Chapter 3 – SENSOR SYSTEMS FOR GENERATING 3D POINT CLOUDS

This chapter provides a review of the different sensors and sensing techniques that can be used to collect 3D point clouds and aid the extraction of features. Two different physical processes have been exploited to measure the range and consequently determine 3D position: triangulation and time-of-flight. Time-of-flight sensors are active (electromagnetic or acoustic) and measure the time taken for a pulse to travel from the source to the feature of interest and back to the sensor. Triangulation is based on geometry and can be active or passive; it requires accurate registration of two positions separated ideally by a large baseline. As a general rule of thumb triangulation can provide accurate range information to approximately 20 times the baseline and as such is of more value for operation at relatively short range. Active systems have greater applicability to operation at longer range but with added complexity.

3.1 PASSIVE SENSING

Stereo electro-optic imaging sensors can be used to collect imagery which can be registered and used to extract range information. Although stereo imaging cameras are becoming commercially available [8] they are not designed to extract point clouds but rather for direct 3D display using stereo imaging systems exploiting polarization, colour (anaglyphs) or viewing angle (autostereoscopic).

A passive stereo imaging system has a choice of sensors with visible and near-infrared cameras (CCDs or CMOS) offering the best angular resolution but unless amplified by using for example electron multiplication (EMCCD) then these sensors can only be operated in the day-time under solar illumination. Infrared sensors operating in the midwave (3 – 5 µm) or longwave (8 – 12 µm) offer greater day/night capability and superior penetration through atmospheric obscurants such as fog and mist. However practical demonstrations of infrared stereo are much less common, simply because of the increased cost of the sensors. Moving to longer wavelengths there is potential to use RADAR imaging to provide an all weather capability and even image inside buildings; progress in these areas is discussed below.

The automatic extraction of features can benefit from the different imaging modes. The simplest discriminant is of course intensity which can be used to segment different features in a video stream. Colour (or frequency) is another common discriminant with RGB cameras in the visible waveband widely used and multi-band infrared cameras becoming available. Extending the spectral theme, hyperspectral cameras can discriminate fine spectral differences and aid the process of feature extraction. Polarisation is a further process that can be exploited. Natural illumination is polarized and depending on surface structure and angle of reflection many objects will retain this polarization. It is well known that polarisation can be used to discriminate between man-made objects such as vehicles and natural objects such as bushes and trees. However possibly the most powerful discriminate for automatic feature extract is position. Position in two planes (or angles) is easily measured using a focal plane array but the third dimension (range) is a much more difficult challenge and is the topic for the remainder of this chapter.

3.1.1 3D Point Clouds from 2D Images

Most passive systems work analogously to the human visual system. The location of 3D objects is derived from parallax, i.e., small changes in the direction from sensors to object. Since depth information is lost when taking a picture, one cannot exploit a single image without additional information. A single sensor has to move in order to determine the 3D structure by triangulation or one needs a second sensor. Passive imaging sensors measure directions, not distances. Therefore, the objects can be reconstructed only up to a spatial similarity transformation, which is specified by seven degrees of freedom (three translation parameters, three rotation parameters and one scale parameter). Using two or more synchronized cameras

mounted on a stereo rig offers the possibility to introduce the scale by measuring the length of the base line used.

For the orientation of the images and the determination of 3D coordinates of objects, mathematical models of the scene and of the sensor(s) are needed. Usually these models assume that the observed scene is static and that the extracted 3D points lie on opaque, non-specular surfaces. Concerning the camera, one typically assumes a straight-line-preserving projection, that is, straight lines in the world, e.g., building edges, are mapped to straight line segments in the image. Furthermore, it is assumed that a unique projection centre exists, so that the light rays between object and image points pass through a single point, i.e., perspective geometry is present, cf. Figure 3-1. To ensure the compliance of this camera model, it is advisable to perform sensor calibration either by a lab calibration or, if possible, by self-calibration on the fly.

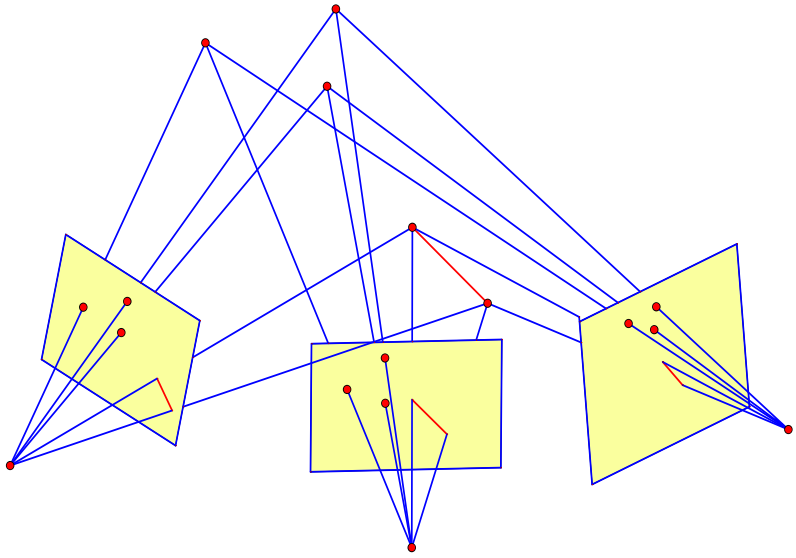

Figure 3-1: Triangulation Principle Depicted Using Three Images Captured from Two Different Viewpoints.

Approaches that automatically acquire 3D points and/or 3D line segments from sequences of images at unknown locations, using projective geometry and automatic correspondences, are available (see standard textbooks [9], [10], [11] for instance). For convenience, the presentation in what follows is restricted to 3D points. The general procedure is first to establish correspondences between image features of consecutive images of a sequence. These correspondences are then the base for the subsequent parameter estimation step that determines the positions of the 3D points and the poses of the moving camera simultaneously. Figure 3-1 depicts the situation for two images captured from two different viewpoints. Given the poses for the two camera viewpoints, the 3D points can be determined by triangulation (spatial forward sectioning, corresponding rays intersect at the 3D point sought). Conversely, given the positions of the 3D points, the poses of the camera can be determined by spatial resection. Since neither the camera poses nor the 3D point position are known, both parameter groups have to be determined simultaneously. Techniques to do this are known as Structure from Motion (SfM) in computer vision, Simultaneous Localization And Mapping (SLAM) in robotics and bundle adjustment in photogrammetry.

