor Testbed

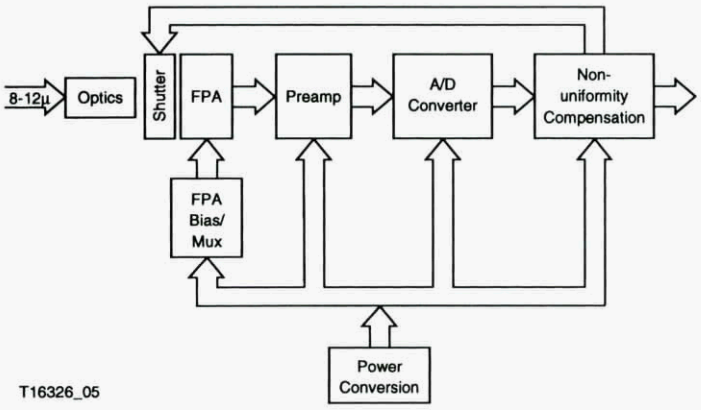
LOCUSP Camera With 2-FOV Lens



System Description

- Low cost uncooled FPA technology
  - > 80,000 pixels
  - ≤ 0.05°C NETD
- Chopperless operation
- 15° x 9°, 5° x 3° dual FOV
- 1.0 mrad, 0.33 mrad IFOV
- Alternate optics interface
- Manual/auto gain control
- 24 Vdc, 13 watts operation
- RS170 video out
- 7.3 lbs without optics
- 5.1" x 5.3" x 7.1" (H x W x L) without optics

Block Diagram



March 1993 Imagery

Figure 9. Uncooled IR Sensor

Image Research Laboratory

- Concept development
- General-purpose computer (VAX, Sun, etc.)
- HOL programmable
- Nonreal-time
- Limited confidence in algorithm performance

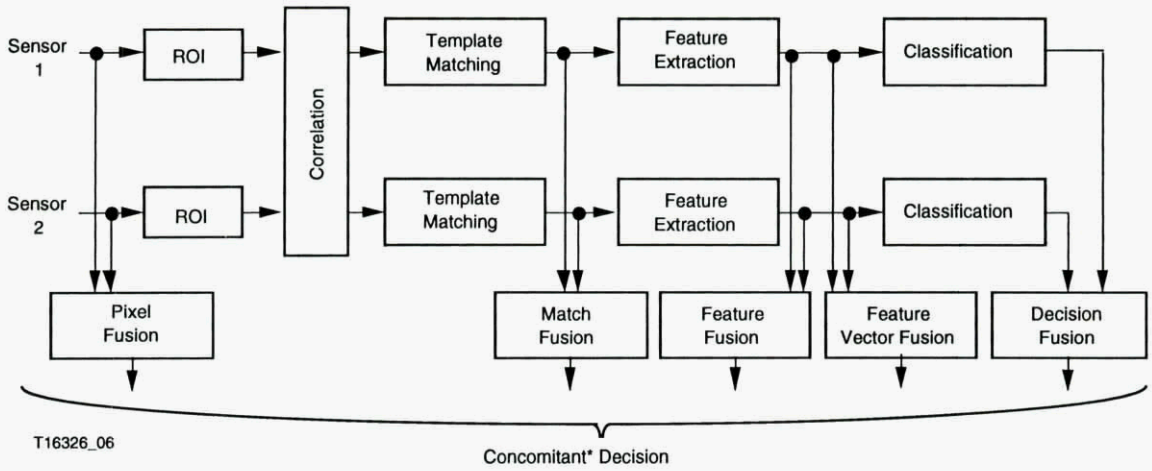
Flexible Testbed Processor

- Concept validation
- Semi-custom hardware
- HOL programmable
- "Real-time"
- High confidence in algorithm performance

"Miniature" Processor (ALADDIN)

- Operational scenario validation
- Rapidly reconfigurable modules
- Form/fit hardware
- HOL programmable
- Real-time

Figure 10. Real-Time Processor Evolution



***Concomitant:** "Something That Accompanies Or Is Collaterally Connected With Something Else" - Webster's New Collegiate Dictionary

Figure 11. Concomitant Dual Model Fusion Algorithm

4. SURVEILLANCE APPLICATIONS

The Desert Storm experience highlighted the strategic benefits of battlefield surveillance and situational awareness. Future conflicts will very definitely involve the rapid deployment of light forces. These light forces are deficient in anti-armor capabilities and lack protection against the heavy systems they may encounter before additional reinforcements arrive. The lethality and survivability of such light forces requires surveillance systems—unattended ground sensors capable of long duration operation. Such systems are applicable to forward observation for fire control, mobile reconnaissance (ground based or airborne), and covert intel operations.

The key technical problem in such a battlefield surveillance sensor system is one of reliable, wide area target detection with low false alarms and real-time threat imagery transmission over an RF communication link. This problem is exacerbated by the fact that military communication channels are typically bandwidth constrained. Intelligent bandwidth compression techniques which combine automatic target recognition with conventional video bandwidth compression can often be used to get over the bandwidth bottleneck.

Figure 12 depicts a conceptual battlefield surveillance sensor system. It contains an acoustic sensor which is on all the time. On detection of any activity, it serves to "wake-up" and turn-on the rest of the system. Such power management is necessary to enable long duration operation on a battery. The main targeting sensors are an imaging IR sensor and low light level TV. A GPS is provided for accurate location of the sensor and an electronic compass and laser rangefinder provide precise target location with reference to the surveillance sensor. A processor is provided for sensor signal processing and a radio transmits the image data back to a control station.

Uncooled IR is an ideal choice for such a surveillance system. The uncooled IR sensor is compact, low cost, and low in power. Since it is uncooled (does not require replenishment of liquid Nitrogen), it is amenable to long duration unattended operation. It is rugged and can be air dropped. The broadband response allows multi-color operation for enhanced target detection.

5. SEEKER APPLICATIONS FOR WEAPON GUIDANCE

Seekers are currently being developed in the U.S. for fixed high value targets (e.g., bridges, bunkers) as well as mobile targets (e.g., SSM/TEL, tank, APC). The great scud hunt during Desert Storm highlighted the need for precision strike against targets.

Examples of weapons intended for fixed high value targets include the JDAM and JSOW. JDAM is a unitary bomb, which in its baseline configuration incorporates INS/GPS guidance. JSOW is a dispenser for cluster munitions or a unitary warhead, which also embodies INS/GPS guidance. Both JDAM and JSOW can have their precision strike capability significantly enhanced by the addition of a simple, low-cost terminal seeker.

Such a guidance approach to precision strike is illustrated in Figure 13. Given apriori knowledge of the location of fixed high value targets, INS/GPS guidance can be used to deliver the weapon within a small error basket—typically 13-30M CEP. A terminal seeker can then be used to target a precise aimpoint to get 3M CEP. Since the search area for the seeker is greatly reduced by the INS/GPS, a low-cost endgame seeker can be utilized for precision strike.

A variety of seeker technologies have been explored for the fixed high value target seeker. Table 3 lists the pros and cons of many of these seekers—the bottom line is that there is no acceptable solution. A dual mode MMW/IR seeker provides an attractive alternate choice for the fixed high value target seeker. A dual mode seeker keys on the fact that the seeker only has to search a small area localized by the INS/GPS. The MMW operates effectively in adverse weather at long range to provide

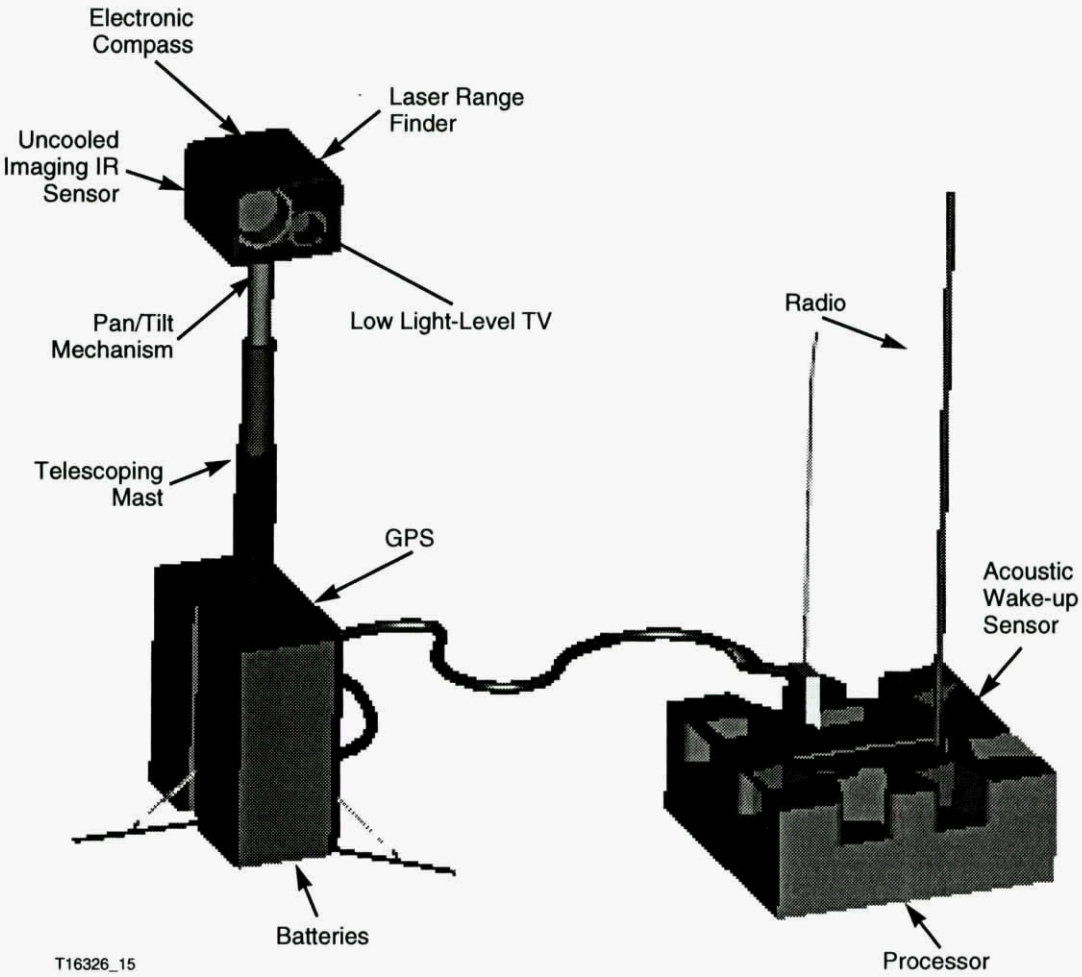
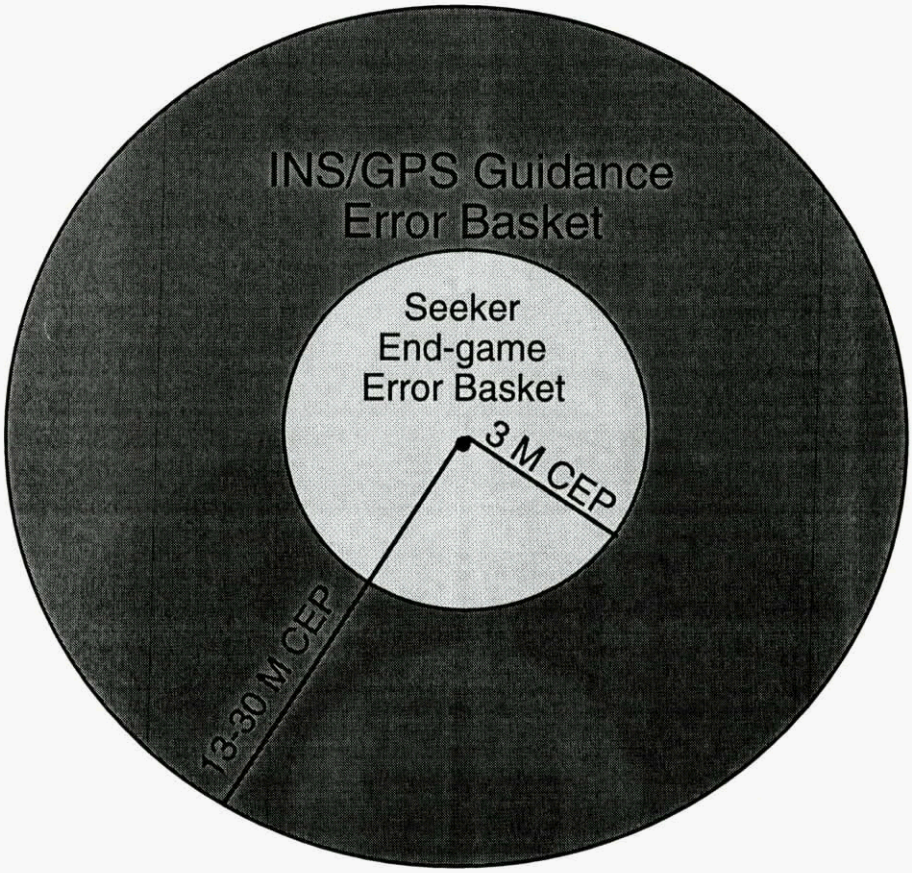


Figure 12. Battlefield Surveillance Sensor System



T16326_07

Figure 13. Precision Strike Approach using INS/GPS in Combination with a Terminal Seeker

Table 3. Fixed High Value Target Seeker Candidates and Their Trades.

Seeker Candidate Technology	Advantages	Disadvantages
MMW (real beam)	• Mature technology • Adverse weather performance • Low cost	• Does not meet accuracy requirement (CEP)
IR (cryogenically cooled)	• Mature technology • Meets accuracy requirement	• High cost • Poor adverse weather performance
SAR	• Good adverse weather performance • Comes close to meeting accuracy requirement	• High cost • Requires offset aimpoint which affects terminal accuracy
Ladar	• Meets accuracy requirement	• High cost • Poor adverse weather performance • Technology not fully mature

Dual-Mode MMW/ IR

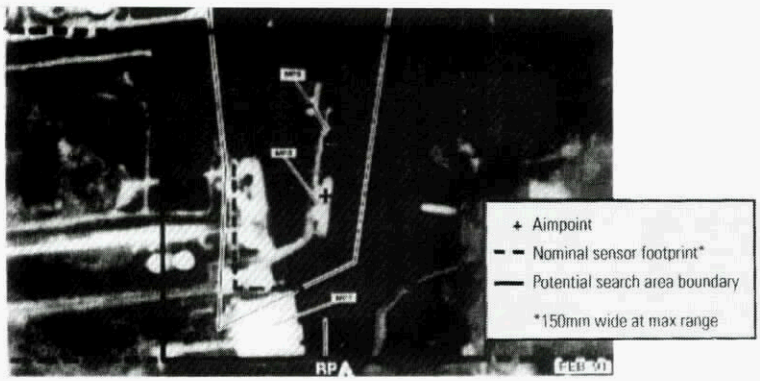
- MMW guides IR to target basket in adverse weather
- IR provides desired aimpoint accuracy
- Low cost

T16326_08

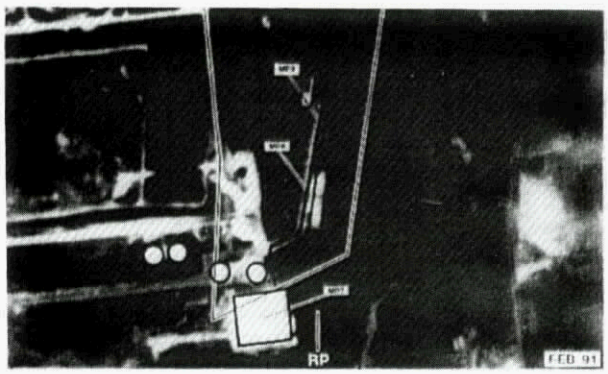
an initial target fix; since the MMW is scanning a small localized area, it delivers adequate performance to get into a target basket until the I^2R effective range is reached. The MMW then hands off to the high resolution I^2R for the endgame precise aimpoint tracking. In clear weather, the I^2R would get the hand-off at a longer range. Such a sensor resource management is well-suited for an autonomous fixed high value target seeker, and a terminal accuracy of 3m CEP is possible.

A key challenge in the development of high value target seekers is the establishment of a methodology and infrastructure for target area mission planning. The general strategy for mission planning is depicted in Figure 14. One starts out with, for example, a Basic Target Graphic of the target area. Salient target features are extracted by the mission planner to generate a scene description file. This file is then transformed to the domain of the seeker sensor (IR, MMW, etc.) by sensor modeling tools. The transformed file is then loaded into the seeker for target acquisition using correlation and scene matching algorithms. The challenge of developing mission planning techniques will greatly benefit from the advances in sensor modeling and computer graphics. The commercial infrastructure set up by the videogame and entertainment industries can also provide benefits here. Oddly enough, the current generation of children growing up with Nintendo may provide the recruiting ranks for the future developers and operators of mission planning workstations!

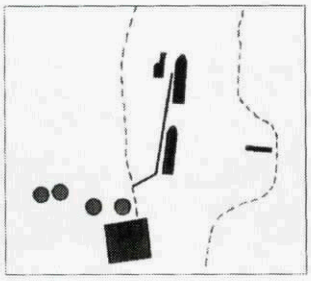
Alliant Techsystems has built and flight tested MMW seekers for high value target acquisition. Figure 15 shows a POL refinery being acquired by the MMW seeker in a flight test. IR terminal aimpoint algorithms have been developed and demonstrated in the lab (see Figure 15). These algorithms rely on template matching between a stored reference image and the sensed image.