Basically, the correspondences of image features can be supplied in two ways: Given automatically extracted image points in one image (interest or salient points), the corresponding image points in a second image can be determined by feature tracking (see, e.g., [12]). These methods require images taken from viewpoints close to each other (short baseline) in order for the automatic correspondences to work, which makes them noise-sensitive and numerically unstable. Alternatively, the image features can be extracted in all images

separately and then the correspondences established by matching. These methods (see [13]) have attracted much attention especially for wide-base-line applications.

Although bundle adjustment yields an optimal solution, it is not suitable for large data sets due to limitations in computational resources. A workaround is the application of sliding window approaches for filtering [14], which allows online processing. Figure 3-2 shows three representative images from an infrared image sequence captured by an airborne platform (for details refer to Section 4.4). The extracted and plotted interest points correspond to hot and cold spots respectively [15],[16]. The estimation results for the corresponding point cloud and the camera poses are shown in Figure 3-3.

Figure 3-2: Three Representative Infrared Images Showing an Urban Area with Plotted Interest Points.

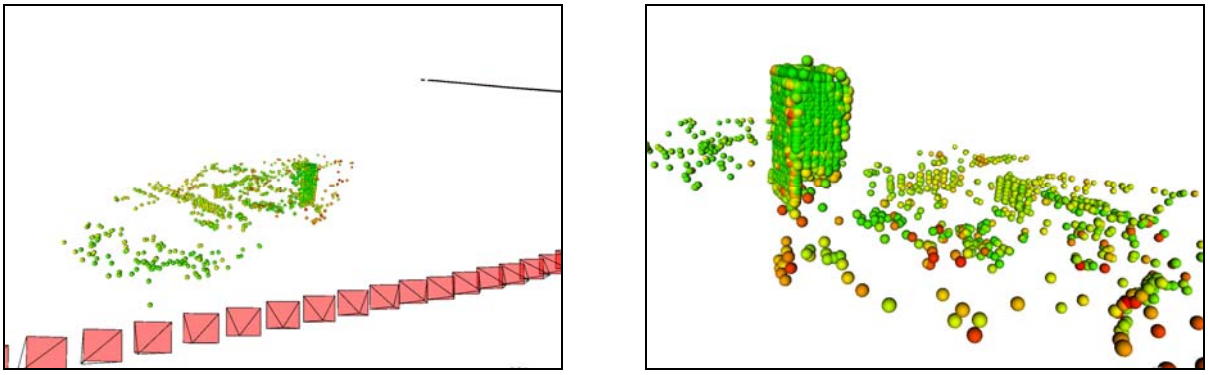

Figure 3-3: Two Views of the Reconstructed Urban Scene. The colours of the 3D points refer to positional precision (green = more precise, yellow = less precise, red = still less precise); the tetrahedra in the foreground depict the estimated poses of the used camera.

The spatial distribution of the 3D points depends on the image contents. Usually one gets a sparse representation with a heterogeneous distributed point cloud. This cloud of 3D points has no structural information among the points. Specifically, there is no topological information to represent connectivity among points. In geometric modelling, it is necessary to convert this cloud of points into a surface representation such as a mesh model. (cf. Section 4.3). Approaches to obtain a dense surface representation have gained much attention recently; cf. [17] for instance. For derivation of these digital surface models or elevation models, one needs smoothness assumption about the surface.

3.2 ACTIVE SENSING

Time-of-flight sensors are the most common active sensors but before describing these it is worth noting that triangulation can also be used in active mode to measure range. If the baseline between the emitter and detector are accurately know then the angle of arrival can be measured and the range calculated.

An excellent review of active triangulation and a comparison against time-of-flight systems is given in reference [18]. The advantage over purely passive systems (structure from motion) is that the point of interest is clearly defined by the laser spot, which can be scanned to construct a 3D image. However the range precision deteriorates with the square of the range to the target so these systems are best suited to relatively short ranges (out to 30 m) and the system complexity makes the purely passive system more attractive. For operation out to many kilometers and to remove the correspondence and disparity problems associated with stereo vision [19] time-of-flight systems need to be employed.

3.2.1 Time-of-Flight Systems

The pulsed TOF LADAR system transmits a laser pulse of short duration to the target of interest and measures the time for the reflected pulse to return to the sensor. Given a measured round-trip time for the pulse, t and the speed of light, c, the range to the object point is calculated using:

$$R = \frac{ct}{2} \tag{3.1}$$

The important system parameters can be understood by considering the energy incident on the target and relating it to the energy returned to the imager. The basic physics behind a TOF system is described by the laser RADAR equation [20]:

$$E_R = \frac{4KE_T T}{\pi \phi^2 R^2} . \Gamma . \frac{T}{4\pi R^2} . \frac{\pi D^2 \eta_R}{4} \tag{3.2}$$

where the energy returned to the sensor (E_R), is dependent on the four processes captured by the four parts of equation 3.2. The first part describes the energy density at the target, which is dependent on the transmitted energy (E_T), the atmospheric transmission (T), the beam divergence (ϕ – beam width in radians) and the range, R. K is a factor to account for the beam profile at the point of interest within the beam. The second physical process is the reflection of the radiation from the target and this is captured by the laser cross-section, Γ. The scattered light propagates back towards the sensor with an area that expands as R^2. Finally the reflected energy is intercepted by the objective lens of the sensor (Diameter, D). η_R is the receiver efficiency and includes the losses associated with system transmission and those dependent on the detection technique, which could include loss in coherence or polarisation if appropriate.

The equation above is used if the target is small with respect to the field of view of the sensor. However, in the case of interest here, the laser divergence is chosen to match the detector field of view. For a scanned system with a single detector the laser beam quality is important and a diffraction-limited beam may be needed. In a flash illumination system with a 2D array of pixels the laser beam can have poorer spatial mode quality. If Lambertian (isotropic) scattering is assumed and the laser and sensor are coincident then equation 1 can be reduced to:

$$E_R = \frac{E_T T^2 \rho D^2 \eta_R}{4R^2} \tag{3.3}$$

where ρ is the total hemispherical target reflectance. This equation clarifies the dependence on the square of the atmospheric transmittance, objective diameter and the inverse square dependence upon range.

If a detector array is used then the energy in a pulse is shared between the pixels, which necessitates high energy, whereas for a single detector system a high repetition rate source is needed to provide a rapid scan. Typically the laser power for both systems will be comparable but delivered in a different format.