Basic Target Graphic (BTG)

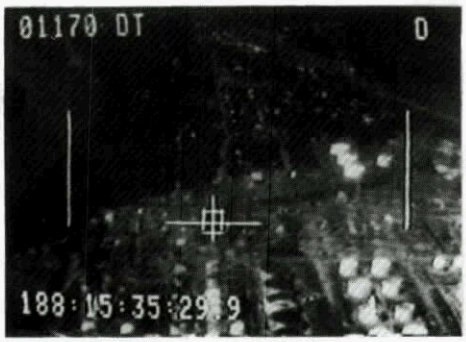


User Selected Features

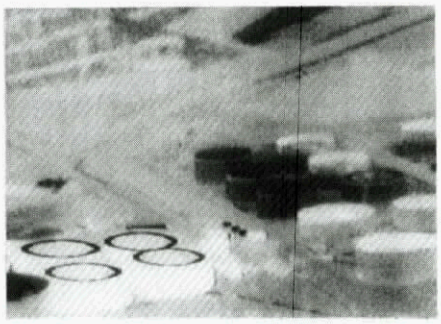


Scene Description File

Figure 14. Target Area Mission Planning



MMW Target Acquisition



IR Target Acquisition

Figure 15. Fixed High Value Target Seeker Application of a Dual Mode Seeker

Relocatable targets such as mobile missile launchers present a different set of problems with their ability to fire and move. Precise location is virtually impossible to determine even after the missile is fired. Their ability to quickly move after firing requires more assets to be used to attack the launch system. The problem is further complicated by their single target nature thus negating the ability to use target clustering techniques for target acquisition. Another complicating aspect of mobile missile launchers is their ability to appear as dim targets, i.e., low observable, countermeasured, camouflaged, and usually in deep hide, thus requiring highly complex sensors, just to acquire them. The ever-increasing availability and capability of these weapons will require a large number of assets to be dedicated or diverted to this mission in order to neutralize its effectiveness. As seen in Desert Storm, the effectiveness of this approach is limited. Hence, a "Scud Buster" must be accurate, countermeasure resistant, all weather, and autonomous while maintaining low cost. Multi-sensor seekers provide a cost effective solution in this case as well—however, the problem is more complex than that of hitting fixed high value targets.

A variety of seeker technologies have been investigated for the mobile target seeker as well. Table 4 outlines their pros and cons. Once again, none of these technologies provide an acceptable solution.

As an alternative to the above technologies, dual mode MMW/IR provides an attractive option. The complementary synergy of MMW and IR balance out each other's weaknesses to provide a total performance capability. By fusing MMW and IR information, their synergy can be exploited to obtain quantum improvements in the performance needed to meet the mobile target seeker requirements [4]. Figure 16 shows actual field test results obtained by Alliant Techsystems (under Army programs) to quantify the benefits of dual-mode MMW/IR synergy over MMW-alone or IR-alone. *It is this synergy which holds the key to the detection of cold, stationary targets, and sparsely located SSM/TEL targets in the mobile target strike mission.* Submunitions such as the U.S. Army's BAT capitalize on this synergy to target elusive SSM/TELs.

The reliable recognition and identification of tactical targets will greatly benefit from the emergence of model based multi-sensor ATR algorithms. One such promising approach is the Integrated Multiresolution Algorithm (IMA) paradigm. The IMA paradigm is contrasted with the classical ATR paradigm in Figure 17. The IMA paradigm is based on the premise (modeled after the human cognitive process) that the ATR function should proceed on a coarse-to-fine basis, with multiple hypotheses (targets, aspects, and depressions) considered at early stages of the process, leading to specific hypotheses at the later stages of the process. This is an integrated ATR approach which, in a single, seamless paradigm, provides detection, recognition and identification as a by-product at successive levels of refinement in a multi-resolution tree. Unlike the classical ATR paradigm, a top-down approach such as IMA completely eliminates the successive information reduction steps, resulting in a two-fold benefit. First, it prevents propagation of erroneous decisions made in earlier stages. Second, it avoids scenario dependence by eliminating information reduction steps which are usually tied-in to local statistics.

6. CONCLUSIONS

Multi-sensor systems will have ubiquitous applications in surveillance systems and smart weapon precision seekers. Multi-sensor systems provide adverse weather operational capability and countermeasure resistance. Multi-sensor systems provide low cost solutions. It takes the right system engineering approach to balance the complementary capabilities of sensors in attaining a low cost solution. A family of multi-sensor systems can be engineered to provide the right cost versus performance operating point for a variety of missions.

7. ACKNOWLEDGMENT

Parts of the R&D work reported here were supported by contracts from the U.S. Army and this support is gratefully acknowledged. The author wishes to thank numerous colleagues responsible for the technical results reported here. Ms. Sue Geronsin expertly prepared the camera-ready version printed here.

Table 4. Mobile Target Seeker Candidates and Their Trades

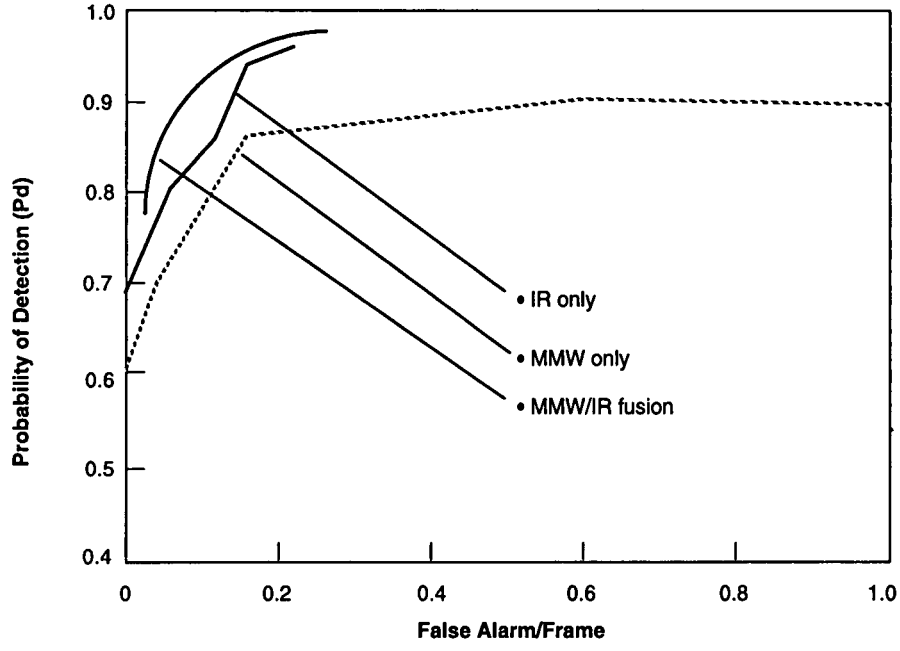
Seeker Candidate Technology	Advantages	Disadvantages
MMW	• Mature technology • Wide area coverage • Adverse weather capability • Cold target detection capability	• Unacceptably high false alarms in wide area search • CM susceptibility (especially for high value targets) • Inadequate target classification/ aimpoint capability
IR	• Mature technology • Good target classification/aimpoint capability	• Adverse weather vulnerability • CM susceptibility (especially for high value targets)
LADAR	• Good target classification/aimpoint capability	• Technology not mature • High DTUPC cost • Adverse weather vulnerability • Inability to search wide areas in mission timelines

T16326_09

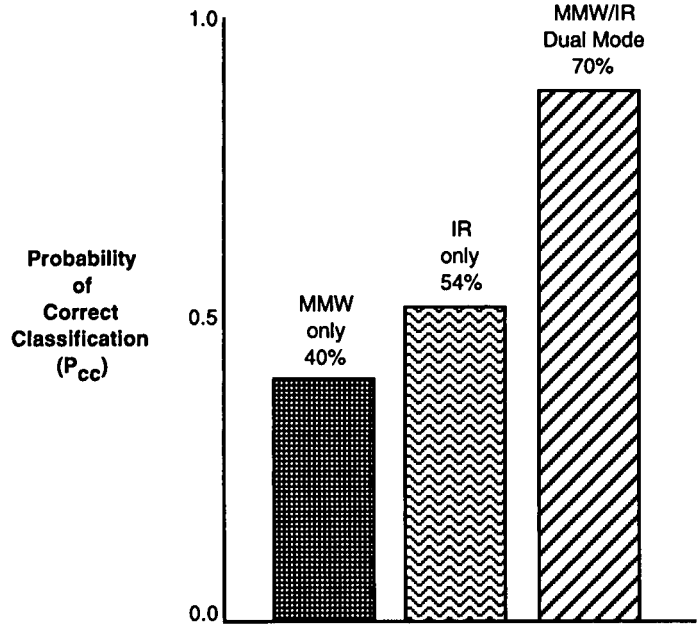
Mobile Target Seeker

Dual-Mode MMW/IR

- MMW/IR fusion synergy provides
  - Wide area search with reliable clutter rejection
  - Cold/stationary target detection capability
  - CM resistance, particularly needed for time critical targets
  - Accurate target classification
- MMW provides adverse weather capability
- IR provides precise aimpoint and tracking
- Dual mode provides graceful performance degradation when one sensor is rendered ineffective



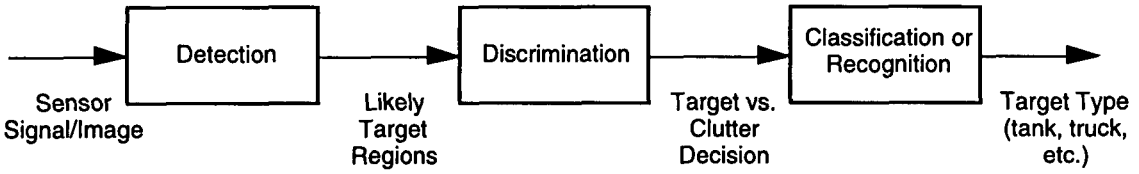
(a) MMW/IR fusion synergy provides high probability of detection with low false alarm rate in wide area search.



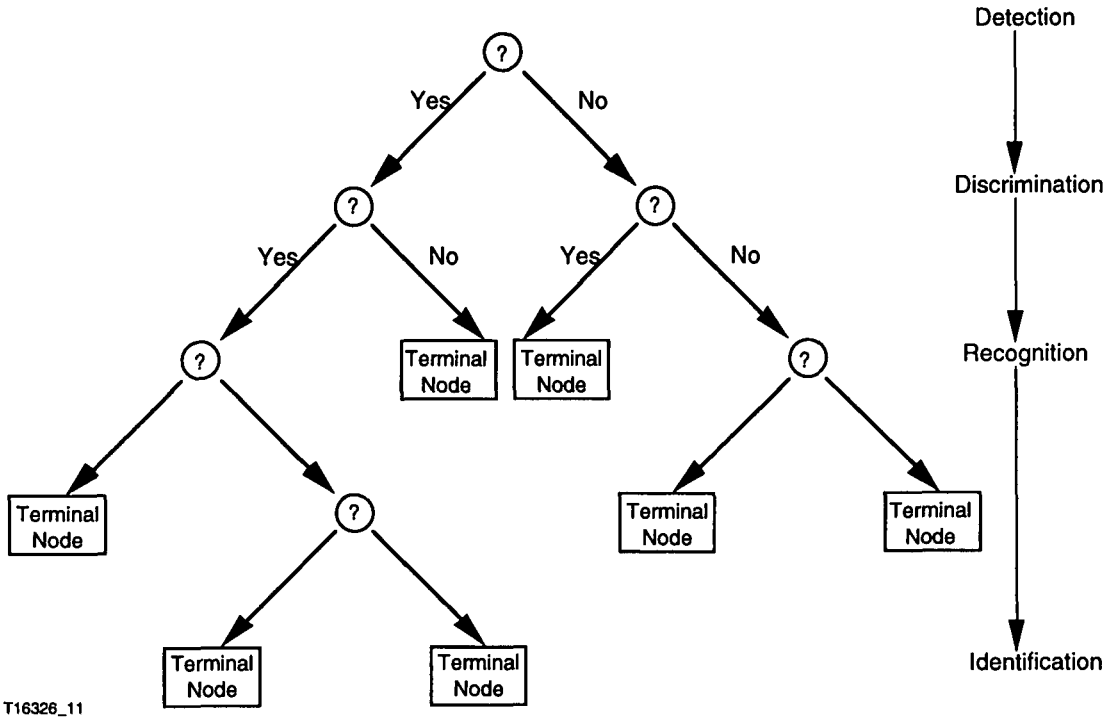
(b) MMW/IR fusion synergy enhances target classification.

T16326_10

Figure 16. Dual Mode MMW/IR Synergy Holds the Key to the Detection of Cold, Stationary, and Time Critical Targets



Classical ATR Paradigm



Integrated Multi-Resolution Algorithm (IMA) Paradigm

Figure 17. ATR Algorithm Paradigms

8. REFERENCES

1. B.R. Suresh and J. Carter, "Cost Effective Dual-Mode Seeker and Its Application," Precision Munitions, Signatures, and Simulators Conference, May 1992, Eglin Air Force Base, Florida.

2. M.A. Gallo, et al, "Low Cost Uncooled IR Sensor for Battlefield Surveillance," SPIE Technical Conference 2020 on IR Technology XIX, July 1993, San Diego, California.

3. R.A. Belt, et al, "The Aladdin Processor — A Miniaturized Target Recognition Processor with Multi-GFLOP Throughput," Proceedings of the SPIE Conference on Architecture, Hardware, and FLIR Issues in ATR (Vol. 1957), April 1993, Orlando, Florida.

4. K. Siejko, et al, "Dual Mode Sensor Fusion Performance Optimization," 6th National Symposium on Sensor Fusion, April 1993, Orlando, Florida.

Sensor Data Fusion for Air to Air Situation Awareness Beyond Visual Range

C. A. Noonan
British Aerospace Defence,
Military Aircraft Division,
Warton Aerodrome (W392D),
PRESTON PR4 1AX,
U.K.

SUMMARY

The modern air superiority aircraft is faced with ever increasing threats and more capable targets. To be effective against them, it needs to know what and where they are as early as possible during any encounter. It must do this during day and night, in all weathers, in hostile counter measure environments and in the presence of clutter.

The aircraft will receive large amounts of information from multiple sensors and data communications systems. If the information is to be used effectively to the benefit of the mission, it must be aligned, correlated, consolidated and presented to the crew in a meaningful form.

A model is offered of tactical situation awareness processing as it might be embedded in a future avionics system. It shows Sensor Data Fusion in relation to sensors and communications, to situation and threat assessment and to sensor and mission management. It shows the flow of data around the sub-system as it creates and maintains the tactical situation database and ranks the information therein in order of importance to the mission.

The requirements placed on sensor data fusion by air superiority operations are discussed. These are dominated by particular features of the tactical aircraft platform and its mission.

A computer test harness, developed by BAe Defence, Military Aircraft Division, is described, along with built in tools which calculate test statistics. The harness was developed as part of a programme of studies carried out by the Mission Systems group of the Product R&D team within the Systems Engineering Department at BAe's Warton unit in Lancashire, U.K. Examples of the results which were obtained when an air to air sensor data fusion model was evaluated are reproduced.

1 INTRODUCTION

Since the early days of air warfare a pilot's ability to maintain an awareness of the tactical situation has been a dominating factor in the outcome of air combat engagements. Until relatively recently this ability was a "sixth sense", possessed by some (the aces) and not by others. Modern multiple-sensor and communications systems have the capability to provide all the data necessary to establish and maintain Tactical Situation Awareness and to extend its scope far beyond the visual range. However, effective use of these sensors can increase the potential to gather information to such an extent that "information overload" becomes a real problem. Even in multiple crew aircraft, the assimilation of data from multiple, disparate sources can become a significant task in a target rich environment if no computer assistance is provided.

The increasing sophistication of the (airborne and ground based) threat, the complexity of the sensor system and a potential requirement for single crew aircraft has lead to an emerging need to provide for pilot assistance through the use of computer aids. In future aircraft, automation of Tactical Situation Awareness processing will reduce workload and enhance performance by providing a single, comprehensive air picture to the aircrew which requires the minimum of human effort to assimilate.

This text is concerned with Tactical Situation Awareness processing for some future air superiority aircraft. In particular, with those processes known as Sensor Data Fusion. Its objective is to identify those problems which are characteristic of, and sometimes unique to, the air superiority role.

2 TACTICAL SITUATION AWARENESS PROCESSING

Figure 1 shows a model for Tactical Situation Awareness processing. It comprises sensors, data fusion, situation assessment and sensor management.

The system shown in Figure 1 is a generic one. It might be said to represent many classes of military and civil Situation Awareness processing systems. However, it is an adequate model on which to base an overview of the air superiority case and to outline the environment in which Sensor Data Fusion will operate in the future.

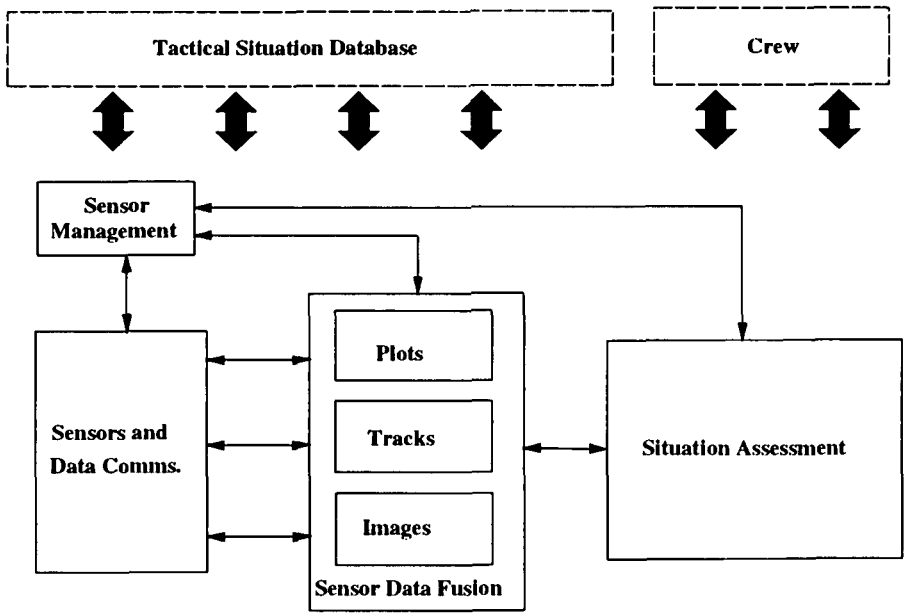


Fig. 1
Tactical Situation Awareness

2.1 Tactical Situation Database

The Tactical Situation Database is an on-line store of information. Some of it will be long-term and slowly-changing, such as fixed ground threats or the expected enemy force composition. It is sufficient to manipulate and update such information at a ground installation and provide it to the aircraft in the form of a data-cartridge at the start of each mission. The Sensor Data Fusion process would make use of such data but would rarely update or amend it.

Another class of information may be described as medium-term and infrequently updated, such as long range threats, targets and friendly forces. They would be beyond the coverage of the aircraft's sensors or, if it is conducting a mission as part of an operational group, beyond the group's sensor coverage. Such information would be received periodically from a Command and Control centre via the data communications network. The Sensor Data Fusion process would correlate and align this information, in time and to the local spatial datum, and ensure that the database is always fully up to date. It may use this data, via Sensor Management, to cue the sensor system when first detection becomes possible.

The final class of information in the database may be described as short-term and on-line, such as targets and threats which lie within sensors and weapons range. Often they will be visible to multiple sensors simultaneously and, in group operations, to several group members simultaneously. The alignment, correlation and consolidation of this data forms the prime task of the Sensor Data Fusion Process. In a target rich environment, it can be highly processor intensive.

In addition to these types of update, the crew may intervene to change the Tactical Situation Database at any time. Eg in response new information received via voice communications.

2.2 Situation Assessment

Situation assessment is concerned with priorities and resources. Priorities may stem from several sources. The most important of these are threat (what poses the greatest danger to me?), target (what are my best opportunities to attack?) and crew intervention (what does the crew want to do next?). By quantifying these, the Tactical Situation Awareness system can adapt to provide the highest quality information where it is needed most.

Threat prioritisation can be based on track parameters provided by the Sensor Data Fusion Process such as predicted closest point of approach, time to closest point of approach, inferred threat type etc. Target prioritisation may be based on firing opportunity calculations and predicted kill probabilities. Threat or target prioritisation may be overridden by the crew at any time and in the case of target prioritisation crew selections will be the norm.

Resources are limited on a small aircraft. Availability of fuel, armaments, processing capability, etc. all impinge on the its ability to follow a particular course of action. Situation

assessment will monitor resources against the current Tactical Situation and indicate to the crew when critical resource levels are reached.

2.3 Sensor Management

The Sensor Management process completes the Tactical Situation Awareness loop. It is concerned with controlling the behaviour of the sensors to adapt it to the current Tactical Situation. The extent to which this is possible depends on the sensors. Some are intrinsically controllable and others are relatively passive and fixed in their behaviour. In general, an air superiority aircraft will carry a mix of these types.

Sensor Management will control the search behaviour of the sensors. Its objective will be to maximise the probability that threats and targets are detected, acquired, tracked and identified as soon as it is physically possible to do so. It will do this by adapting the sensors' response to information available from external sources. Depending on the quality of the external information it may initiate wide volume or narrow searches (a process known as cuing) or direct track initiation, for one or more sensors, using external data to initialise Sensor Data Fusion (a process known as priming).

Once threats and targets have been detected and acquired, Sensor Management tries to ensure that prescribed levels of information quality are achieved, and then maintained, in the tactical situation database. If the information, regarding a particular target or threat, is found to be inadequate in any way, sensor resources may be concentrated in its region. Conversely, if the information exceeds required quality levels, sensor resources may be freed for use elsewhere.

All of the above actions will be subject to crew intervention, indirectly via the priority list and directly via cockpit sensing controls.

3 THE IMPACT OF THE AIR SUPERIORITY ROLE ON SENSOR DATA FUSION

The requirements placed on sensor data fusion by the air superiority role are dominated by particular features of the aircraft platform and its mission. For many surface based and large platform based sensing tasks, the numbers, types and spatial distributions of sensors may be chosen to deliver constant high probabilities of detection and tracking over a given region. On an air superiority aircraft, this is not the case. The aircraft is designed to optimise vehicle performance features and the sensor fit is strictly limited by constraints of space, weight, power consumption etc. Thus, an approach to data fusion is required which is tailored to the sensor fit and the likely sensor deployment strategies.

The following sections consider the limitations imposed by the constrained sensor fit, the nature of the targets and the short duration of airborne engagements.

3.1 The Impact of Sensor Fit

The aircraft is equipped with a sensor suite to provide targetting data to its weapons system, to provide early warning of attack and to provide situation awareness to the crew beyond the visual range. Usually, the prime sensor will be an air to air RADAR. This will provide accurate range and radial velocity measurements and fair sight line measurements. It may have additional modes for cooperative and non-cooperative target identification. Data from the RADAR may be used for targetting medium and short range missile systems and for controlling gunfire.

The RADAR may be supported by an infrared sensor referred to as Infrared Search and Tracking (IRST). An IRST will provide accurate sight line and, sometimes, angular velocity measurements. It will not provide directly measured range or radial velocity. It may have modes for non-cooperative target identification, but these will usually require crew intervention and may only be practical on multiple crew aircraft. Data from the IRST may be used for targetting in place of RADAR data but, unless supported by externally measured range data, would introduce extra uncertainty into the fire control processing.

The third common on-board sensor is the Electronic Support Measures (ESM). This comprises a RADAR Warning Receiver supported by enhanced direction finding. The ESM measures the characteristics of the RADAR emissions made by other aircraft or surface installations. These may be used for non-cooperative target identification. It also provides relatively poor sight line data. The purpose of the ESM is to give early warning of hostile platforms and weapons engagements and to aid the identification of the emitting platform.

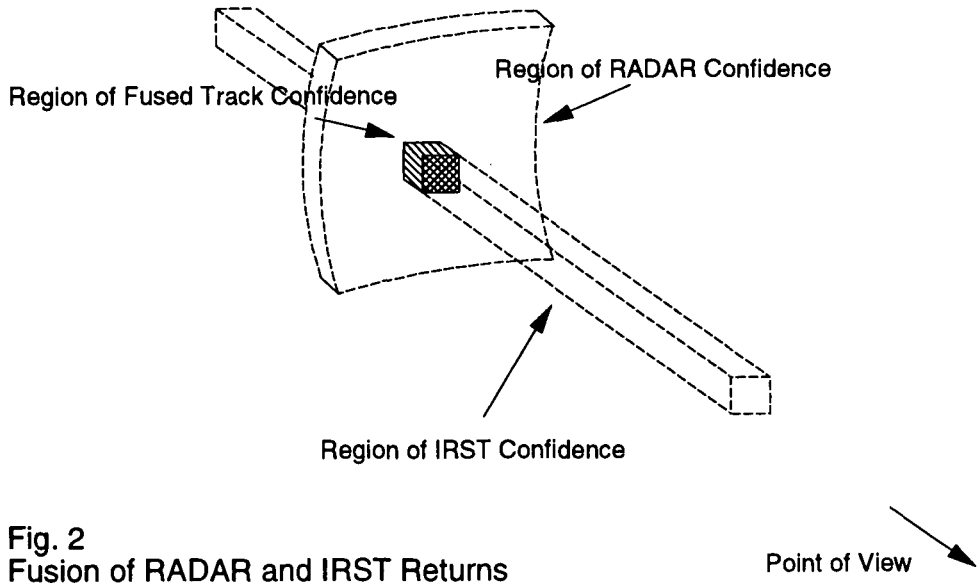


Fig. 2
Fusion of RADAR and IRST Returns

The air superiority platform may carry other sensor types but the above, in conjunction with data communications, form a sufficient basis for examination of the properties of the Sensor Data Fusion process.

Such a sensor suite is referred to as non-commensurate. This means each sensor produces an accurate and reliable estimate of a different characteristic or feature of the target and coarser estimates (or no estimate) of other target characteristics. If the data fusion process can associate different sensor views of each target together correctly, the rewards in track quality and confidence will be great.

Such a system is said to offer high spectral band width. This means that it is observing the targets and threats, simultaneously, in multiple bands of the electromagnetic spectrum. This has the advantages of high probability of detection because the aircraft has more chances to "see" the target. It also produces large synergistic gains as a result of data fusion because lots of different types of information are being gathered. The result is fused tracks of high dimensionality.

An example of this, which is quoted frequently, is the fusion of RADAR and IRST measurements. As discussed above, RADAR contributes an accurate estimate of range and radial velocity and IRST contributes accurate sight line and angular velocity. The effect of data fusion on position estimation may be visualised as shown in Figure 2.

In addition to the great improvement in location accuracy, fusion yields estimates of speed and heading to an accuracy that is not available from either sensor individually. This is because RADAR suffers from uncertain angular velocity estimation and IRST produces no radial velocity. Both are required to estimate speed and heading.

For the purposes of this discussion of the sensor suite, it is sufficient to regard Sensor Data Fusion as comprising two main tasks. They are data consolidation, producing single estimates of target state and identity from multiple inputs, and data association, working out how many targets are present and which sensors are observing them. In later sections, a more detailed breakdown will be introduced. With this type of sensor suite, data consolidation is "easy" and approximate methods produce results which are good approximations to the optimum.

However, data association, in the presence of ambiguity, may be difficult. Optimal methods can produce uncertain results (and approximate methods even worse ones). Data association must be carried out on the basis of the data common to all sources and the quality of the result will be governed by the coarsest information which is used. This is particularly important when fusing data for identification purposes. Reliable identification often requires ESM returns and identification performance may be governed by our ability to do association on the strength of poor ESM sight line data.

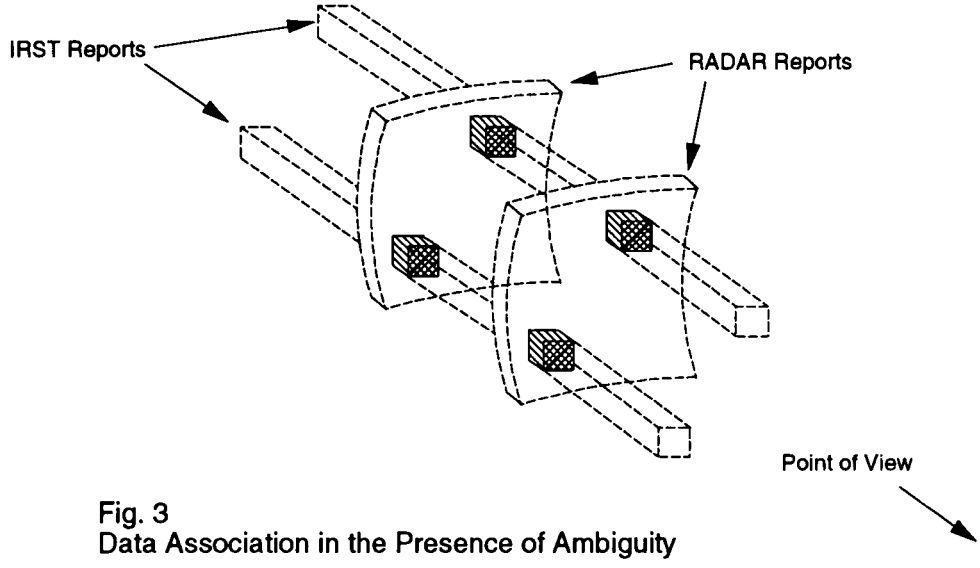


Fig. 3
Data Association in the Presence of Ambiguity

Returning to the RADAR and IRST example, association cannot use the accurate RADAR range and radial velocity information because the IRST produces no corresponding measurements and the accurate sight line and angular velocity measurements from IRST must be compared with the coarser equivalents from the RADAR. This means that the quality of the result is governed by the RADAR angular accuracy. These difficulties may be visualised as shown in Figure 3.

This represents one of the simplest target configurations leading to ambiguity. Two targets, flying close together, are detected and reported by both sensors. The data fusion process must select the appropriate pair of fused tracks from the four candidate tracks. Simple logic indicates that there are only two feasible pairs. Unfortunately, the likelihoods that each pair is the true one will have similar values unless some other information can be used to discriminate between them.

Clearly, a data association algorithm, making optimal use of all the available information, is required in order to resolve matters as quickly as possible. In addition, a strategy of pooling track data from unresolved groups into group reports may reduce uncertainty and deliver a stable solution until full resolution is achieved.

The situation is complicated further by the variations in coverage and sensing characteristics of the various data sources. The example in Figure 3 was assumed to arise in a region where RADAR and IRST were operating at similar, high levels of detection probability. When dealing with multiple data sources, such a situation would be relatively unusual and unresolved groups would frequently occur in which each data source reports different numbers of targets.

Sensing characteristics may also vary with environmental conditions, with target type and target behaviour. IRST detection may be affected by the weather or by target aspect and the ESM performance is dependant on the sensor management strategy of the target (the more emissions it makes, the easier it is to detect and identify). Furthermore, the fusion platform may allocate different sensors to the task of maintaining the information quality for any track or group in response to changes in the situation.

Uncertainties, such as these, lead to a requirement for flexibility in the Sensor Data Fusion process. In addition, it must be highly sophisticated in its approach to data association.

3.2 The Impact of Target Behaviour

Airborne targets may be highly manoeuvrable (fighters in particular) and it is often necessary to dedicate the on board sensors (RADAR and IRST) to the maintenance of high quality, frequently updated, tracks on a small number of high priority targets in order to minimise the probability of track loss. This leads to the concentration of sensor resources in a small region and the potential loss of wider situation awareness and surveillance. This is represented in Figure 4.

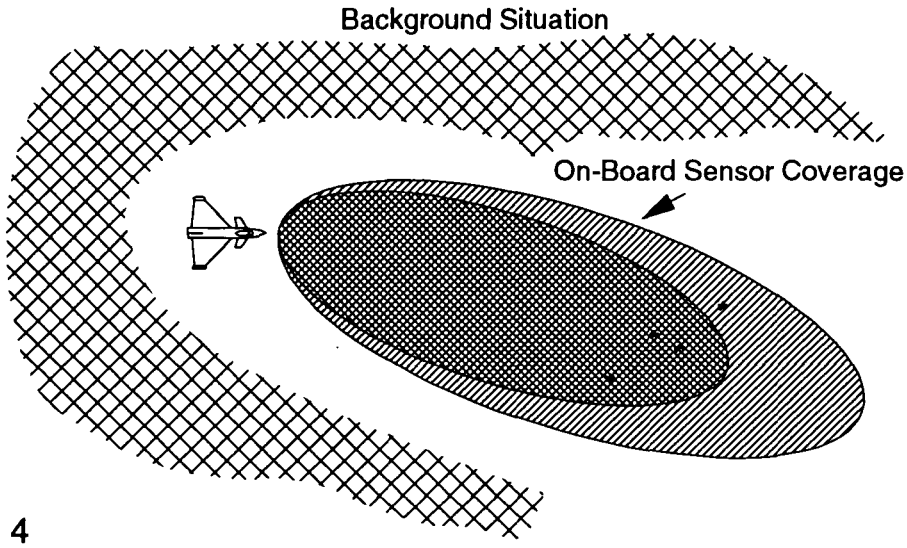


Fig. 4
Sensor Deployment for Priority Tracks

While this is going on remotely-sensed, communicated data will provide background situation awareness, in conjunction with the ESM, to satisfy the platform's needs for surveillance and early warning.

3.3 The Impact of "Time Lines"

The aircraft and its target are capable of manoeuvring at high speed for short periods. For this reason, the periods during which the aircraft is actively prosecuting engagements or evading threats (the time lines) are relatively short in air to air combat. During these periods the system must track and identify targets and threats quickly and with high levels of confidence in order to deploy weapons and defences in a timely manner. This may have a big impact on sensor management and on the strategies and algorithms employed by data fusion.

In systems where time lines are long, ambiguities of the type represented in Fig 3 may be resolved over relatively long periods of time as more and often better-resolved sensor reports become available. This luxury is usually not available in the air superiority environment where a fast response is demanded. In the future, such a response may be provided by intelligent sensor management seeking, actively, the necessary extra information and concentrating the sensor resource in the regions of greatest ambiguity.

In addition, with the emergence of covert operational concepts and low-observable airframe technology, the time lines are likely to shorten further. The covert concept is one in which, through widespread use of passive sensors and controlled inter-platform cooperation and data exchange, detectable active sensor emissions are minimised. This has the effect of reducing the amount of information available to the ESM which in turn reduces the overall probability of detection and identification confidence. Low observable airframe technology reduces the apparent RADAR cross section resulting in a reduced detection and tracking capability from the RADAR.

On the one hand, there are advantages to our aircraft if it employs covert sensing. It will be more likely to surprise its target and the target will be less likely to employ adequate defences. The disadvantage is a reduction in the opportunities to detect and track the target. On the other hand, there are advantages for the enemy in covert operation and any given sensor suite, employed by our aircraft, will take longer to track and identify threats and targets. The response of the system designer must be to use all the information available to the utmost.

4 IMPACT ON SYSTEM DESIGN

All of the above factors influence the design of Sensor Data Fusion systems for air superiority aircraft. This section will consider some architectural options associated with sensor data fusion and attempt to re-evaluate them in the context of the specific needs and characteristics of air to air operations.

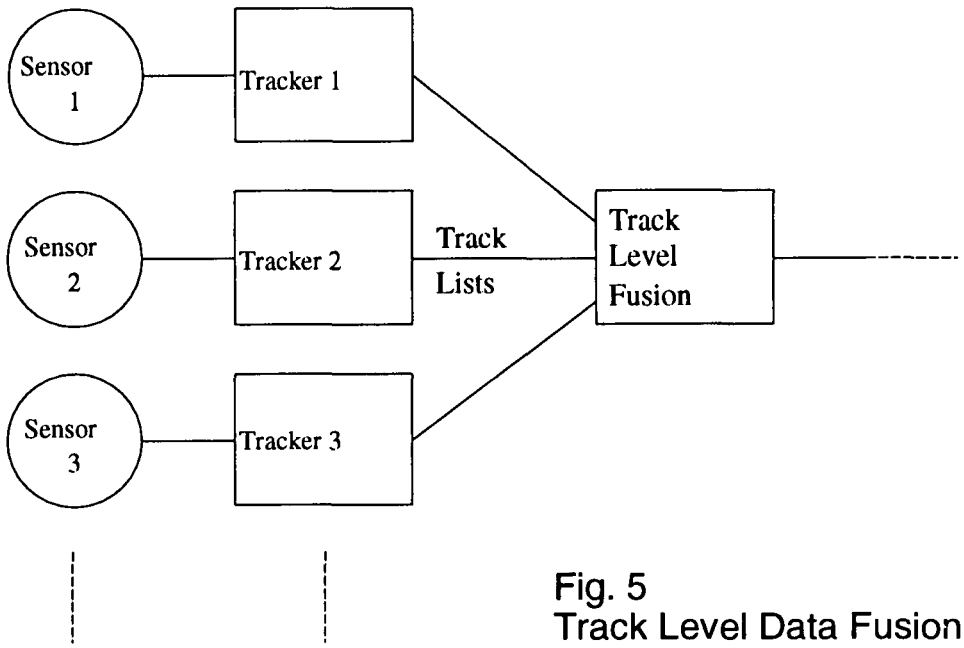


Fig. 5
Track Level Data Fusion

4.1 Architectural Options

In the field of Sensor Data Fusion, architecture refers to the strategy used to organise the data sources, processes and data interconnections into a coherent system.