3.2.1.1 2-Dimensional (Flying Point) Scanning Systems (LIDAR)

Light detection and ranging (LIDAR) or laser detection and ranging (LADAR) sensors have been developed that exploit various different detection and modulation processes. Direct detection LIDAR is by far the most common but frequency and amplitude modulation techniques have also been developed and demonstrated to great effect.

Direct detection systems transmit a short laser pulse (~3 ns) and digitise the returned signal; the time delay is used to measure the range to the target. With a fully digitised return signal processing techniques can be used to extract not just the range to the target but the structure within the pixel. For example first and last pulse return can be used to measure the height of a forest canopy and the depth of the forest floor (for further discussion see Section 3.2.2). Typical performance figures for some airborne LIDAR systems are tabulated in Table 3-1. Modern airborne systems operate at a pulse repetition frequency of around 100 kHz, where the repetition frequency is reduced as the range is extended. Typical systems operate with an average power of a few watts and can achieve a range out to a few kilometres dependent on the reflectivity of the target material. Generally, the trend is towards eye-safe wavelengths around 1.5 µm in order to operate with higher power and extend the range. Range precision down to a few centimeters is possible with absolute accuracy of ~10 cm more realistic. Beam divergence and pointing accuracy typically limit the angular resolution to 0.1 mrad (10 cm at 1 km).

Table 3-1: Provides an Overview of Commercial Airborne LIDAR Systems and Achievable System Performance. The table is based on information collated in reference [21].

Manufacturer	Airborne Hydrography	Leica Geosystems	Optec	Riegl	Topo Sys GmbH
System	Hawk Eye II [22]	ALS60 [21]	Orion [23]	LMS – Q680 [24]	Falcon II [25]
Wavelength	532 nm	1064 nm	1064/1541 nm	1550 nm	1560 nm
Pulselength	5 ns	5 ns	< 7 ns	< 4 ns	5 ns
Pulse Frequency	64 kHz	20 – 200 kHz	50 – 200 kHz	80 – 400 kHz	83 kHz
Echoes/Pulse	4	4	4	Full waveform at 1 GHz	9
Range Precision	3 cm	3 – 4 cm	< 5 cm	4 cm	1 cm
Spatial Precision at 1 km Range	50 cm	20 cm	25 cm	10 cm	20 cm
Operational Height	500 m (typical)	200 m to 5 km	200 m / 1 km / 2.5 km	1.0 – 2.5 km	1.6 km

One of the rapidly growing markets for LIDAR technology is ground-based mobile mapping whereby a LIDAR system is mounted on a moving vehicle and operated in a 360 degree scanning mode. Many mobile services [26] and systems are becoming available such as Lynx for Optech [23] and HDL-64E from Velodyne [27]. These systems offer vertical coverage up to ~30 degrees with a distance accuracy of ~2 cm out to a range of approximately 100 m. They are ideal for imaging the façade of buildings and other areas inaccessible to airborne survey. At the cheaper end of the market the LD-LRS series of sensors from SICK [28] are very popular and were used extensively in the DARPA urban grand challenge for autonomous navigation.

Frequency modulated systems measure range by impressing a linear frequency chirp upon the transmitted beam then mix the returned signal with a local replica to determine the frequency shift and hence the range.

This technique has been demonstrated to great effect with 1.5 µm fibre laser technology. Erbium fibre laser technology is attractive for LIDAR applications as it promises a compact, reliable laser source however direct detection systems need short pulses and the high peak power is incompatible with the small fibre core. QinetiQ in the UK avoided this problem be developed a chirped pulse system that achieves excellent range precision with a relatively long frequency chirped laser pulse. This approach avoids the high peak powers that would otherwise damage the fibre allowing them to operate at considerably longer range than traditional LIDAR systems [29]. A representative image collected at a range of 18 km is shown in Figure 3-4.

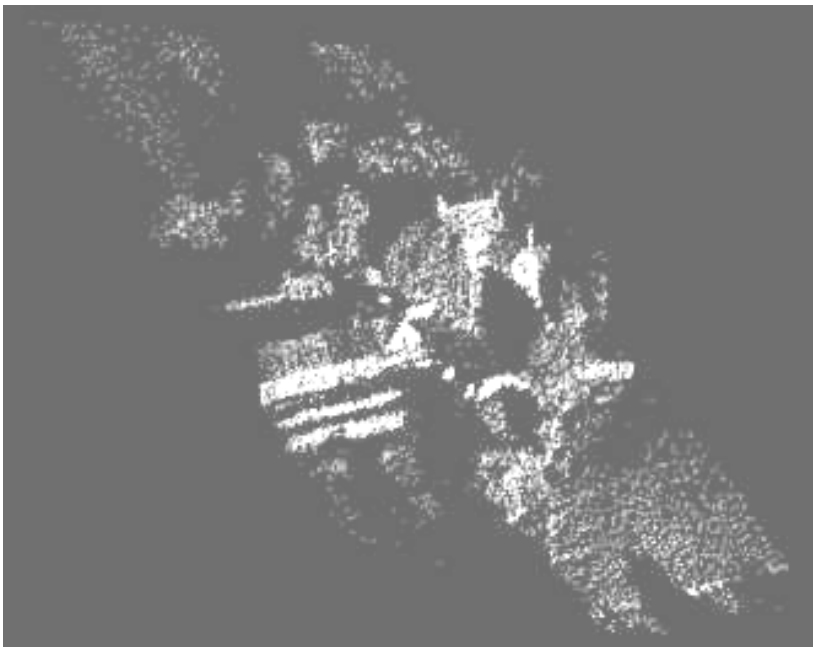

Figure 3-4: Farm House and Outbuildings at 18 km (Ground-Ground Demonstration) – Reproduced with the Permission of QinetiQ Limited.

Amplitude modulated systems transmit a beam in which the signal level is modulated as shown in Figure 3-5. The phase change, θ between the transmitted beam and received signal is measured and the range to the target is calculated using:

$$R = \frac{\theta\lambda}{4\pi} \tag{3.4}$$

where λ is the wavelength associated with the modulation (rather than the carrier). LIDARs based on amplitude modulation provide excellent results for short range operation but ambiguous results are obtained at range intervals equivalent to a 360 degree phase change. For example with a modulation frequency of 10 MHz unambiguous results are obtained out a range of 15 m ($\lambda/2$ as the beam must travel to and from the target) but targets at 5 m and 20 m will be ambiguous. Various techniques have been developed to mitigate the range ambiguity such as the use of multiple frequencies whereby the range ambiguity is extended to the lowest common multiple. Focal plane arrays based on the amplitude modulation technique are becoming widely available and will be discussed below.