If technology were capable of producing a single sensor, to provide all the information required by the aircraft with all round coverage, all weather, day and night operation and not prone to countermeasures, there would be no need to do any Data Fusion and architecture would not be an issue. Given that no such sensor exists, multiple data sources must be accommodated. Similar arguments can be applied to the desirability of one or several fusion processes. If it were practicable to gather all sensor measurements, from on-board and external sensors, together at a single fusion process and guarantee not to exceed the capacity of the host computer or the data interconnections, this would probably offer the most attractive architectural option irrespective of the application. However, system designs must accommodate existing networks and infrastructures and this usually means including a capability to use data which has been pre-processed into tracks prior to arriving at our aircraft.

For these reasons data fusion architectures raise important issues and they are discussed in the following sections. In general, the discussion relates to the fusion of kinematic state data (location and motion). Fusion of identity data is relatively insensitive to architecture differences provided that the data does not undergo severe truncation as a result of transmission between processes.

Three common architectural models for sensor data fusion are discussed in many books on the subject. These include References [1], [2] and [3]. The models are "track level", "plot level" and "combined/hybrid" fusion. In addition to these, a fourth option known as "information filter" fusion is considered. The information filter is discussed briefly in reference [5] and the way it may be used to distribute the data fusion process is discussed in reference [4].

4.1.1 Track Level Fusion

The first architectural model to be considered is track level fusion, represented in Figure 5. It is also commonly referred to as the "sensor level tracking model" and the "autonomous model". It operates at the highest level of all and all data is assumed to arrive in the form of tracks which have been pre-processed in a dedicated tracker at source. Whilst communication between sensors, to assist tracking, is possible, each track is formed largely on the basis of the detections made by a single sensor. Within the Tactical Situation Awareness processing model used here, the main source of feedback to the sensors is via Sensor Management.

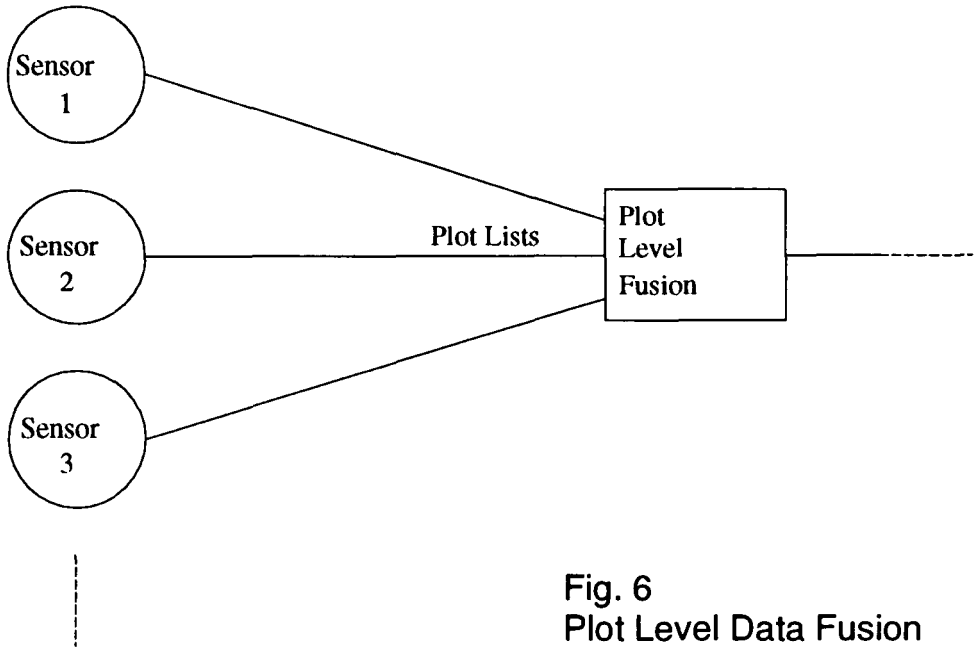


Fig. 6
Plot Level Data Fusion

This approach is recognised as having the advantages of:

- o relatively light bus data traffic,
- o low processing density as a result of simple parallelism,
- o good survivability as a result of its distributed nature,
- o allowing each tracker to be tailored to maximise the contribution of its sensor,
- o a data fusion process independent of sensor specific features.

It has the disadvantages of:

- o requiring special processing to allow for error correlations,
- o producing less accurate/continuous tracks due to its non-optimal approach.

Track level fusion is flexible due to its distributed nature and because it communicates target data at a relatively high level. This has advantages for the system integrator and the system maintainer. The Sensor Data Fusion process can be kept independent of sensor specific features and variants of a system, with sensor fit changes, should be easier to create than with other architectural models. Furthermore, air to air sensors tend to be highly specialised and incorporate specialised track processors. This architecture accommodates them well.

The less accurate tracks and any error correlations, not accounted for in the processing, lead to reductions in tracking performance. However, due to the non-commensurate nature of the sensor suite used by the air superiority aircraft (discussed in section 3.1), this approach will produce good approximations to optimal performance in this case. This is not generally true of less specialised applications of data fusion.

Non-optimality may be an important issue when considering system validation and product liability. It may be necessary to predict, accurately, when and how the performance of such a system will differ from an optimal one.

4.1.2 Plot Level Fusion

The second architectural model to be considered is Plot Level Fusion, represented in Figure 6. It may also be referred to as the "central level tracking model". It shows the sensors producing streams of plots or individual target measurements which are processed by a single tracking filter which produces the situation database directly.

This approach is recognised as having the advantages of:

- o more accurate/continuous tracks due to its optimal approach,
- o freedom from error correlation problems,
- o being more amenable to sophisticated data association techniques.

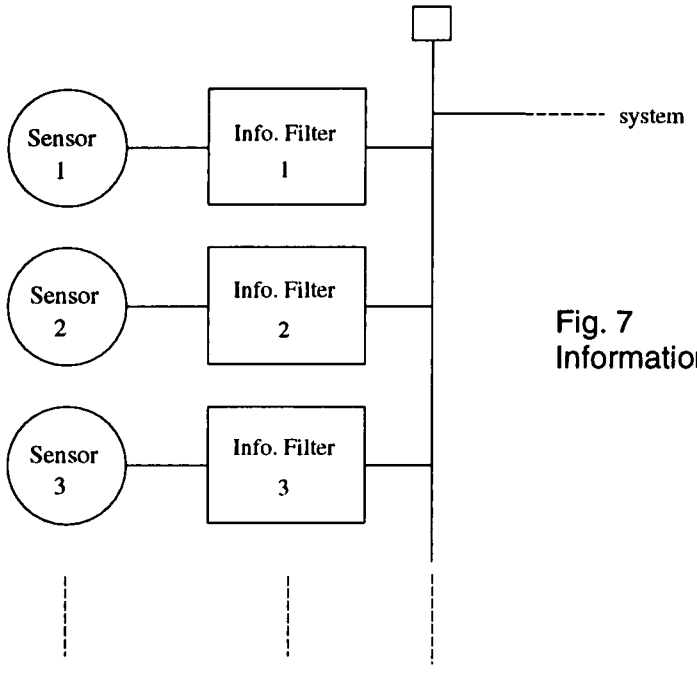


Fig. 7
Information Filter Data Fusion

It has the disadvantages of:

- o susceptibility to sensor degradation,
- o dependence of the Sensor Data Fusion process on low-level features of the sensors.

Plot level fusion produces more accurate, more continuous tracks. This has advantages where data association is difficult because the more accurate tracks result in smaller association gate sizes. This is good because, the smaller the gate is, the easier it is to find the track when the next update is due.

This is particularly significant when tracking manoeuvrable airborne targets at short range, where the size of the data association gates may become significant in relation to the sensors' instantaneous fields of view. However, at longer ranges this effect will be much smaller.

The amenability to sophisticated data association techniques (e.g. multiple hypothesis approaches) is a big advantage of the plot level fusion approach. The fact that the non-commensurate sensor suite is prone to the type of ambiguities discussed in section 3.1 makes data association particularly important.

Unlike the track level model, all of the processing is concentrated in a single processing centre. The processing demands of data association increase rapidly (factorially) as the number of targets and sensors increases and, for some implementations, can generate unsupportable processing requirements when faced with a target rich scenario. Thus, the full benefits of sophisticated data association might not always be realised. Careful analysis of these factors, on a case by case basis, is advisable.

Optimality may offer important advantages when considering system validation and product liability. Again, processing overload can degrade system performance and is worthy of consideration in any such analysis.

The disadvantage of plot level fusion is its susceptibility to sensor degradation. If the performance of any individual sensor is degraded, the situation database will be prone to degradation. Provided sensor degradation can be detected, other architectural models allow degraded data to be separated, more easily, from unaffected data.

4.1.3 Combined/Hybrid Fusion

The third common architectural model is the combined fusion model which performs plot and track level fusion in parallel. It may also be referred to as the "hybrid model".

This architectural model has none of the disadvantages described above and all of the performance related advantages. However, the costs in bus data traffic and processing can be high and some method must be devised to choose between the two solutions when they differ.

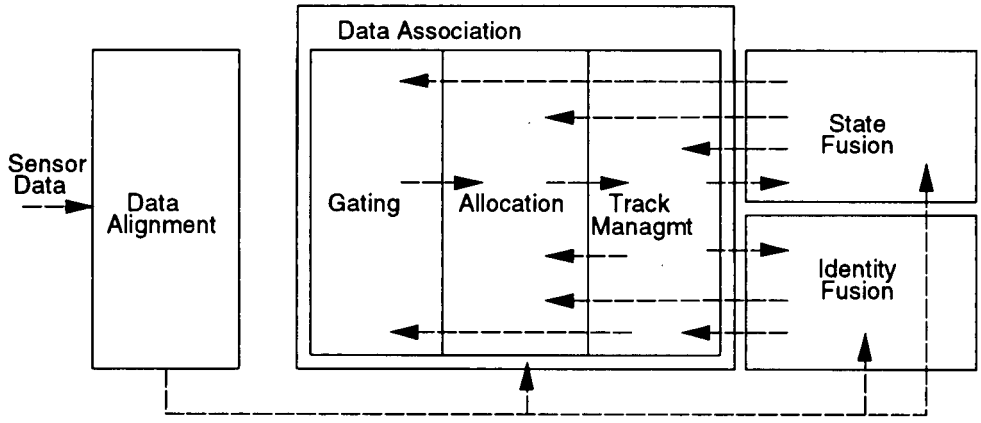


Fig. 8
Data Fusion Information Flow

The great advantage of this approach is that data can be directed towards the most appropriate fusion algorithm on the basis of an adaptive decision process. Thus, no data, suitable for optimal processing, need be processed sub-optimally and when optimal treatment of a highly ambiguous data association problem would lead to processing load problems and truncation of the algorithm, a quicker approximate approach will be available. For these reasons, a pragmatic approach which selects and mixes fusion algorithms on an opportunistic basis may offer the most attractive implementation options for future Situation Awareness processing systems.

4.1.4 Information Filter Fusion

There exists a fourth architectural option which is worthy of evaluation. It has been demonstrated in the field of robotics. Instead of a conventional tracking filter, each sensor processor runs an information filter as represented in Figure 7.

It is not intended to explain and justify the information filter here because it would take up too much space. Reference [4] offers a full and clear definition. However, the properties of the information filter approach are worthy of consideration.

Information filter fusion has the properties of algebraic equivalence to a Kalman filter based, plot level fusion process and physical distribution similar to a track level fusion process. It is said to require greater computing and communications resources than track level or plot level fusion and it introduces processing redundancy into the system. What it offers in return is distribution of the fusion task without sub-optimality and the ability to reconfigure the sensor suite, on-line, without loss of data or interruption to track continuity.

The information filter form is also capable of dealing with identity fusion and sensor management. These are discussed in references [6] and [7]. It would be fully feasible to connect an information filter to a higher level hybrid system.

4.2 Data Fusion Processing

The previous sections dealt with the processing environment external to Sensor Data Fusion, with the nature of the sensor suite and with the way sensors, fusion processes and interconnections may be organised. This section deals with the internal organisation of a data fusion process. It concentrates on properties rather than on algorithmic details. It considers a data fusion process comprising six sub-processes: data alignment, gating, allocation, track management, state fusion and identity fusion. Gating, allocation and track management collectively perform the data association task.

The process is represented in Figure 8, along with the most important internal information flows.

4.2.1 The Fusion Sub-Processes

- o <u>Alignment</u> resolves time, reference frame and point of view differences in the incoming sensor returns.
- o <u>Data association</u> performs two interdependent functions. Namely, estimation of the number of targets currently "visible" and allocation of sensor data to the target tracks.

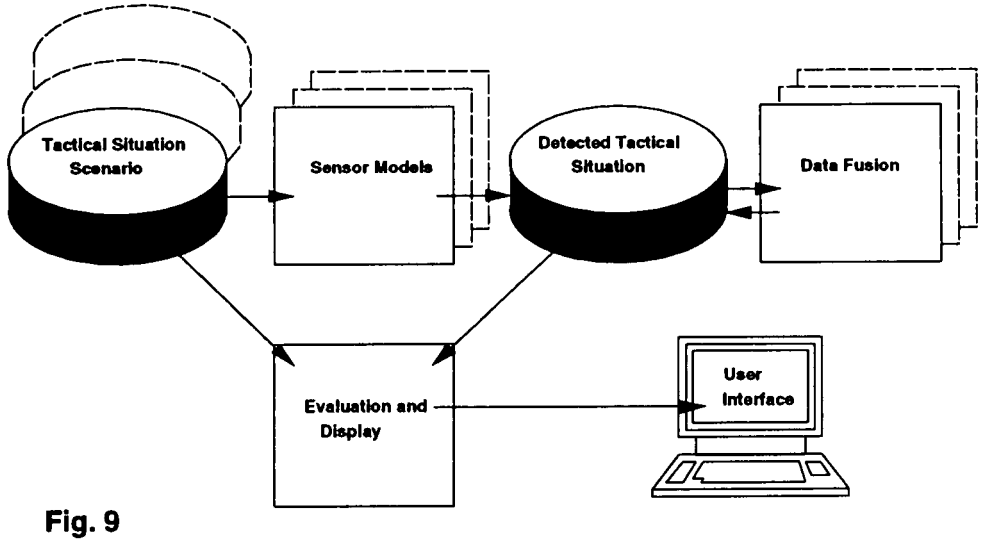


Fig. 9
The Tactical Sensor Data Fusion Test Harness

- o Gating is a data reduction filter which excludes highly unlikely combinations from the data association process. Small gates lead to missed (true) allocations and large gates to false allocations. In general, the gates are chosen to minimise the sum of missed and false allocations.
- o Allocation groups sensor returns with target tracks. Common approaches optimise some separation or likelihood metric. Allocations may be one to one (nearest neighbour) or many to one in a probabilistic weighted mean.
- o Track management initiates new target tracks, maintains existing ones and removes discontinued ones. It may provide "history" data for incorporation in the allocation metric (Eg by some multiple hypothesis approach).
- o State fusion consolidates target kinematic data.
- o Identity fusion consolidates target identity data.

4.2.2 Processing Performance Issues

In general, each new sensor update will encounter the functions of the data fusion processing model in sequence from left to right on Figure 8. Clearly, inaccuracies introduced during the early stages have the potential to degrade the subsequent processes.

This can manifest itself in several ways. For example, inaccurate data alignment can cause poor data association performance which can in turn cause data from several targets to be fused into a single, ghost track. This ghost track then degrades the next round of data association and a vicious circle begins.

On the other hand, whilst approximate state fusion based on correctly associated data will degrade the other functions somewhat, when the approximation is a good one the effects will be insignificant.

5 TEST HARNESS

To evaluate Sensor Data Fusion systems, the Mission Systems R&D Group at BAe Defence, Military Aircraft Division developed a computer test harness. The Data Fusion Test Harness was developed over the period from 1990 to 1992. Its main purpose is to evaluate systems for tactical aircraft but it is not limited to those applications. It is in current use, contributing to evaluations of systems for R&D purposes and for aircraft projects. The test harness concept is represented in figure 9.

It is written in C, and runs under CSTools on a Meiko transputer system with an embedded SPARC processor board. The SPARC communicates to a network of Sun work stations which host the control panels and some of the output displays.

The main outputs are a display showing various plan views of the tactical situation driven by the transputer graphics system and graphs, tables and text displays hosted by the work stations via the X windows system.

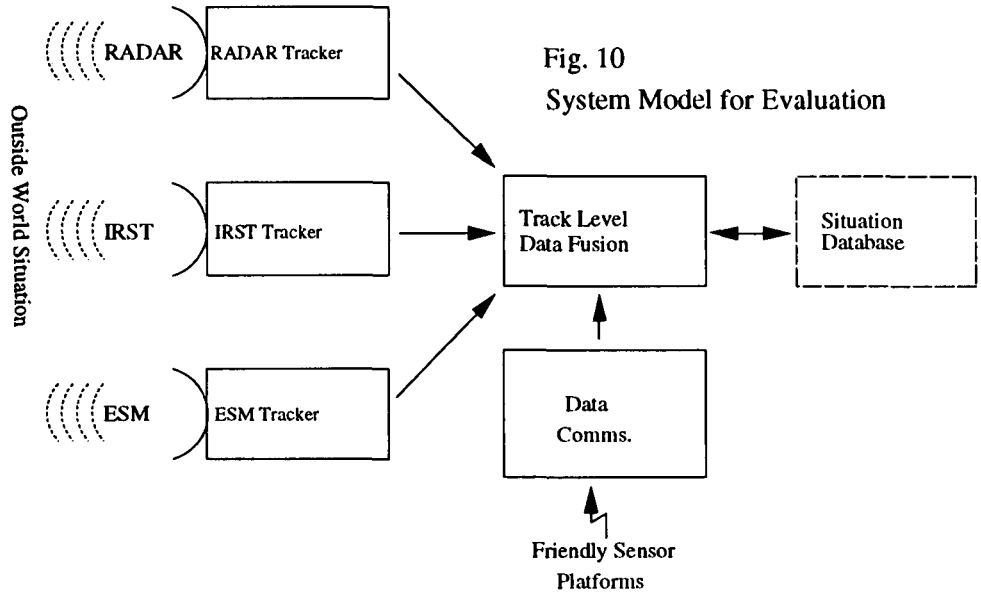


Fig. 10
System Model for Evaluation

The scenarios are interchangeable and are prepared off-line. They can range from simple m-on-n air to air encounters to full battle scenarios with formal Combat Air Patrols, supported by Airborne Early Warning, facing mass attacks. Alternatively, a scenario could represent a ground attack aircraft flying through an enemy's surface to air defences. The scenarios are stored as data files and are activated by a time stepping simulation controller.

The sensor models and the way in which they communicate with the fusion processes define the system model (analogous to the architectural models discussed in section 4.1). Figure 10 provides an example system model.

For any given system model, where it is appropriate, several data fusion algorithms will be included in the harness. Usually this excludes alignment but includes different algorithms for the other sub-processes shown in Figure 8.

When the simulation is active, the user can interrogate individual tracks in any of the track lists, via the tactical display, and select various measures of performance (MoP's) of the current fusion algorithm. The MoP may be output in graphical or tabular form.

6 EXAMPLE SYSTEM EVALUATION

It is neither feasible nor intended to reproduce, in this text, a complete set of system evaluation results covering all architecture models and processing options. Rather, examples will be given which demonstrate how such evaluations have been, and continue to be, carried out. In particular it will be shown how some of the characteristics of Sensor Data Fusion, in the air superiority aircraft, were measured for a track level fusion system. The approach taken transfers readily to other system models.

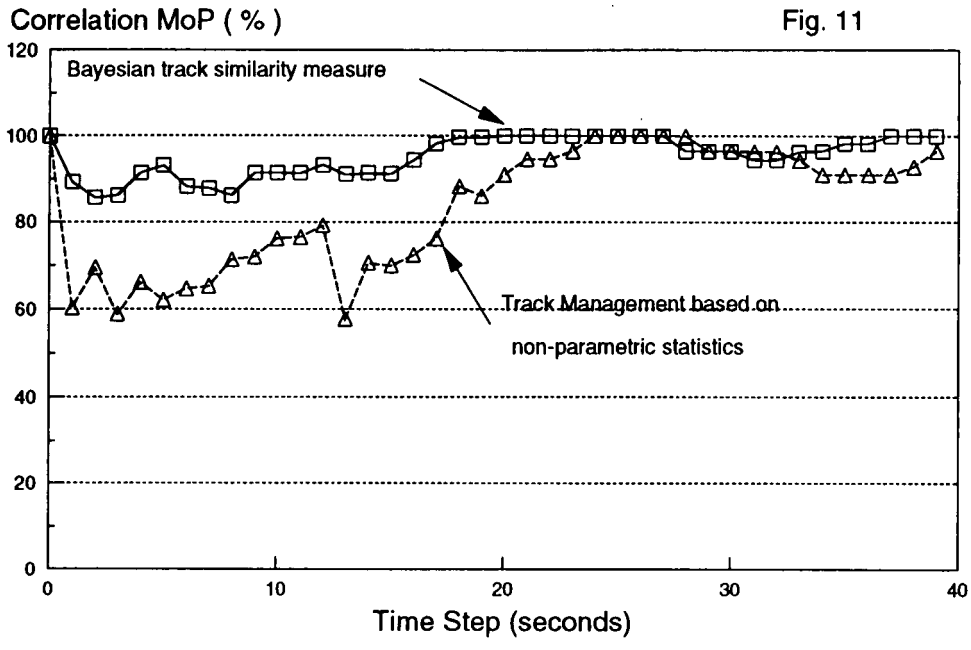
6.1 A Model for Evaluation

The results will be drawn from an evaluation of a computer simulation of a system model with the sensor fit and processing architecture shown in Figure 10. This represents parts of a system for Tactical Situation Awareness in an air superiority aircraft. The scenario and the sensor detections were simulated, the trackers and the fusion algorithms were real.

The on-board sensors comprised RADAR, IRST and ESM. Each sensor had a dedicated tracker. The model incorporated a data communications network which connected it to other similar models which observed the same situation. The local tracks and the communications network tracks were fused in a track level fusion process.

The data communications network comprised simplified sensor models for each friendly aircraft in the scenario, which reported their information in a fixed polling sequence.

Correlation of the network data was assumed to take place externally at a Command and Control centre, whilst correlation of the network data with on-board sensor data was assumed to be a task for the data fusion system under evaluation.



6.2 Example Results

This section will show how some of the properties of Sensor Data Fusion, discussed in the previous sections, were measured and it will reproduce some results. In particular, the following aspects will be addressed:

- o data association performance,
- o state estimation performance.
- o processing load distributions and budgets,

The measurements were made using a four-on-four air to air combat scenario. This began with two four-aircraft formations closing on each other and entering the coverage of the on-board sensors and ended with them breaking formation and manoeuvring to engage in close in combat. The period covered represents the Beyond Visual Range sensing phase of the encounter. This scenario was arranged to maximise the Sensor Data Fusion processing task. All eight aircraft were using active RADAR, IRST and ESM, and the friendly formation were reporting all their on-board sensor information along with their own positions and identities on a regular basis via the communications network.

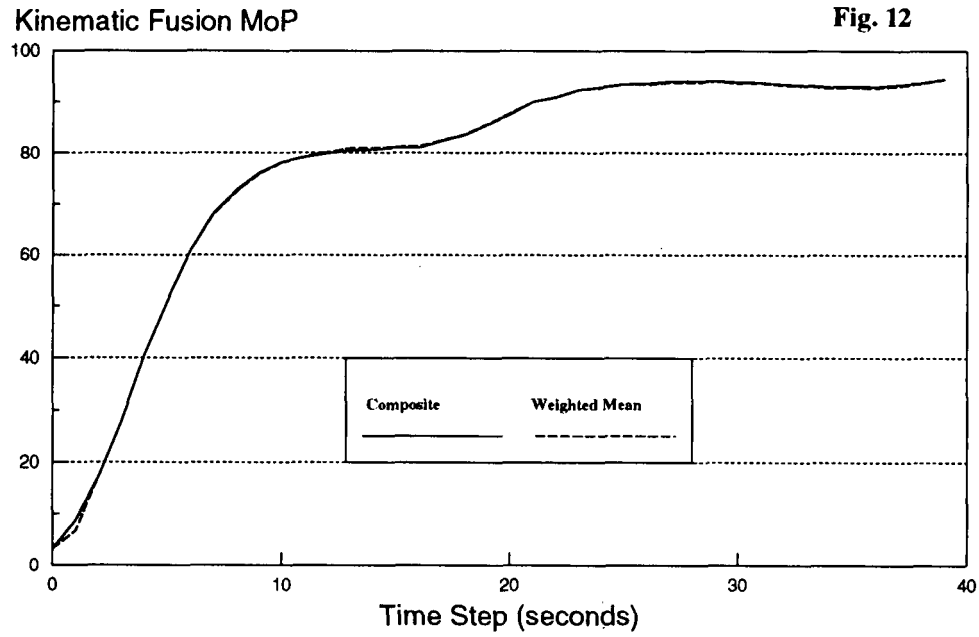
First consider data association performance. It is predicted in section 3.1 that data association performance will be sensitive to the choice of algorithm in the presence of ambiguity. Figure 11 shows the performance of two data association algorithms exposed to identical sensor returns. The data association MoP used here is the proportion of sensor tracks which are allocated to "perfect" fused tracks, ie. fused tracks comprising all data for a single target and no spurious data.

One algorithm generated Bayesian probabilities for all feasible track combination hypotheses, incorporating full association history, and then extracted the maximum likelihood set of allocations.

The other algorithm selected maximum likelihood allocations on the basis of current sensor reports only, and then used a non-parametric statistical ranking process to create, maintain and remove fused target tracks.

The shape of the graph is due to features of the scenario. From 0 to 10 seconds the on-board sensors are acquiring the targets and the amount of information in the system is increasing rapidly. It continues to increase from 10 to 20 seconds when the data communications network starts to exchange the newly acquired data. Then at 30 seconds new ambiguity is introduced when targets appear to cross one in front of the other.

The Bayesian approach is theoretically more attractive than the non-parametric statistical one because it uses the available data more fully. It can be seen that the results bear this out.



In general, the Bayesian algorithm is superior, showing less disturbance from the true situation and faster recovery. However, both algorithms are shown to be capable of fully resolving the scenario, the non-parametric statistical approach is simply slower to converge. Unfortunately, this is a major disadvantage for an air superiority aircraft. Given a less ambiguous scenario, less difference would have been observed between the algorithms. When evaluating systems of this sort, the scenario is important. The fusion task it poses must incorporate difficult situations to fully exercise the algorithms under evaluation.

Next, consider kinematic state estimation. The assertion was made in section 3.1 that it is intrinsically easy with this type of sensor fit. Figure 12 shows the results when the position estimation performance of two algorithms was compared. Here the state estimation MoP is best expressed mathematically:

$$\frac{1}{3}\left(e^{\frac{-\Delta_r^2}{\sigma_{Rr}^2}}+e^{\frac{-\Delta_\theta^2}{\sigma_{R\theta}^2}}+e^{\frac{-\Delta_\psi^2}{\sigma_{R\psi}^2}}\right)\times 100\%$$

The Δ's are the separations between the estimated and true positions in range, azimuth and elevation and the σ^2's are the RADAR measurement error variances. This is averaged over all fused aircraft tracks. This MoP has the properties of approaching zero as target and track approach infinite separation and approaching 100% as they approach coincidence. On this scale, on average, a single RADAR measurement would score 37% .

Two state fusion algorithms were compared. A composite of the measurements from the contributing sensors known to have the best measuring accuracy in each dimension (Eg range from radar, angular position from IRST) was compared with the optimal (least squares) weighted mean of the contributing tracks.

In this case the optimal weighted mean is theoretically more attractive than the composite because it uses all the data. However, it can be seen that the results are nearly indistinguishable. Again, the shape of the graph is due to the same scenario events.

Another feature of the graph which is of interest is the slightly poorer performance of the weighted mean around the 30 second mark. Due to the extra ambiguity in the scenario during this period, the sensor tracking performance was slightly degraded. This led to a small degradation in the weighted mean performance which was avoided when the composite algorithm was used.

This demonstrates a characteristic of optimal and near optimal approaches, namely they perform best only when the incoming data is reliable and conforms to the assumptions made about it. In this case, the assumption that all the data arose from the same target is implicit in the use of an optimal weighted mean algorithm. The composite is more robust because it relies only on the best (most accurate) measurement dimension of each sensor.

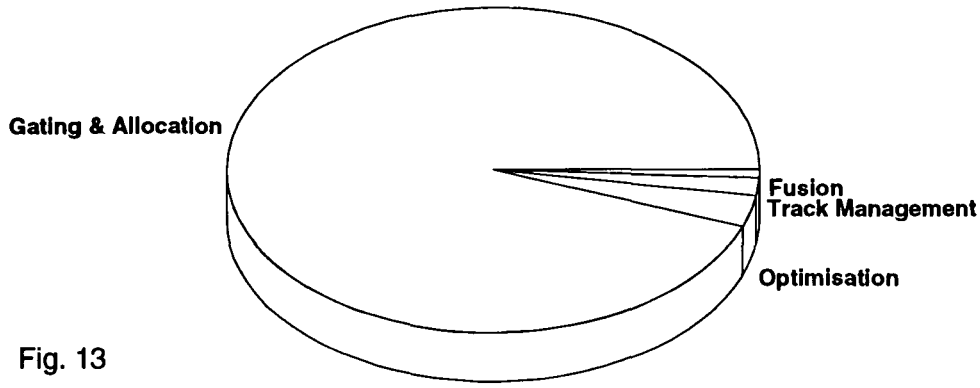


Fig. 13
Typical Processing Distribution

Finally, consider processing load. The question is discussed in section 4.1.2. The predicted factorial increase in data association processing as the number of targets and data sources increases may be observed. Figure 13 shows the processing distribution resulting from the Bayesian association algorithm and the optimal weighted mean state estimation algorithm, applied to the scenario described above. It was observed at a stage where all sensors and the comms network had acquired and were reporting on all targets (and friends) within their coverage. The algorithms used were the nearest to optimal, and hence the most processor intensive, in their class. The processing load MoP used here was simply the time taken by the computer to perform each task as a proportion of the time taken to perform the total task.

The striking thing about this result is the very small proportion of the computer's time which is actually spent fusing data. Most of the time is spent doing data association. Other results show that the apparent concentration of processing on the data association functions becomes more extreme as the numbers of targets and data sources increase, and that this result holds true when less optimal algorithms are observed.

7 CONCLUSIONS

The results, reproduced in this paper, are a small sample of the work completed to date on the Data Fusion Test Harness. They show how the characteristic features of Sensor Data Fusion system performance may be measured.

When a typical air superiority aircraft sensor fit was considered, it was seen to be highly constrained. To maintain background situation awareness whilst prosecuting a multiple target engagement would require support from other friendly sensor platforms via a data communications network if both tasks, surveillance and the engagement, were to be fully effective. In air to air combat, the capability to exchange data with other similar and dissimilar platforms offers many advantages.

The characteristics of the sensors also had an impact on the ability to perform multiple sensor identification. In particular, better ESM sight line accuracy would make identity fusion much more reliable.

Consideration of the architectural options revealed that a future air superiority aircraft system will, in all probability, have to perform some of its fusion at track level in order to utilise data from existing sub-systems or communications infrastructures. It was also noted that the aircraft will encounter situations where plot level fusion (or the algebraically equivalent information level fusion) of its local sensors would be an advantage. When resources allow, the most appropriate architecture for such a Sensor Data Fusion system is likely to be a Combined/Hybrid form, which will enable Track Level, Plot Level and Information Level Fusion to take place as and when they offer the best potential performance.

The results show that, for a future air superiority aircraft, equipped with non-commensurate sensors, algorithmic complexity in the fusion of the data has a lesser impact on system performance than the determination of the data relationships. For the communications network and tactical data exchange this means getting the alignment right and, in all cases, the data association must be good. The implication is that resources must be allocated to the data association task. This means system design and development

resources must be allocated to the provision of a first-class algorithm and on-line processing resources must be sufficient to allow the algorithm to resolve difficult situations.

8 ACKNOWLEDGEMENT

This text has been developed from Reference [8]. Many thanks to Mike Everett and Richard Freeman for their contributions to the original paper and their assistance in reviewing this revised version.

9 REFERENCES

[1] Multiple Target Tracking with RADAR Application (Artech House, S. S. Blackman, 1986).

[2] Multisensor Data Fusion (Artech House, Edward Waltz & James Llinas, 1990)

[3] Mathematical Techniques in Multisensor Data Fusion (Artech House, David L. Hall, 1992)

[4] Communication in Decentralized Data-Fusion Systems (University of Oxford Department of Engineering Science, S. Grime, H. F. Durrant-Whyte & P. Ho, OUEL 1900/91)

[5] Tracking and Data Association (Academic Press, Yaakov Bar-Shalom and Thomas E. Fortmann, 1988)

[6] An Experimental Evaluation of Distributed Identity Fusion (BAe Sowerby Research Centre, R Deaves and P Greenway, 1994, Proceedings (SPIE) International Symposium on Photonics for Industrial Application vol. 2355)

[7] Sensor Management using the Decentralised Kalman Filter (BAe Sowerby Research Centre, P Greenway and R Deaves, 1994, Proceedings (SPIE) International Symposium on Photonics for Industrial Application vol. 2355)

[8] Sensor Data Fusion for Air to Air Situation Awareness Beyond Visual Range (BAe Defence, Military Aircraft Division, C. A. Noonan, M. E. Everett, R. C. Freeman, 1993, AGARD Conference Proceedings 539, Pointing and Tracking Systems).

MULTI-SENSOR DATA FUSION IN COMMAND AND CONTROL AND THE MERIT OF ARTIFICIAL INTELLIGENCE

René G. Zuidgeest
National Aerospace Laboratory (NLR)
Anthony Fokkerweg 2
1059 CM Amsterdam
The Netherlands
Phone: +31-20-5113654
fax: +31-20-5113210
E-mail: rgzuidg@nlr.nl

SUMMARY

The human operator observing the real world is confronted with a huge amount of data from multiple sensor systems observing that world. Multi-sensor data fusion (MSDF) is one of the emerging fields in advanced information processing, concerned with fusing sensor data from these multiple sensor systems. Automated multi-sensor data fusion can help the operator by processing sensor data into concise and surveyable information, that is more useful than every sensor system separately can provide.
The merit of MSDF can be increased by employing the knowledge of the human operator about the real world, the sensor systems and the fusion process. With the aid of this knowledge, automated MSDF can assign meaning to sensor data and is able to reason about the observed world at a high level, comparable to what humans do.
Artificial intelligence provides techniques to represent this knowledge and to reason with it. These techniques are discussed in the context of a generic framework comprising a world model and fusion processes. These techniques can contribute to an effective updating of the world model and can support its fusion processes. In addition, a global distributed fusion architecture based on the framework is proposed. As specific domain of fusion, battlefield surveillance is considered.
This paper shows the potential use of artificial intelligence in multi-sensor data fusion.

1. INTRODUCTION

Multi-sensor data fusion (MSDF) can be considered an important field in advanced information processing [1]. MSDF is the process of combining sensor data in space and time in such a way that it provides more relevant information than each sensor system separately is able to.