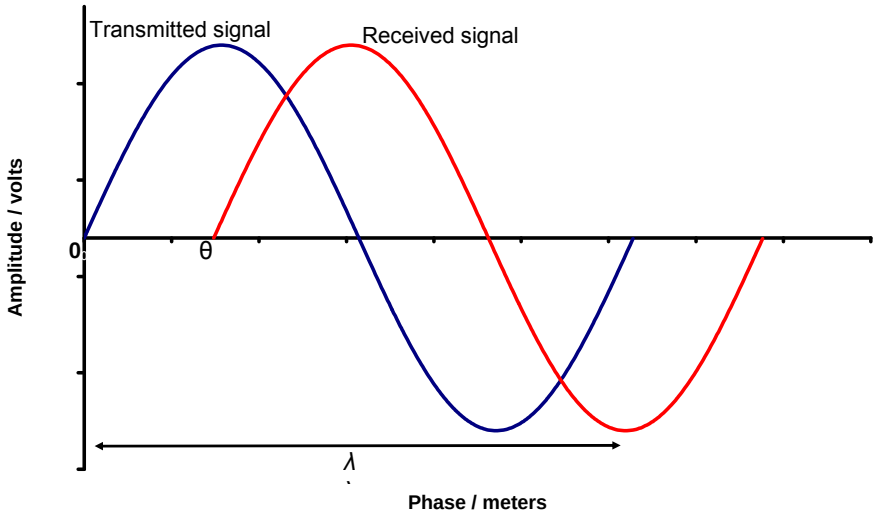

Figure 3-5: Transmitted (blue) and Received Signals in an Amplitude Modulated LIDAR System.

3.2.1.2 1-Dimensional Scanning Systems

A natural progression from the two dimensional LIDAR scanning is to scan in only one dimension and use a sensor array. This mode of operation has the potential to achieve a significant increase in the area coverage rate from these airborne scanning systems. A 1-dimensional scanning system can be realised in a number of different ways:

- A linear array of detectors and scanning either azimuth or elevation.
- A gated focal-plane array in which the position of the gate can be scanned with respect to transmitted laser pulse.

In the gated imager a short laser pulse illuminates the target and a gated camera (possibly intensified) is used to capture range slices. Andersen et al. [30] have used laser pulses of 500 ps duration at 532 nm and a gated and intensified camera to capture facial profiles. Relative range accuracy down to 1 mm was achieved at distances up to 500 m.

There are a number of examples of temporal scanning to construct a 3D point cloud however very few examples of LIDAR systems operating with linear detector arrays in a line scan-mode. This is quite surprising as there are many technical challenges in moving from a traditional single-point LIDAR system to a staring 3D imager and a line-scan system could simplify many of these problems. Reducing the physical size of timing circuitry such that it can fit into a pixel is a significant challenge. In a linear array there is extra space that can be used to control the timing circuitry and extract the signal from the pixel. Increasing the laser energy from that required for spot illumination to burst illumination is also quite challenging. Illumination over a strip could be more easily achieved. The linear array only makes sense if the other dimension is provided by platform or sensor motion but in many of the applications of interest (airborne LIDAR, mobile mapping) this is exactly how the system would be operated.

3.2.1.3 3D Staring Arrays (Flash LIDAR)

In a 3D staring array each pixel in the array captures the range to the target. This is challenge both for the illumination source which must illuminate a large area and for the focal plane array designer who must include timing circuitry within each tiny pixel.

One of the first and most impressive demonstrations was achieved under the DARPA JIGSAW programme [31] which demonstrated 3D imaging of a vehicle under a thick forest canopy by exploiting the small gaps between the leaves and image registration from multiple viewing angles. The JIGSAW sensor is based on a 32 x 32 focal plane array in which each detector is operated in the Geiger avalanche mode such that a small signal (approaching a single photon) returned from the scene triggers the avalanche process and stops the timer. The pixel size at 30 µm diameter is relatively small but the pixel pitch at 100 µm means that the sensor has a poor fill factor. To compensate for the low fill factor the laser illuminator consists of an array of beams (32 x 32) generated from a single beam using a diffractive optic. The system operates with a frequency doubled Nd:YAG laser operating at 532 nm with pulse duration of 300 ps and repetition rate of 16 kHz (~16 million pixels/sec). A range resolution of 40 cm has been demonstrated.

The JIGSAW programme achieved very impressive results but from relatively short ranges (~200 m). Other groups have focussed on a higher density of pixels on the focal plane array and operation at longer range, which generally means operating with an eye-safe laser source.

ASC (Advanced Scientific Concepts Inc [33]) has developed a 3D staring array camera system (TigerEye 3D). The advanced Readout Integrated Circuit (ROIC) incorporates both analogue and digital processing on each pixel to capture the leading edge of the returned signal and sample the returned waveform. Each pixel can achieve 15 cm range precision reducing to a few centimetres after processing the waveform. The ROIC has been hybridised with InGaAs avalanche photodiodes and operates in the eye-safe waveband at ~1.5 µm. The detector array has been implemented in a 128 x 128 array on a 100 µm pitch. The active pixels are smaller than the pitch at 20 µm to reduce the noise so a lenslet array is used to improve the quantum efficiency. This removes the requirement for an array of laser beams and the system can be used with burst illumination laser. The detector array has been implemented in the camera system shown in Figure 3-6 and used to collect 3D movies, an example image is also shown in Figure 3-6. The system operates out 1 km range using pulses of 7 mJ at 30 Hz [32] which corresponds to a transmitted power of 0.2 W for 500 x 10^3 voxels/sec. A comparison of scanning LIDAR versus 3D flash LIDAR has concluded that the flash imaging can be up to 4 times quicker [34] for equivalent transmitted power levels.

Figure 3-6: 3D Flash Illumination Imaging System and Sample 3D Image of a Truck.

Selex in the UK has been developing a 3D time-of-flight focal plane array based on avalanche photodiodes in MCT [35],[36]. The goal is to develop a long range 3D imaging capability that can supplement thermal imaging systems and provide enhanced resolution at long range. MCT was chosen as the diode material because it can be operated in avalanche gain mode in the eye-safe waveband with relatively low voltages and it has a very high electron-to-hole mobility ratio which provides nearly ideal (noise-free) avalanche gain. Operation in avalanche mode is needed to achieve the range performance and achieve compatibility with airborne laser designator technology. The requirements drive the specification towards small pixels and

operation with laser pulses typically 10 – 20 ns. A novel timing circuit has been developed to enable a demonstration with pixels of 24 µm diameter.