The increasing importance of automated MSDF is driven by a technology push as well as by a market pull. The technology of sensor systems is rapidly growing, more and more sophisticated and complex sensor systems are coming available on the market. They provide a huge amount of data, creating a need for advanced information processing through MSDF.
On the other hand, the real world is getting more and more complex [2]. Dissimilar sensors operating in different spectral regions are required to detect the full variety of objects present in the real world [3]. For a human operator monitoring the real world through a set of dissimilar sensor systems of increasing complexity, it is a significant problem to fuse the sensor data, to assess the real world and decide on proper reactions within a limited time frame. Because of excessive data, ill-digested information and stress, wrong interpretations about the situation in the real world might be made that may have disastrous consequences.

MSDF can be applied to various domains. Domains of research at NLR are air traffic control [4], multi-radar tracking employing uncertainty techniques [5], navigation based on Kalman filters [6], battlefield surveillance [7], air defence, and remote sensing.
World wide, research into MSDF is mostly performed in a military context. One domain of applied MSDF and where this paper focuses on is command and control (C^2). In this domain, commanders take decisions on the basis of fused information from various sensor systems located on and observing a battlefield. An example of a naval application in this field is the SIAP-project [8]. Other applications are AMUID performing battlefield analysis on basis of sensor information [9], and ECRES [10, 11] and IDA [12], performing the same function, but on the basis of intelligence information (e.g. human reporting).

The wide range of delicate applications (e.g. human lives are involved in air traffic control and command and control) justifies the research into MSDF. Currently, the field of artificial intelligence (AI) is in the spot-light to support MSDF. Sensors provide only numerical data of measurable quantities (e.g. signal strength, polarisation). Processing of this numerical data and performing calculations (such as calculation of the position of an object) is necessary. However, a great

deal of data can be transformed to a higher symbolic level and consequently can be reasoned with in a more abstract way comparable to what humans do by using explicit knowledge about the domain. AI is a surplus value to MSDF, especially in advanced sensor control and allocation, identification of objects, assessment of the situation in the real world and prediction of future states of that world by using knowledge about objects (i.e. their structure, their relation with sensor information, their behaviour and the contexts in which they act, etc.). Systems employing AI technology could serve as an intelligent interface transforming excessive and complex (sensor) data in real-time into surveyable and relevant information for the operator [13].

Chapter 2 provides a general functional architecture of an MSDF system based on a command and control model, applied to battlefield surveillance. Battlefield surveillance is considered as the continuous observation of the battlefield area to provide timely information for command and control functions.
Chapter 3 globally describes the world model that includes MSDF and knowledge about the observed world.
Chapter 4 discusses a specific set of AI techniques for representation of that knowledge and reasoning with it. These techniques emerge from knowledge-based systems (KBSs); neural networks are not considered in this paper. KBS techniques have been preferred because of their relative maturity, their ability to explain their reasoning process in a comprehensive manner (might be important in order to convince the operator) and the ease with which explicitly represented knowledge can be modified. Neural network techniques lack these important features. However, neural networks and KBSs can be complementary, where neural networks reside at a lower level of information processing than KBSs. Integration of these two techniques might provide interesting results.
Chapter 5 describes a global distributed architecture for MSDF for C^2 networks where KBS techniques and distributed AI play an important role.
Finally, Chapter 6 presents concluding remarks.

2. BATTLEFIELD SURVEILLANCE AS CONTEXT FOR MSDF

The basis of the application of MSDF in the domain of battlefield surveillance as presented in this paper is a generic C^2 model. Four command levels are identified in this model: highest, intermediate, lowest and executive. These levels have their equivalents in the Air Force, Navy and Army C^2 structure.
The data flow between the levels is cyclic. First, global tasks are generated by the highest command level, which are worked out and decomposed by the lower command levels, up to the executive. If the tasks have been executed, reporting is done all the way up to the highest command level. The C^2 cycle is closed when these reports have been assessed by the highest command level. Time and data are the most important factors that distinguishes the levels: the lower the level, the more time critical and the more detail in the data and information.

The C^2 *functions* for one command level are given in Fig. 1. Five main functions are distinguished, presented in the inner ring. A commander is tasked by a higher command level. In the context of these tasks, the current battlefield situation is analyzed. After the analysis, decisions are taken (how to implement the task) and available resources are allocated. Then the orders are prepared to task a lower command level. In the execution function, interaction takes place between the two command levels; sometimes the orders need to be readjusted because the situation has changed during preparation of orders or the commander had incomplete or wrong information. After the orders have been executed, the results are reported, a reassessment based on the report information and new sensor data is made and reporting is done to the higher command level.

The five main functions can be decomposed in a number of *processes*. These processes are displayed in the outer ring of the C^2 model. Because this paper mainly covers the first function, only the processes (1) collection of information from sensor systems and intelligence, (2) composition of the battlefield (including MSDF) and (3) analysis and assessment of the battlefield are discussed here. Note that the processes are performed in the context of the task issued by the higher command level.

The function situation analysis containing fusion and interpretation of data is currently done in the human's mind. However, because of the large amount of data which is made available by current technology and the inherent complexity of the data, it becomes more and more difficult to obtain and combine the *relevant* information out of this data stream and evaluate it properly within certain time constraints. AI provides tools and techniques for automating at least part of human knowledge. Therefore, AI can support automation of fusion processes which are now performed by the human operator. By automating low-level routinely tasks, the human operator or commander can focus on more important tasks like high-level assessment and decision-making based on interpreted fused sensor data.

Fig. 2 depicts a general architecture for battlefield surveillance incorporating automated MSDF. The architecture is based on the situation analysis function of the C^2 model. It consists of a number of geographically distributed platforms with mounted sensor systems observing the battlefield. These platforms provide symbolic sensor reports about observed events, detected objects, etc. (i.e. the sensor reports are the sensor system's output of object detection and signal-to-symbol transformation processes). These sensor reports are sent to a fusion centre (*information collection*) where they are spatially and temporally aligned (*scenario composition*) to provide a battlefield description which is then

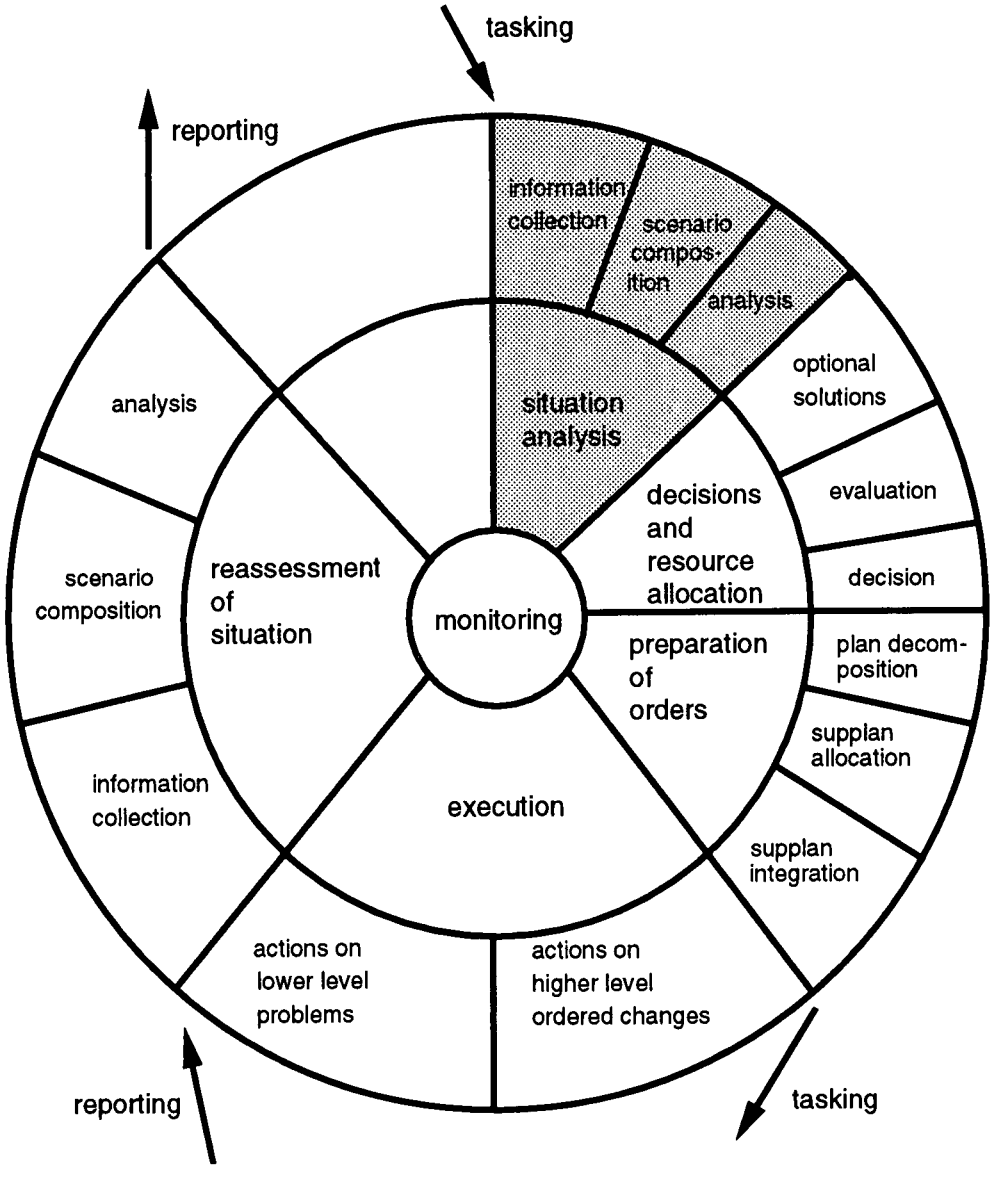


Figure 1. *A generic C^2 model with its functions and processes at one command level* [1].

interpreted (*analysis*) resulting in the battlefield situation description. This battlefield situation description is presented to the operator through a man-machine interface.

Based on the fusion and interpretation results, it might be needed to direct the sensor systems in order to obtain an optimal battlefield situation description. The fusion centre as well as the operator can issue requests for additional sensor information (e.g. focus on specific area) to the sensor manager. It constructs and maintains a global temporal plan in which sensors/platforms are allocated and distributes it to the sensor platforms that implement the plan.

In the following chapters, the fusion centre node containing the battlefield description is worked out in more detail.

3. WORLD MODEL AND DATA FUSION

This chapter focuses on the world model and the fusion process. These elements are located in the fusion centre node of Fig. 2. The world model is a reflection or simulation of the real world in time and space. In the context of battlefield surveillance important aspects to be represented in a world model are military objects (their structure, behaviour, and context), terrain and weather circumstances, sensing systems (their capabilities and limitations) and the relationships (e.g.

[1] This model has been developed at NLR by R.P. De Moel and B.J.P. van der Peet for an expert meeting on *Computing Technology relevant to Time Critical Command and Control Applications* (IEPG/P-3/SG-6/WG).

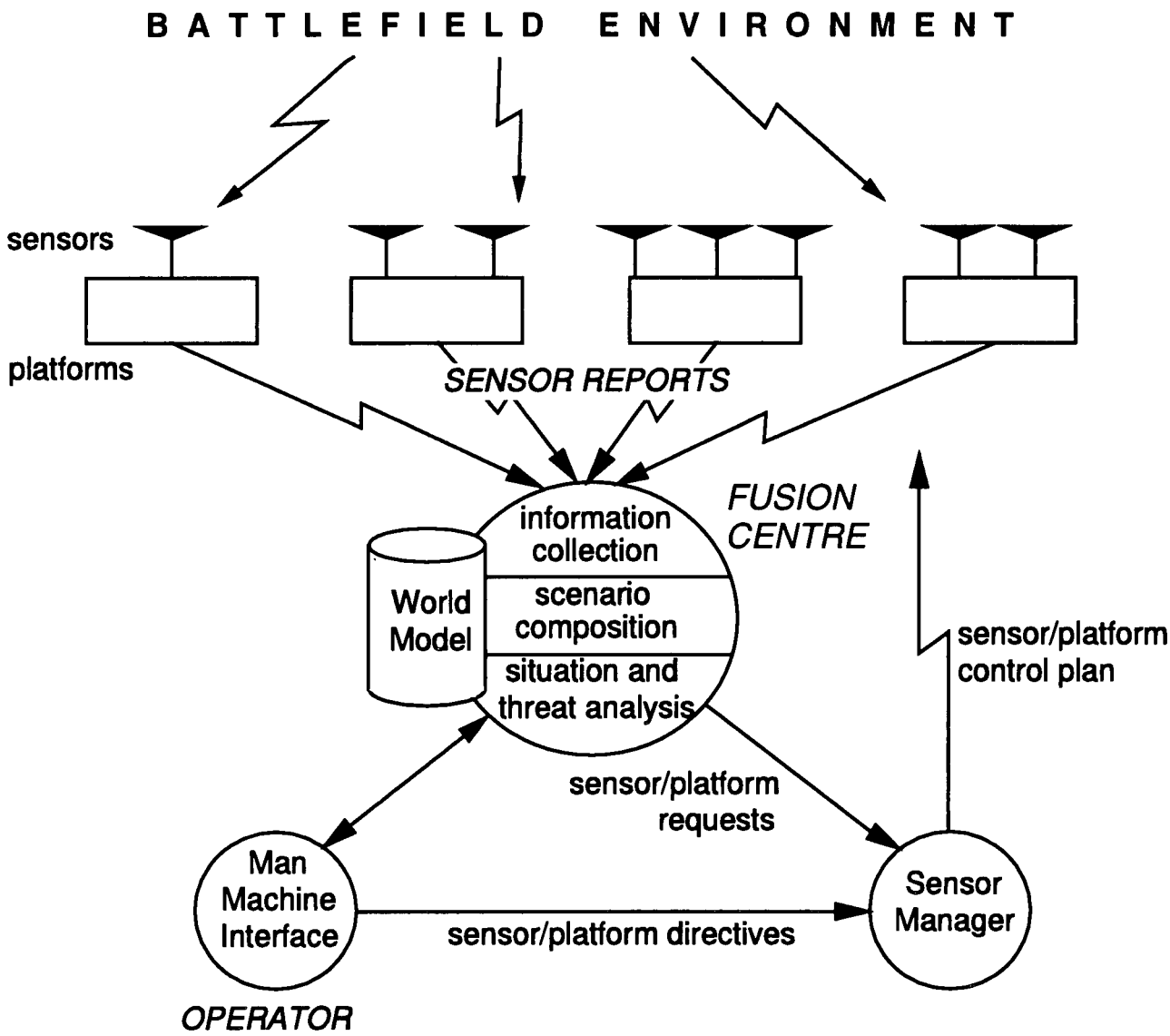


Figure 2. *A general system architecture for battlefield surveillance incorporating MSDF.*

causal effects) between these aspects (e.g. signature of object sensed by specific sensor under certain terrain/weather conditions). In addition, the world model contains inference knowledge to fuse sensor data, assess and predict object status and observed situation, identify objects, and deduce more abstract, relevant and concise information about the real world in an effective and efficient manner.

The world model has two main input streams which are categorised as bottom-up or top-down. Bottom-up data is a continuous stream of data about the real world such as (pre-processed) sensor data, weather reports, terrain conditions, and intelligence data [14]. Top-down data is more discrete and consists of requests for information (from position of an object to complex *what-if* questions) issued by a human operator through a man-machine interface or by an automated system such as a resource planning system for counteractive actions. These requests initiate a top-down, goal-directed search in the world model to extract or infer the required information. In fact, the set of possible requests for specific types of information defines the purpose of the world model, i.e. to provide valid answers to questions about the real world, and hence defines - given the application domain - the construction and contents of the world model.

For the purpose of this paper, the world model is assumed to be based on a blackboard concept [15, 16]. The blackboard concept is well-suitable for problem domains in which large amount of different source data, large number of competing hypotheses, different levels of abstraction and multiple symbolic representations are involved [17]. For the problem domain MSDF in C^2 and in particular battlefield surveillance, the blackboard model consists of a blackboard information structure representing hypotheses about the real world at different levels of abstraction, and a number of knowledge sources about the different sensor systems

(e.g. ESM, Radar, IR), the battlefield environment (i.e. terrain, weather), the military domain (objects, tactics, etc.) and their inter-relationships. This chapter discusses the world model in terms of this blackboard structure and the possible fusion processes that can be incorporated in knowledge sources operating on that structure.

As information structure, a hierarchical representation fits best in which each level in the hierarchy is an abstraction of the lower. The two main reasons are the hierarchy in the fusion process (e.g. fuse sensor data into *one* object, see Section 3.2) and in the military domain. Four levels have been identified: (1) sensor level, (2) object level, (3) recognition level, and (4) relational level.
The sensor level describes the object measurements (e.g. contour and temperature), represented in sensor reports. Correlation of similar sensor reports in time results in sensor report tracks.
The object level contains information about objects on the battlefield and is the result of spatial and temporal correlation of sensor reports and sensor report tracks at the sensor level.
The recognition and the relational level comprise the tactical level in military terms. The recognition level contains military relevant information of single objects on the battlefield (e.g. identity). The relational level describes relations among objects, resulting in the detection and recognition of units or battle formations. These two levels contain information which is fully abstracted from sensor-dependent data. At these levels, *concepts* like division, tank, and the relationships among them are represented.

The knowledge sources operate on one or two levels of the blackboard (i.e. the levels of the world model). Knowledge sources can be specialized in fusing similar sensor report tracks into one object track, identifying objects from the object level to the recognition level, recognizing units, identifying representative objects within units and monitor them only, etc. These knowledge sources are responsible for the fusion of the data to higher levels of abstraction.

The next sections will discuss each level and the relationships and the fusion processes between these levels.

3.1. The Sensor Level

The sensor level describes object measurements (e.g. signal strength, Doppler speed, size) and characteristics about the measurement (e.g. type of sensor system, resolution, position of sensor system), which are both represented in *sensor reports*. These are the output of object detection and signal-to-symbol transformation processes of the various sensor systems mounted on platforms. The sensor level is the least abstract level of the world model. A sensor report is a low-level description of a phenomenon (a potential object or target or event that has a high degree of discrimination in relation to its environment, e.g. a hot spot indicating an engine of a tank or an explosion).

A sensor report consists of a number of attributes. Attributes concerning measurement characteristics depend on the sensor system and platform from which the sensor report originates. These are: platform ID and position, sensor system ID and type, time stamp and others like sensor performance, accuracy, and a preliminary confidence value of the observed phenomenon.
Other sensor report attributes concern measurement(s) about the detected phenomenon. These attributes, called *object features*, are symbolic representations of the signal features extracted from phenomena in the real world by a sensor system. The kind of attributes and their dimensions depend on the sensor system type and position of the platform. Table 1 shows the relation between various sensor system types, the measured signal features and the resulting object features.

Tracks of sensor reports acquired at successive times, but having similar signal/object feature values are initiated and maintained, resulting in sensor report track hypotheses. Such a hypothesis represents the belief that a set of successive sensor reports are manifestations of the same object in time. In principle, a track is formed by sensor reports from the same sensor system or similar sensor systems, because they have a common format and attributes and are, therefore, easier to correlate.

This type of fusion of sensor report into tracks happens only at the sensor level. From sensor to object level, individual sensor reports or sensor report tracks are fused into an object or object track.
Sensor report or sensor report tracks are correlated or associated to object tracks on basis of spatial data, radiometric data or the context of the sensor report. Sensor reports are correlated if their spatial references are very close or because their non-spatial object features are similar. An example of spatial fusion is the fusion of an IR and a radar sensor report track of the same platform with overlapping spatial references. An example of non-spatial fusion is the cross-section of ESM sensor reports having similar values for the non-spatial attributes (e.g. common frequency) in order to determine precise position. An example of contextual fusion is that a sensor report is part of a pattern of sensor reports (e.g. representing a column), which makes correlation based on context possible (e.g. on basis of the relative position in a column). The extent to which fusion of sensor reports and tracks can be successfully performed depends on a number of parameters such as acquisition time, object activity, density and discrimination, and sensor characteristics and performance.

3.2. The Object Level

The information at the object level consists of objects (or events). The objects at object level in the world model are hypotheses, expressing the belief that a set of sensor reports or sensor report tracks are concerning the same real-world object. Different sensor report tracks (possibly acquired from different types of sensor systems) may refer to the same object. At object level,

Signal feature	Object feature	Sensor system
propagation time	range	radar, laser
azimuth/elevation	azimuth/elevation	radar, IR, TV, laser, ESM
Doppler shift	velocity	radar, laser
reflected power level	Radar Cross Section	radar
polarisation	material	radar, laser
video image	contour	IR, TV, laser
frequency	frequency	ESM
modulation	modulation	ESM
	classification[2]	ESM, radar

Table 1. *Relation between signal features, object features and sensor system types.*

these tracks are fused. Moreover, for each tracked object, more abstract information is inferred on basis of knowledge about object features and how they are sensed by sensor systems and manifest in sensor data. This abstract information does not contain any specific sensor data. Examples of such information are mobility, fire capability, relation with other objects (context), etc. If enough sensor data is acquired and information is inferred, an object can be classified and promoted to the recognition level.

At the object level, tracks and features of objects are maintained and predicted. Contextual knowledge plays an important role with respect to accuracy. For example, in case of extrapolation of tracks, if an object is following a road for ten minutes (e.g, *track mode*: "road following") then the track can be extrapolated to the next crossing met. Moreover, a line of bearing (e.g. an ESM sensor report), can be crossed with the road to determine precise position under the assumption that the object is still following the road.

3.3. The Recognition Level

Also at the recognition level, single objects on the battlefield are described, but the information describing the objects is much more abstract and contains more military relevant information, like identity of an object and related potential capabilities (e.g. threat).

The relation of the recognition level to the object level is that an abstract name or class, for example *tank*, has been associated to the object attributes; in other words, a conceptual meaning is assigned to the objects. The recognition level consists of hypotheses about the classes to which the objects belong. This association of objects to classes has economical advantages regarding processing and memory, because the concept tank incorporates much implicit information (e.g. has a barrel, has treads, is mobile and armoured) that was explicitly represented at object level. Reasoning about a tank is easier than reasoning about a large set of - partly sensor-dependent - attributes representing the object.

Recognition is done on the basis of two types of information: (1) structural and behavioral information of the object, and (2) contextual information of the object.

The hypotheses at recognition level depend on these two types of information. Recognition based on structural and behavioral information is, for example, recognizing a tank by respectively its shape obtained from a TV sensor and the fact that it is moving. An example of recognition based on contextual information is the recognition of an object as tank on the fact it moves inside the borders of an area in which a *tank* company is operating.

3.4. The Relational Level

The relational level represents inter-object relations and the (tactical) situation of the area under observation. In the world model, battle formations (units) are represented by means of groupings of (military) objects and units. The relational level consists of several sub-levels, corresponding to the military hierarchy. These sub-levels are bottom-up (derived from [18]): platoon, company, battalion, regiment, and division level.
Each unit in the world model is a hypothesis expressing the belief that units (or objects) at lower levels together form one coherent unit in the real world. At which level that hypothesized unit is represented, depends on the contents of the real-world unit (types of objects and numbers) and its tactical behaviour. An hypothesis contains also track information. The maintenance of this

[2] Today, certain types of sensors (in particular radar and ESM) can perform preliminary classification of the detected objects. This is not considered as a signal feature, but, of course, this information should be provided in the output vector as object feature.

information is less time critical than object level information, because the dynamics of a unit is less than of each single object. This makes tracking of units without tracking of every individual object possible, which is less difficult and saves computing resources.

Tactical, strategic and doctrinal knowledge (knowledge about the order of battle) plays an important role at this level in order to determine the status of the situation. Also, this knowledge can be used to detect and recognize units and objects according to their tactical behaviour in the order of battle (type of inference based on contextual information and knowledge).

The purpose of this level is two-fold: (1) provide the operator or commander with high-level, surveyable and comprehensible information about the battlefield for subsequent assessment and decision-making, and (2) provide the lower levels and knowledge sources the necessary context to assist and direct their inferences.

4. USING AI TECHNIQUES TO MAINTAIN THE WORLD MODEL

In Chapter 3, the world model and fusion processes incorporated in knowledge sources have been described. The information describing the battlefield situation and the knowledge used to predict possible future situations and to infer new information from old information and newly acquired sensor reports can be supported by a number of techniques from the field of AI. This chapter discusses a number of candidate AI techniques that can support the maintenance of the world model and the provision of answers on questions to this model.

4.1. Representation Techniques

Chapter 3 mainly focused on the structure of the blackboard (i.e. the hierarchical representation of hypotheses), but the representation of the knowledge in the knowledge sources was not discussed.
It is unlikely that a common representation of knowledge for all the sources can be used. The lower levels (especially the sensor level) include much numerical processing algorithms and knowledge is likely to be represented implicitly in program code inside sensor data processing modules. However, at the higher levels, representation techniques from the field of AI are applicable. The most well-known are: (1) *semantic networks* suited to describe conceptual relationships (tank *is a* vehicle) and contexts, (2) *frames* to describe and model real world objects and their structure, (3) *production rules* to describe causal relationships (*if* enemy division moves to city *then* city is in danger), and (4) *scripts* to describe sequence of events (e.g. to represent military doctrine and tactics; special actions, e.g. crossing a river by a division, can reveal its organization).

4.2. Inference And Control Techniques

A number of inference techniques exist in deducing new information from old information using explicit knowledge. The most well-known techniques are *forward reasoning* and *backward reasoning*. Forward reasoning is typically data-driven that is heavily applied to lower levels of information processing. Acquired sensor data is quickly processed and prepared for higher level inferences. At these higher levels, more goal-directed inference techniques are applied in order to work towards a solution satisfying some goal (e.g. answering an information request from the operator) and to control the combinatorial explosion effect inherent to data-driven techniques. It is important to design a control method that keeps the number of inference steps towards an optimal solution restricted (i.e. keeping the combinatorial explosion under control) by finding the right balance between the application of data-driven and goal-driven inference techniques.
Another inference technique is *inheritance*. Inheritance is based on hierarchical relationships such as *is a* and *has a* (e.g. a tank *is a* vehicle and therefore inherits the quality that it is movable from the class vehicle, or a division *has a* battalion and therefore the velocity of the battalions is inherited by the division where they are part of).

Without sufficient control the basic inference techniques can lead to a combinatorial explosion of facts and sub-goals. Real-time performance of a knowledge-based system requires control techniques to control the inference and search processes through knowledge and information. Domain-independent search and control techniques (e.g. A^*-algorithm) as well as domain-dependent techniques (e.g. expert military knowledge for meta-level control and demons that are only activated in case of specific events) have to be employed to satisfy time constraints [19, 20].

4.3. Techniques Dealing With Imperfect Information

The fact that sensor systems do not provide accurate information, due to internal functioning or external conditions (e.g. weather, terrain, ECM, deception) has implications on the beliefs in the world model. The hypotheses in the world model are not a priori true and might be in contradiction to one another. This opens a discussion on how to handle imperfect information.
Information is imperfect if it has one or more of the following characteristics.

1. The information is <u>incomplete</u>: not everything is known. Nevertheless, conclusions may have to be drawn, possibly in the form of hypotheses. Incompleteness may be caused by: sensor coverage limited in space and time, lack of information about enemies, and non-generality of knowledge: exclusions.
2. The information is <u>uncertain</u>: a proposition can not be said to be true or false. Instead, only some indication of the 'belief' in a proposition can be given. Causes may be: incorrect information (e.g. false), incomplete evidence, (ir)reliability of sensor

reports, changing information (when, how?), and elapsed time (no update information).

3. The information is inexact: the information itself is intrinsically vague. This can be caused by: sensor system inaccuracies and inadequacies, processing inadequacies, navigation errors, time delays, vague expressions (temporal and spatial), etc.

Various approaches have been developed to formalize reasoning with imperfect knowledge. These approaches focus on combination and propagation of uncertainty in inferences, on detecting and resolving contradictions and on belief revision [5]. It appears that all these techniques have their problems.
With respect to uncertainty management, formalization of combination and propagation of uncertainty values in inferences appears to be difficult [21]. For example, the *Bayesian inference theory* [22] suffers from problems about assessing subjective prior distributions by humans, even if they do not know much about it, and requires evidential data to be mutually independent. The *Dempster-Shafer theory* [23] is an improvement in this respect, but its formulas are complex and therefore suffers from a computational problem that could be a drawback with respect to real-time performance.
Another technique is *fuzzy logic*, which has been developed to enable reasoning with fuzzy, vague notions (*fuzzy sets*) as people do [24, 25]. Fuzzy reasoning is not specifically geared towards real-time performance, and defining fuzzy sets and the fuzzy logic for a specific application also appears to be problematic.

Other techniques that are complementary to the techniques discussed previously focus on the fact that the real world is a dynamic world, its state changes continuously with time. This requires a continuous monitoring of the integrity of the world model. Detected contradictions (e.g. between newly acquired information and current information) need to be resolved through revision of the beliefs (i.e. hypotheses in the world model).
Two techniques dealing with revision of beliefs are *truth maintenance* and *non-monotonic reasoning*.
A Truth Maintenance System (TMS) [26, 27] maintains the beliefs (possibly a multiple set of hypotheses, representing different beliefs about one situation) on which future inferences will be based. A TMS serves as a kind of administrative registration system. It is able to detect contradictions in the set of facts and conclusions from which they were drawn.
A non-monotonic reasoning process can withdraw specific conclusions and facts (assumptions) causing the inconsistencies (it is called non-monotonic because the set of conclusions might be smaller than before). There is a lot of research going on in developing non-monotonic logics [28, 29], and their integration with TMSs [30, 31]). Temporal reasoning [32, 33, 34, 35] is closely related to truth maintenance and non-monotonic reasoning and much research is performed in integration of these fields into one single reasoning system that detects and resolves contradictive information in time [36, 37]. Each technique has its problems like NP-completeness of TMSs, the generality of research on non-monotonic logics and reasoning), and the frame problem in temporal reasoning.

5. GLOBAL DISTRIBUTED FUSION ARCHITECTURE

In the previous chapters, the world model within a central fusion node in the context of battlefield surveillance was discussed. The framework of the world model was based on a blackboard architecture in order to provide a surveyable insight into the information streams and fusion processes, and applicable AI techniques. For C^2 applications, one central fusion node will not be adequate. An architecture needs to be designed that in particular fulfils the following requirements.

1. *It shall support the fact that sensor systems are geographically distributed.* Communication protocols, type of data sent over (such as raw data, processed data or interpreted data), throughput, and distributed sensor management play an important role.
2. *It shall include the option that fusion might happen on platforms as well as in C^2 nodes.* Fusion might be partly performed on the sensor platform. In this case, interpreted data (e.g. object tracks) is sent over as the result. Reasons for this might be
   - to reduce data throughput which is less than if raw sensor data is sent over, or
   - to have interpreted data locally available for immediate action or accurate local sensor management.
3. *It shall be modular, reconfigurable and scalable.* This in order to be flexible with respect to
   - investigation of fusion of specific sensor combinations in an isolated manner,
   - number of deployed sensor systems,
   - characteristics of the battlefield and theatre of operations,
   - the expected sensor data throughput,
   - sensor data processing happening at different sites and/or on different machines, etc.
4. *It shall have inherent parallelism*, so that architectural elements can have private processors.
5. *It shall incorporate robustness.* If a fusion function or sensor system fails, then the system shall still be operational.
6. *The architectural framework shall be general*, and applicable to many war theatres, C^2 network structures, and application domains (e.g. crew assistance for combat aircraft [38], naval domain or air defence).

The requirements rule out a centralized or sequential processing architecture. A decentralized architecture [39, 40], based on loosely coupled, a-synchronous, course-grained, semi-autonomous agents (Fig. 3) may be better. The architecture comprises platform nodes, data fusion nodes, assessment and sensor management, and man-machine interfacing.
The platform nodes are located on and around the battlefield (ground or airborne) observing the battlefield through a number of mounted sensor systems. They broadcast sensor reports which are processed by the various fusion nodes.

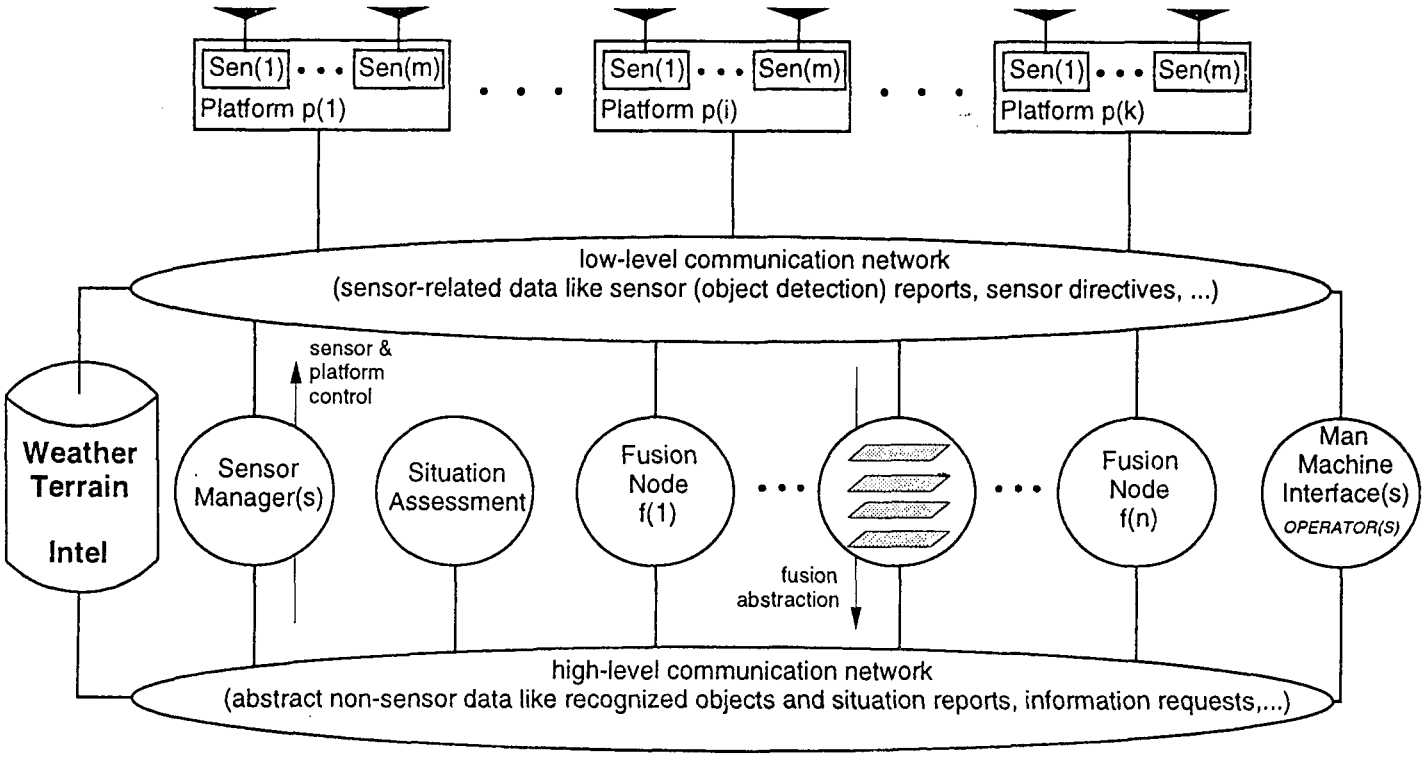


Figure 3. *Distributed architecture incorporating MSDF.*

The main function of the fusion nodes is interpretation of sensor data. All fusion nodes incorporate the world model framework as discussed in Chapter 3 and rely on conventional as well as advanced information processing techniques such as those discussed in Chapter 4. Every node might be a specific instance of a generic fusion shell, in which specific knowledge is entered that mainly depends on the set of sensors to be fused and terrain under observation; the knowledge in the world model of a fusion node that is specialized in fusing heli-borne ESM and radar reports, differs from knowledge that is specifically focused on fusing ground IR and radar. The higher levels, i.e. military and unit level, have a common representation of information, because these do not include sensor-dependent information. This enables the fusion nodes to exchange this high-level information across a communication network. In this way, the nodes can assist each other with the interpretation process. For example, one node might ask another whether it saw also a tank at a certain position *X*. If so, accuracy of the position and certainty about the identity (it is a tank) can be increased.