A focal plane array has been developed in half TV format (320 x 256) camera, integrated with a big-sky laser and demonstrated in outdoor trials. The timing circuitry on each pixel responds to the centre-of-gravity of the laser pulse and range precision of less than 30 cm can be achieved. A picture of the 3D laser imaging rig and an example of the imagery collected is shown in Figure 3-7. The current technology is based on Gen II (loop-hole) CMT technology but research is progressing towards Gen III technology which will enable larger format arrays (full TV) and multi-functional performance; mid-wave thermal imaging and 3D range measurements on a single focal plane array.

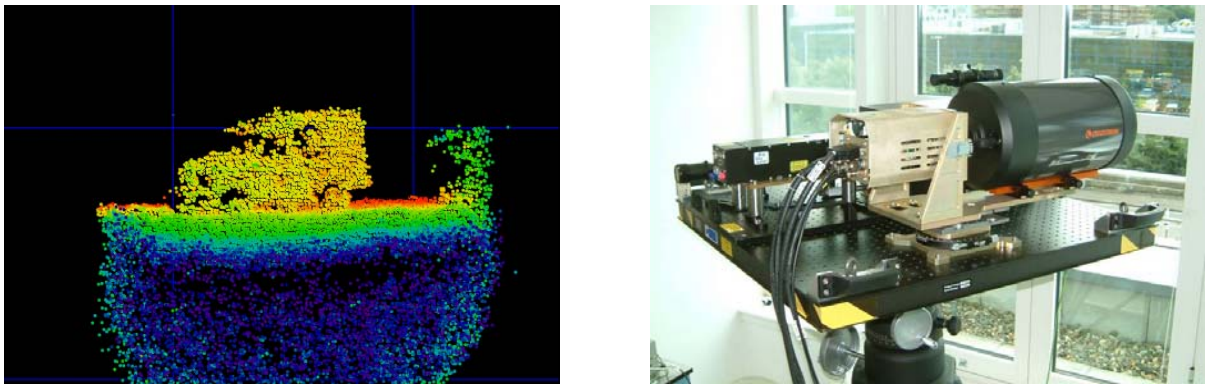

Figure 3-7: UK 3D Laser Imaging Camera System and Sample Imagery of Land Rover.

There has been a flurry of development activity on 3D cameras based on amplitude modulated time-of-flight focal plane arrays. Detector array developments, that allow the array response to be modulated at the same frequency as the illumination source, have greatly simplified the system design allowing relatively large arrays (VGA) with small pixels (40 µm) to be produced. Typical modulation frequencies of ~15 MHz are used to achieve range accuracy below 1 cm out to a depth of approximately 7 m. The technology is focused towards the vehicle safety and the computer gaming launch and current cameras operate with typical integration times of 10 ms and over a wide field-of-view (up to 60°). Of particular note is the acquisition of 3DV systems but Microsoft [37] which has plans to market a new 3D gaming environment with its Xbox 360 called Kinect [38] with integrated RGB camera and microphone array. This 3D imaging technology is designed for short range operation but with suitable modifications to the optics, modulation frequency and illumination source it should be feasible to exploit this new class of sensors for ground-based urban mapping at longer range.

Table 3-2: Comparison of Performance Parameters for Amplitude-Modulated Camera Systems.

Company	3DV Systems	Canesta	Mesa-imaging	PMD Technologies	Primesense
Camera	Zcam		SR4000	Camcube 2.0	PrimeSensor
Array Size	640 x 480	160 x 120	176 x 144	204 x 204	640 x 480
Pixel Pitch			40 µm	60 µm	
Operational Range	0.5 – 2.5 m	0.3 – 1.5 cm	0.8 – 8.0 m	7 m	0.8 – 3.5 m
Modulation Frequency			15 MHz	18 – 21 MHz	
Resolution	1 – 2 cm	0.3 – 1.5 cm	1.5 cm	3 mm	1 cm
Frame Rate	60 fps		54 fps	25 fps	
Illumination Wavelength		870 nm	850 nm	870 nm	

3.2.1.4 Through-Wall Synthetic Aperture RADAR

The principle of operation of a Through-Wall Synthetic Aperture RADAR (TWSAR) consists in transmitting a RADAR signal towards a scene of interest, a building in this case, and collecting the signal that is reflected back from the wall, and from the targets and clutter behind the wall. One type of setup consists in using separate transmit and receive antennas. These antennas can be mounted on a rail system (Figure 3-8a) or on a vehicle in a side-looking configuration (Figure 3-8b). As the system is moved (along the rail or by driving the vehicle), a chirp signal is periodically transmitted through the transmit antenna towards the wall. The returns collected by the receive antenna are summed coherently to form a synthetic aperture.

a) b)

Figure 3-8: TWSAR Prototype – a) Mounted on Rail; and b) Mounted on a Mobile Lab.

Through-wall RADARs have the potential of mapping interior room layout, including the location of walls, doors and furniture. They could provide information on the in-wall structure and detect objects of interest concealed in buildings such as persons and arm caches. However, imaging through walls presents many technical challenges. Through-wall RADARs must cope with signal transmission and reception through inhomogeneous media and require a large dynamic range. The later requirement results from a desire to detect weak signals coming from behind the wall in the presence of direct, strong signal reflections from the external wall surface.

Typical synthetic aperture RADAR systems produce 2D images that provide range and azimuth information. Figure 3-9 shows one such image. It was acquired by driving the vehicle-mounted TWSAR testbed shown in Figure 3-8b along a building wall [39].

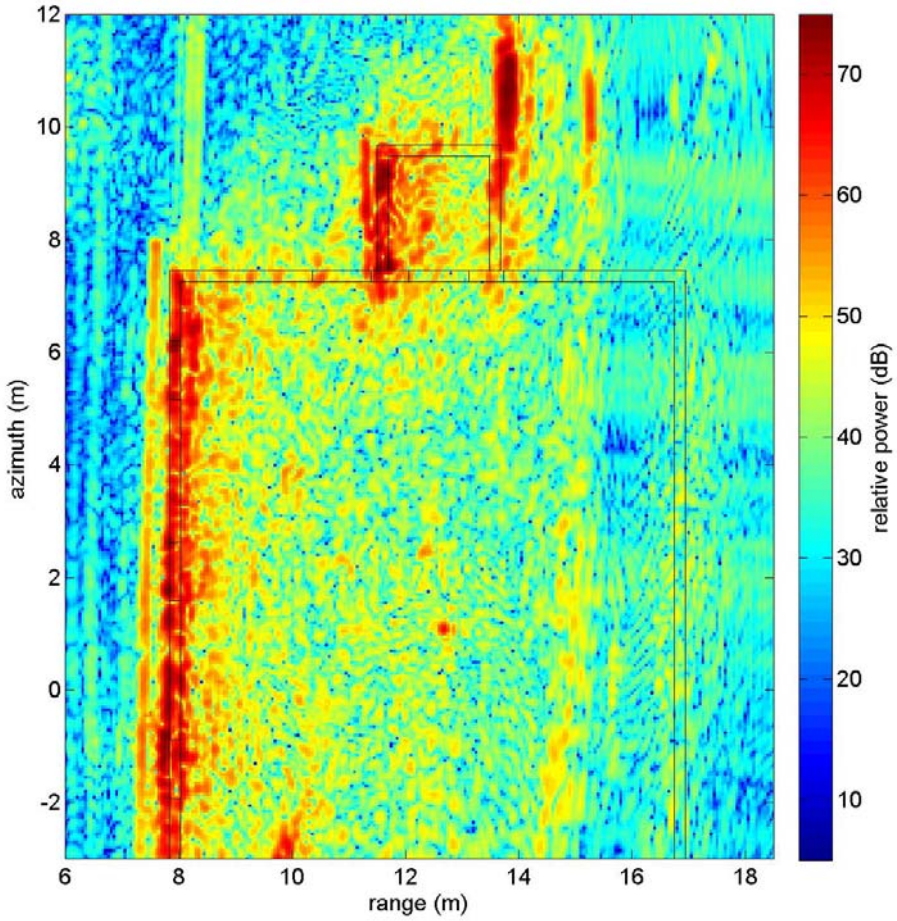

Figure 3-9: 2D TWSAR Image of a Wood Building with Building Layout Outlined in Black.

Research is ongoing to develop 3D TWSAR systems capable of producing 3D images that would provide information in the elevation dimension. Efforts are also made to elaborate wall compensation strategies and to develop tools for target detection and classification [40].

3.2.2 Estimating Distances through Waveform Analysis

In this section we address some issues related to measuring range with LIDAR systems. We will consider two aspects: signal processing techniques for detecting pulses in received waveform signals, and practical issues related to LIDAR measurements of an urban scene.

3.2.2.1 Detecting a Pulse

Here, we will give a short discussion about the basic signal processing aspects on how a LIDAR system operates in terms of extracting pulses and estimating range from the received laser light. As most LIDAR systems used for 3D urban mapping operate using the time-of-flight principle, we will limit the discussion here to such systems. For a comprehensive overview on state-of-the-art within this topic, as well as specifications of several commercial LIDAR systems refer to [41].

In order to describe the principle of operation, first assume that a laser pulse is emitted at time t_0. Furthermore, assume that a sufficiently large proportion of the laser light is reflected off at least one surface in the path of the laser beam and back to the detector. The reflected pulse reaches the detector at time $t_0+\Delta t$. Given the speed of light c, the range can be estimated through $d = c\Delta t$. Obviously, in order to obtain good range data it is crucial that Δt is estimated accurately.

A number of approaches are proposed for solving this task, of which three seem to be especially frequent: peak detection, constant-fraction detection and correlation-based detection. The first two techniques typically work satisfactorily if the Signal to Noise Ratio (SNR) is good, which is why one of them is often found in single-pixel (scanning) systems. As the noise level increases, however, the performance of these techniques degrade and the benefits of using the more computationally demanding correlation-based techniques such as the matched filter become more prominent [42]. In the matched filter, the received signal is correlated with the emitted signal which generally increases the robustness of the range estimates in the presence of noise.

In practice the received signal is a sampled and quantized version of the actual physical signal. The simplest way to estimate Δt is to select the position of a sample as the true position, be it the peak value or the first value above a certain threshold. However, since the true range value is likely to lie somewhere between two samples, this kind of approach will produce an expected range error that is proportional to the distance corresponding to the time between two samples. In order to increase the resolution and accuracy of the range estimate beyond the sampling resolution, some additional computations can – or should – be performed, e.g., through local weighting.

The above-mentioned techniques are based on the implicit assumption that each pulse in the signal is the result of only one reflecting surface, i.e., the pulse corresponds to one range value. In practice, this means that the reflecting surfaces in the path of the laser beam must be well separated in space, so that there is no significant overlap between their respective contributions. Now, assume that there are two reflecting surfaces and that the distance between them is small compared to the pulse width. The result is that the pulses reflected off these surfaces will merge into a significantly broader pulse. Obviously, treating that pulse as if only reflection was present is suboptimal. Not only will there be only range estimate, but the range estimate itself is likely to be quite uncertain. Typically, it will be somewhere between the true surfaces, producing virtual "ghost" points in the void between the surfaces. Such an unwanted effect calls for a more careful analysis of the signal. A possible remedy is to remove points stemming from signals that manifest significant pulse broadening; as the received pulse being significantly broadened compared to the emitted pulse indicates that the laser light has hit many surface elements on its way through the atmosphere (e.g., through a tree canopy). Conversely, if a laser pulse hits a single surface that is perpendicular to the direction of the laser beam, its shape is (more or less) preserved and we could expect the range estimate to be more accurate. In Figure 3-10, an example of pulse broadening is shown.

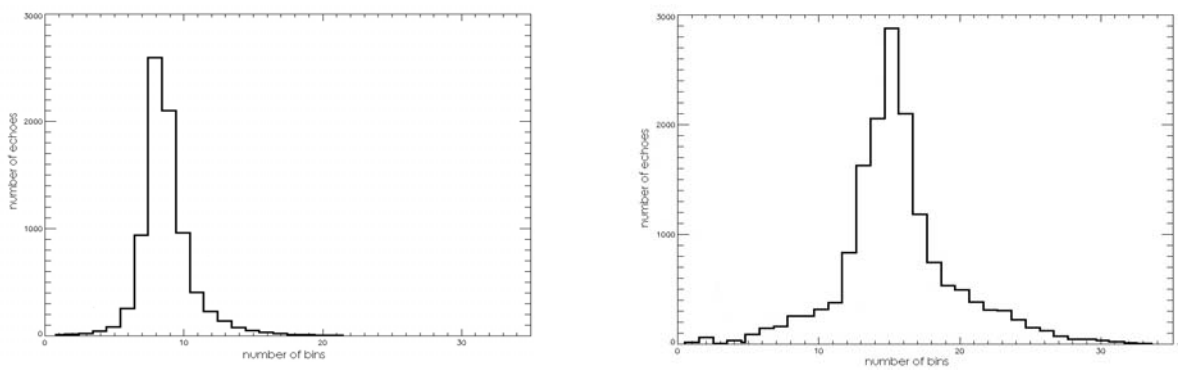

Figure 3-10: Histogram of the Width of Detected Pulses in an Area Containing Buildings (left) and Vegetation (right). Note the general broadening of pulses in the vegetation area due to the superposition of many small reflections from the leaves.