The findings of the fusion nodes are used as input to the situation assessment node. This node takes the high-level information of the specific fusion nodes as input for situation and threat assessment. The assessment node has been added to the architecture because it can provide the operator as well as the sensor management system high-level global information of the battlefield situation which makes well-organized decision-making and sensor management possible. In this context, small-scale situation and threat assessment is performed. An alternative is to distribute this function among the fusion nodes, eliminating centralized assessment. In this case, local assessment might be the basis for local sensor management.

The sensor management system provides feedback to the sensor systems on the basis of the findings of the other processes and the human operator. The nodes (in particular the assessment modules) and the human operator can send the sensor management system requests for sensor data. In this way, the total system is able to anticipate to battlefield situations. It does that by constructing a global plan that allocates and controls the platforms and sensor systems, based on a pre-specified plan and sensor data requests. The global plan will be distributed to the platforms where local sensor managers work out the plan.
The basic functions of the sensor management system are: (1) the monitoring of global sensor system performance, (2) the processing of sensor data requests into a plan, (3) the allocation of platforms and sensor systems and (4) the maintenance of a long-term observation strategy plan in order not to loose global observation by focusing too much and too long on local areas.
The sensor management plan could be constructed on basis of algorithmic techniques as well as advanced techniques. It can use explicit knowledge about sensor systems (e.g. with respect to terrain and weather) and AI-based planning techniques.

The final component of the architecture is the man-machine interface (MMI). Its main goal is to depict in a surveyable manner the battlefield situation, mainly based on the information in the assessment node. If necessary, the operator can consult other systems or

nodes represented in the architecture. Through the MMI, the operator can directly influence the sensor management plan.

The distributed architecture of agents discussed in this chapter opens the world of distributed artificial intelligence (DAI) and processing [41]. This key technology fits very well to the inherent distributeness of the C^2 process. Several alternatives are possible to map this architecture or multiple instances of it [17] on existing C^2 networks. For example, each C^2 node in a network may have co-located a number of sensor systems, one or more fusion nodes to fuse the sensor data, an assessment node, a sensor manager to control the sensors and an MMI. In addition, a communication module should be available in order to exchange information (battlefield information, sensor control plan, etc.) with other C^2 nodes. Important questions to be solved are the way of communication between the various agents and between C^2 nodes (locally as well as globally) and how to perform conflict resolution in order to achieve globally coherent behaviour of the C^2 network of data fusion systems [42].

To conclude this chapter, note that the system architecture fulfils the requirements.

- It copes with geographically distributed sensor systems.
- It is modular, reconfigurable and scalable. The modules can be situated on platforms as well as in C^2 nodes and can be installed on separate hardware. Furthermore, if extra sensor systems are placed in the field, an extra fusion node can be created and added, based on the same framework as the others. If the set of sensor systems or the operational theatres changes, the architecture still holds, the main thing to do is to down load new data and knowledge into the various data bases and knowledge sources about the used sensor systems and operational theatre such as tactical data and terrain/weather data.
- The architecture also incorporates robustness in case of failure. If a sensor system or fusion node fails, the system will still be operational. It is preferred that each element in architecture runs on private hardware, so that a hardware failure will not shutdown the complete system. Beside this, real-time performance is increased because processing is distributed among multiple parallel machines.
- And finally, it is thought that the architecture is flexible enough that it can be applied to many battlefield situations and operational theatres, such as the naval domain or air defence.

6. CONCLUDING REMARKS

This paper describes the potential use of artificial intelligence to enhance multi-sensor data fusion in the field of C^2. A C^2 model has been presented, and the place of MSDF in this model has been discussed. Knowledge about the domain of fusion can contribute to better performance of data fusion. To mention some basic pieces of knowledge that can support dat fusion processes:

- knowledge on sensor systems and how they operate, given the terrain and weather conditions and objects to be detected,
- knowledge on the manifestation of objects in different types of sensor data,
- knowledge on typical contexts of an object (for example with which objects it usually cooperates),
- and knowledge on tactics and doctrines to infer and assess the battlefield situation.

AI techniques using this knowledge can aid in effective control of the platform suite and in directing the fusion processes by only focusing on relevant parts of sensor data instead of processing and fusing all information that is offered by the sensor systems.

Because of the explicit representation of knowledge, it is expected that a fusion application in battlefield surveillance can be relatively easily transformed into e.g. a naval application by "only" replacing the knowledge. Flexibility and adaptability is provided in case of changing military context by replacing the knowledge depending on the theatre of operations.
The framework of a distributed architecture with its multiple communicating agents and incorporated AI techniques remains valid in many other applications, e.g. crew assistance for combat aircraft. For the Air Force, the link between battlefield surveillance and e.g. Close Air Support, Battlefield Air Interdiction and Air Reconnaissance missions should be clear.

Research of application of AI in domains like C^2 should be encouraged, especially with respect to real-time performance and integration into C^2 infrastructure with its existing computer systems which are mostly based on conventional technology [13, 37]. In this respect, techniques of major concern are representation and inference techniques, and techniques dealing with uncertainty. In addition, distributed problem solving techniques and related architectural solutions like the blackboard model are equally important. Further research should actually prove the usefulness of AI and the maturity and applicability of its techniques in the domain of multi-sensor data fusion.

7. REFERENCES

[1] Waltz E., Llinas J., *Multisensor Data Fusion*, Artech House, Inc., ISBN 0-89006-277-3, 1990.

[2] Harris C.J., *Knowledge Based Systems for Real-Time Command and Control Systems*, Large Scale Systems: Theory and Applications 26-29 Aug., Zurich, p477-80, Vol. 2, 1986.

[3] Chaudhuri S.P., Agrawal R., *Artificial Intelligence applications to Command, Control and Communications Systems/Subsystems*, Sensor Fusion II, Proceedings of the Meeting, Orlando, 28-29 March, p85-96, 1989.

[4] Blom H.A.P., Hogendoorn R.A., Schaik F.J., *Bayesian multi-sensor tracking for advanced air traffic control systems*, AGARDograph 301 on

Computation, Prediction and Control of Aircraft Trajectories, 1988.

[5] Donker J.C., *Reasoning with Uncertain and Incomplete Information in Aerospace Applications*, AGARD Symp. on Machine Intelligence for Aerosp. Electronic Systems, Lisbon, May 16, 1991.

[6] M.A.G. Peters et al., *Development of a TRN/INS/GPS Integrated Navigation Systems*, IEEE/AIAA Tenth Digital Avionics Conference, Los Angeles, 14-17 Oct., 1991.

[7] Zuidgeest R.G., *Multi-Sensor Data Fusion and the Use of Artificial Intelligence*, NLR TP 92174 L, Proc. of the 12th Conf. on Artificial Intelligence, Expert Systems and Natural Language, Avignon, June 1-6, 1992.

[8] Drazovich R.J., *Sensor Fusion in Tactical Warfare*, 1983, Advanced Information & Decision Systems, Mountain View, CA.

[9] Spain D.S., *Application of Artificial Intelligence to Tactical Situation Assessment*, 1983, 0531-6863/83 0000-0457 IEEE, Adv. Information & Decision Systems, Mountain View, CA.

[10] Naylor R.T., et al., *Battlefield Data Fusion*, in Harris D.J. (ed.): *Application of Artificial Intelligence to Command & Control Systems*, Peter Peregrinus Ltd., 1988.

[11] Denholm P., *An Enemy Contact Report System (ECRES)*, in: *Intelligence Generation and its Integration into Military Command and Control Systems*, 1988, AFCEA Europe Seminar, Rome.

[12] Edwards P.H., *Fusion of Land Battle Information*, Agard Symp. on Real-time aspects in Air Battle Management 3-6 Oct. 1988, Copenhgn, Denmark, BAe Warton Aerodrome, Preston PR4 1AX UK.

[13] Lehner P.E., *On the Role of Artificial Intelligence in Command and Control*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-16, No. 6, Nov/Dec, 1986.

[14] Goubet M., Desjouis M., *Using Artificial Intelligence for Battlefield Surveillance*, Signal Vol. 44, No. 2, p69-73, Oct, 1989.

[15] Nii P.H., *Blackboard Systems*, AI Magazine Vol. 7-2/3 1983, Standford Univers., CA 94305.

[16] Hayes-Roth B. (Stanford University), *A Blackboard Architecture for Control*, Elsevier Science Pub. b.v., p251-321, 1985.

[17] Adler R.M., Cottman B.H., *A Development Framework for Distributed Artificial Intelligence*, Proc. of the Fifth Conf. on Artificial Intelligence Applications, Miami, USA, 6-10 March, 1989.

[18] Koninklijke Landmacht (Royal Netherlands Army), *VS30-1, Organization and Behavior of the Sovjet Armed Forces* (in Dutch), 1987.

[19] Smith D.E., *Controlling Inference*, Department of Computer Science Stanford University, Report No. STAN-CS-86-1107, April, 1986.

[20] Smith D.E., *Controlling Backward Inference*, Artificial Intelligence 39, 1989, p. 145-208.

[21] Abrahamson B., et al., *Uncertainty management in expert systems*, IEEE Expert, Apr. 1990, p. 29-41.

[22] Duda R.O, Hart P.E, Nilsson N.J., *Subjective Bayesian Methods for Rule-based Inference Systems*, Technical Note 124, Artificial Intelligence Centre, SRI International, Menlo Park, 1976.

[23] Shafer G., *A mathematical theory of evidence*, Princeton, Princeton University Press, 1976.

[24] Zadeh L.A., *Fuzzy Logic*, Computer, April 1988, p83-93.

[25] Hellendoorn J., *Reasoning with fuzzy logic*, Doctor's thesis TU Delft, also appeared as NLR TP 91024 U, 1991.

[26] Doyle J., *A Truth Maintenance System Massachusetts Institute of Technology*, Artificial Intelligence Laboratory, Cambridge, MA, U.S.A., Artificial Intelligence 12, 1979, p. 231-272.

[27] de Kleer J., *An Assumption based Truth Maintenance System*, Artificial Intelligence 28, pp. 127-162, 1986.

[28] Roos N., *What is on the machine's mind? Models for reasoning with incomplete and uncertain knowledge*, Doctor's thesis TU Delft, also appeared as NLR TP 91099 U, 1991.

[29] Lukaszewics W., *Non-monotonic reasoning: formalization of commonsense reasoning*, Ellis Horwood Limited, England, 1990.

[30] Reinfrank M., et al., *Kapri - A Rule-Based Non-Monotonic Inference engine with an Integrated Reason Maintenance System*, SEKI-Report SR-86-03, University of Kaiserslautern, 1986.

[31] Reinfrank M., *Fundamentals of Truth Maintenance*, Linkoeping Studies in Science and Technology Dissertation, no. 221, Linkoeping, 1989.

[32] Allen J.F., *Towards a General Theory of Action and Time*, Artificial Intelligence 23, p123-154, 1984.

[33] McDermott D., *A Temporal Logic for Reasoning about Processes and Plans*, Cognitive Science 6, p101-155, 1983.

[34] Shoham Y., *Temporal Logics in AI: Semantical and Ontological Considerations*, Artificial Intelligence 33, p89-104, 1987.

[35] Zuidgeest R.G., *Introduction to the Theory and Application of Temporal Reasoning*, NLR Technical Report (to be issued), 1994.

[36] Joubel C., Raiman B., *How Time Changes Assumptions*, ECAI 90, Proc. of the 9th European Conference on Artificial Intelligence, p378-383, 1990.

[37] Schang T., Fages F., *Real-Time Expert System for Onboard Radar Identification*, Thomson-CSF, 1988.

[38] Lizza C., Friedlander C., *The Pilot's Associate: A Forum for the Integration of Knowledge Based Systems and Avionics*, IEEE, p. 1252-1259.

[39] Demazeau Y., Muller J.P., *Decentralized AI*, Proceedings of the First European Workshop on Modelling Autonomous Agents in a multi-agent world, Cambridge, England, Aug. 16-18, 1989.

[40] Durfee E.H., *Cooperative Distributed Problem Solving*, The Handbook of Artificial Intelligence, Vol. IV, Addison-Wesley Publishing Company, 1989.

[41] Lesser V.R., Corkill D.D., *The Distributed Vehicle Monitoring Testbed*, AI Magazine, 1983.

[42] Polat F. et al., *Distributed Conflict Resolution among Cooperating Expert Systems*, Expert Systems, Vol. 10, No. 4, November 1993.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-AG-337	ISBN 92-836-0031-2	UNCLASSIFIED/ UNLIMITED

5. Originator

Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly-sur-Seine, France

6. Title

Multi-Sensor Multi-Target Data Fusion, Tracking and Identification Techniques for Guidance and Control Applications

7. Presented at/sponsored by

8. Author(s)/Editor(s)

Multiple

9. Date

October 1996

10. Author's/Editor's Address

Dr. David F. Liang
Head, Space Systems and Technology
Defence Research Establishment Ottawa
Department of National Defence
Shirley Bay, Ottawa K1A 0Z4
Canada

11. Pages

304

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other AGARD unclassified publications is given on the back cover.

13. Keywords/Descriptors

Weapon systems	Inertial navigation
Target acquisition	Inertial guidance
Tracking (position)	Target classification
Data fusion	Target recognition
Stealth technology	Image processing
Signal processing	Simulation
Detectors	Performance evaluation

14. Abstract

Advances in stealthy, high speed, accurate weapon and target systems have imposed stringent requirements on the performance of advanced surveillance, detection, tracking, identification, and classification systems to support military guidance and control applications. The utilization of any single sensor will no longer be sufficient to cope with increasingly demanding operational requirements and challenging mission environments.

This AGARDograph provides an overview of the multi-sensor, multi-target tracking (MS/MTT) techniques and technology, with emphasis towards practical implementations. The areas that are being dealt with are:

— Multi-Sensor Phenomenology and Sensor Signal Processing;
— Data Association and Tracking Techniques;
— Pixel and Symbol Level Image Fusion, Target Classification and Recognition;
— Simulation and Performance Evaluations;
— Data Fusion for Guidance and Control Applications.

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie (1)47.38.57.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

AGARD NON CLASSIFIEES

Aucun stock de publications n'a existé à AGARD. A partir de 1993, AGARD détiendra un stock limité des publications associées aux cycles de conférences et cours spéciaux ainsi que les AGARDographies et les rapports des groupes de travail, organisés et publiés à partir de 1993 inclus. Les demandes de renseignements doivent être adressées à AGARD par lettre ou par fax à l'adresse indiquée ci-dessus. *Veuillez ne pas téléphoner.* La diffusion initiale de toutes les publications de l'AGARD est effectuée auprès des pays membres de l'OTAN par l'intermédiaire des centres de distribution nationaux indiqués ci-dessous. Des exemplaires supplémentaires peuvent parfois être obtenus auprès de ces centres (à l'exception des Etats-Unis). Si vous souhaitez recevoir toutes les publications de l'AGARD, ou simplement celles qui concernent certains Panels, vous pouvez demander à être inclu sur la liste d'envoi de l'un de ces centres. Les publications de l'AGARD sont en vente auprès des agences indiquées ci-dessous, sous forme de photocopie ou de microfiche.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordonnateur AGARD-VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Directeur, Services d'information scientifique
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Goddard Space Flight Center
Code 230
Greenbelt, Maryland 20771

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
Netherlands Delegation to AGARD
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
2700 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

Le centre de distribution national des Etats-Unis ne détient PAS de stocks des publications de l'AGARD.

D'éventuelles demandes de photocopies doivent être formulées directement auprès du NASA Center for AeroSpace Information (CASI) à l'adresse ci-dessous. Toute notification de changement d'adresse doit être fait également auprès de CASI.

AGENCES DE VENTE

NASA Center for AeroSpace Information
(CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
Etats-Unis

The British Library
Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Les demandes de microfiches ou de photocopies de documents AGARD (y compris les demandes faites auprès du CASI) doivent comporter la dénomination AGARD, ainsi que le numéro de série d'AGARD (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Veuiller noter qu'il y a lieu de spécifier AGARD-R-nnn et AGARD-AR-nnn lors de la commande des rapports AGARD et des rapports consultatifs AGARD respectivement. Des références bibliographiques complètes ainsi que des résumés des publications AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
publié par la NASA Scientific and Technical
Information Division
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements and Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 22161
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

REGISTRY
-8 JAN 1997
DRIC

Imprimé par le Groupe Communication Canada
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

AGARD

NATO OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Telefax (1)47.38.57.99 • Telex 610 176

UNLIMITED

DISTRIBUTION OF UNCLASSIFIED

AGARD PUBLICATIONS

AGARD holds limited quantities of the publications that accompanied Lecture Series and Special Courses held in 1993 or later, and of AGARDographs and Working Group reports published from 1993 onward. For details, write or send a telefax to the address given above. *Please do not telephone.*

AGARD does not hold stocks of publications that accompanied earlier Lecture Series or Courses or of any other publications. Initial distribution of all AGARD publications is made to NATO nations through the National Distribution Centres listed below. Further copies are sometimes available from these centres (except in the United States). If you have a need to receive all AGARD publications, or just those relating to one or more specific AGARD Panels, they may be willing to include you (or your organisation) on their distribution list. AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
2700 Amadora

SPAIN
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Goddard Space Flight Center
Code 230
Greenbelt, Maryland 20771

The United States National Distribution Centre does NOT hold stocks of AGARD publications.
Applications for copies should be made direct to the NASA Center for AeroSpace Information (CASI) at the address below. Change of address requests should also go to CASI.

SALES AGENCIES

NASA Center for AeroSpace Information
(CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
United States

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Requests for microfiches or photocopies of AGARD documents (including requests to CASI) should include the word 'AGARD' and the AGARD serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Note that AGARD Reports and Advisory Reports should be specified as AGARD-R-nnn and AGARD-AR-nnn, respectively. Full bibliographical reference [illegible]

RA&I)
rvice

341106++P+UL

Pri[illegible] Group
45 Sacré-[illegible]ada K1A 0S7

UNLIMITED

ISBN 92-836-0031-2