Note also the pulse broadening even occurs at slant incidence towards a single surface. Whether this broadening effect is large enough to cause trouble in practice depends on the size of the footprint of the laser beam and the incidence angle.

Pulses that have been significantly broadened can be subject to more refined analysis aiming at extracting information "hidden" in the signal. Two very useful techniques for this purpose are pulse fitting and deconvolution. The goal in pulse fitting is to find an optimal approximation of the signal using a set of basis functions. In other words, the signal is decomposed into a number of components that (hopefully) correspond to the contributions from the individual surfaces. The basis function is typically defined as the shape of the emitted pulse, or in practice, a nice function sufficiently similar to it.

The received signal $y(t)$ can be described as the emitted signal $s(t)$ convolved with a function that represents the scene $x(t)$. Whereas pulse fitting aims at approximating the signal at hand with a number of basis functions, the essence of *deconvolution* is basically to estimate what the received signal would have looked like if an ideal pulse was emitted, i.e., to estimate $x(t)$. If we are successful in doing so, the resulting signal will contain sharp spikes each of which corresponds to the position of the reflecting surfaces (Figure 3-11). The peaks may then be detected with one of the pulse detection techniques outlined above.

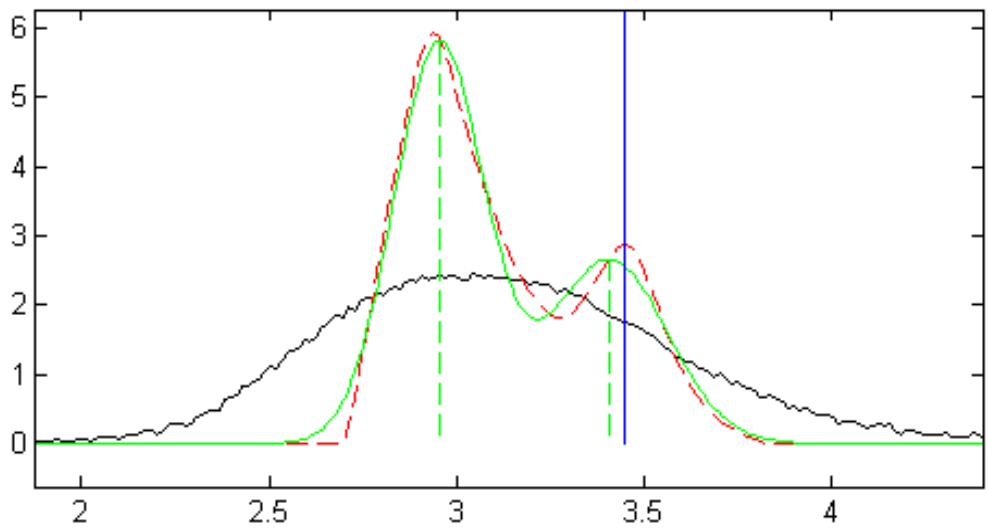

Figure 3-11: Deconvolution Example. The original signal (black) consists of two contributions: one from a hard surface (whose position is shown by the blue vertical line) and one from partially obscuring vegetation. The red dashed curve shows the result after deconvolving the original signal with the function of the emitted pulse. The green curve is the result of fitting two Gaussian functions to the red curve.

3.2.2.2 Measuring Distance in Practice

So far in this section we have discussed various signal processing tools that we have at our disposal for estimating range in the individual waveform signals, from simple peak detection to pulse fitting and deconvolution. Together with data about the pose (position and orientation) of the sensor, obtained from GPS/IMU measurements, we can describe the 3D position of each reflecting surface element in a global coordinate system, resulting in a (huge) point cloud that can be further analyzed (Chapter 4).

Obviously, a prerequisite for any pulse detection technique is that there is a decent signal to work with in the first place; the surface must reflect enough laser light back to the receiver to exceed the noise level. In practice there are factors that sometimes hamper the collection of good data.

One problem concerns certain materials that absorb quite strongly at the laser wavelength and thus often cause data dropouts. Another problem is caused by specular surfaces who act like mirrors reflecting a large portion of the incoming laser light in another, undetectable direction. This can be seen in regions with water, where the number of laser detected hits is generally quite low, especially if the water surface is calm (i.e., mirror-like). Furthermore, internal reflections in certain materials may lead to phase shifts that can result in an attenuation or extinction of the light.

Above, we mentioned the problem of resolving surfaces that lie close to each other. This is typically the case in vegetation and virtually no commercially available system allows for the extraction of individual leaves or branches of trees or bushes. However, LIDAR systems have an inherent ability to "see through" semi-transparent structures, e.g., vegetation, which in this respect often gives them an advantage over passive imaging. For the same reason that the sky can often be discerned through the canopy from underneath a tree, the LIDAR system can "see" the ground from above the tree top. And unlike passive systems that need the same spot on the ground to be visible in multiple images in order to produce a range estimate, the LIDAR system needs only one view on the same spot. Hence, provided that the amount of light reflected from the ground back to the detector is above the noise level, it is fully possible to accurately measure the ground level in vegetation areas.

3.3 REGISTRATION OF 3D POINT CLOUDS

Given the task to create a detailed and accurate 3D model about a certain object (or an entire scene, for that matter), it is desirable to have access to as many accurate and complete measurements of the object as possible. The ability to combine multiple data sets corresponding to the geographical area into one would allow us to obtain a more complete "view" of the object. A typical example from mapping of urban terrain is that data from an airborne platform may be combined with data from terrestrial measurements to capture both the roof and the walls. However, in order to be able to combine several data sets, they have to be aligned with one another. The first idea that may come to mind is to accurately measure the pose (position and orientation) of each sensor and then simply superpose the two datasets using the pose information. In practice, it may be quite risky to rely on such an approach, for at least two reasons. First, the alignment of the data sets can never become better than the accuracy of the pose estimates of the sensors. Second, the accuracy of the pose estimate may be subject to random and significant variations over time, e.g., depending on the quality of the GPS signal, which will make it difficult to assess the quality of the alignment at a given time. Luckily, there are data processing techniques to help in this matter, known as *registration* techniques. The goal of the registration process is to establish the geometrical relationship between multiple data sets so that all data can be represented in a common coordinate system.

As an example, assume that we have a primary point cloud **P** and a secondary point cloud **S** that we want to align with **P**. Normally we will treat **P** and **S** as rigid bodies. Then the registration problem is *to find the rotation matrix **R** and the translation vector **t** that brings* **S** *into the best possible alignment with* **P**.

In the early 1990's, Besl and McKay [43] presented a registration method known as the Iterative Closest Point (ICP), that has become very popular and is probably the most well-known registration method to date. Its main principle of operation is quite straightforward:

1) Pair each point in **S** with the closest point in **P**;
2) Compute the optimal transformation that minimizes the error (MSE) between the paired points; and
3) Apply the optimal transformation to **S** and update the MSE.

These three steps are iterated until a minimum (or a satisfactorily small) MSE is reached. Although very frequently encountered, ICP has a number of shortcomings. For example, it has an inherent tendency of creating a large overlap between **P** and **S** (as that tends to reduce the MSE) which is problematic if only the

parts of the data that actually overlap are relatively small. Since the advent of the original ICP, a number of improvement have been proposed in the literature, e.g., concerning different distance measurement, strategies for closest point search, false match rejection and motion estimation [44].

Another recently suggested solution to the 3D registration problem is based on the normal distribution transform, NDT [45]. Basically, in the 3D-NDT [46] approach the point clouds are divided into a number of cells, or voxels, each of which is described by a normal distribution function representing the probability of finding a data point at a certain position. Not only does this approach mitigate the problem of handling huge data sets (as only a few parameters are needed for each voxel), but it does not depend on computationally demanding nearest-neighbour search of the ICP algorithm.

A registration example is shown in Figure 3-12, where FLASH LIDAR data from Bonnland (part of the SET-118 data pool) have been stitched together with the ICP algorithm. The final result is a joint data set where the position of each LIDAR point is expressed relative to a reference frame (here, the very first frame). The green dots show the estimated position of the LIDAR sensor for each of the acquired frames.

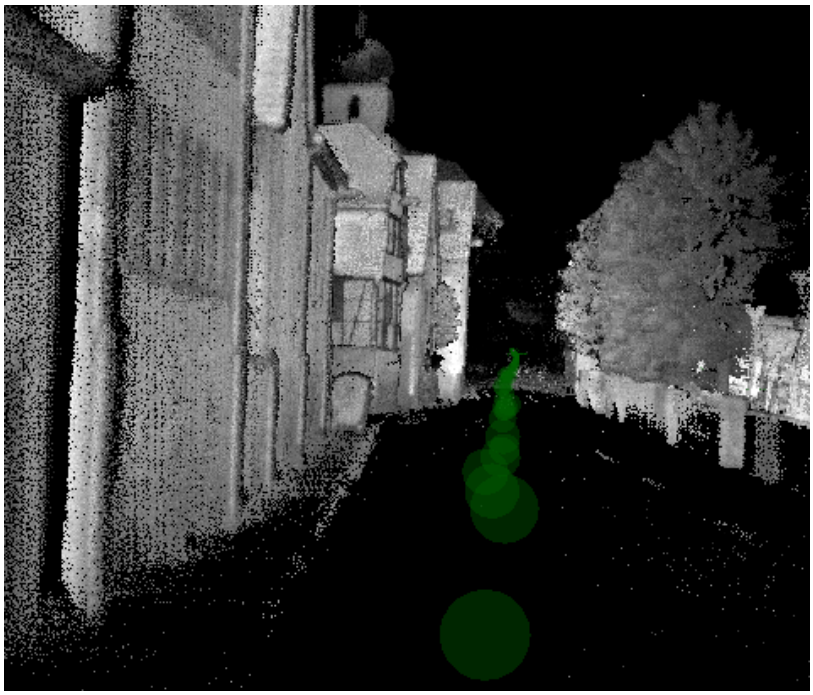

Figure 3-12: A Number of FLASH LIDAR Data Sets Collected from a Car while Driving Down a Street in Bonnland, Registered with the ICP Algorithm. The green dots show the estimated position of the sensor for each frame.

3.4 GEOREFERENCING OF 3D POINT CLOUDS OR MODELS

3D models may be acquired at different time and using different types of sensors. When it is required to represent a set of models in a common reference frame, one solution consists in georeferencing the models. Georeferencing means registering a 3D model against a global coordinates system so as to determine its position on Earth.

One method to georeference a 3D model relies on control points. A control point is a fixed point on Earth (such as a building corner) whose spatial coordinates are measured precisely. To compute the position of a 3D model in a global coordinates system, it is first required to survey the area where the model will be

acquired to get the precise location of a certain number of control points. Correspondences can then be established between control points and their counterpart in the 3D model. The geographic position of the model is obtained by minimizing the position error over all the correspondence pairs. The accuracy of the georeferencing process can be quantified by evaluating the residuals of the minimization procedure.

Another possible solution to georeference a 3D model consists in using an integrated geo-positioning system to compute the trajectory of the vehicle mounted with the sensors used to acquire the model. Typical integrated geo-positioning systems generally combine the use of one or more GPS receivers, a Differential Global Positioning System (DGPS) receiver, an Inertial Measurement Unit (IMU) and a wheel odometry measurement device, all mounted on the acquisition vehicle. During data collection, the position of the vehicle is recorded by the geo-positioning system. Afterwards, the vehicle's trajectory is optimized and filtered using post processing software. Interpolation techniques and coordinates transformations are then applied to compute the global location of each of the acquired 3D points, which results in a georeference a 3D point cloud. This georeferencing method requires both the sensor data and position data to be accurately timestamped.

The position of a georeferenced 3D model is usually expressed using the Global Positioning System (GPS) where a geographic location is referenced by its longitude, latitude and altitude. Models can also be positioned using the Universal Transverse Mercator (UTM) coordinate systems where a location is referenced by its UTM zone and its northing and easting coordinate pair.