Chapter 5 – ASSESSMENT OF 3D MODELS

The assessment of 3D urban models is currently accomplished largely either by application of classical metrics or by vision-based judgments made by individuals or groups of human beings. These metrics are useful for identifying deficiencies of acquisition, to compare different models and to check conformance with specifications [82]. These specifications are currently often geometric issues such as positional accuracy and completeness but will eventually include suitability for fulfilment of various human goals. Automatic evaluation of the quality together with the knowledge about the input data and relevant meta-information, for example, the purpose of the acquisition, are increasingly needed in decision processes about terrain and terrain models.

Classical mathematical metrics, especially the so-called Sobolev norms, which include the widely used quadratic or rms metric (the "root-mean-square" metric, also called the L_2 metric or norm), have often been used as measures of geometric approximation [83],[84],[85]. Weighted quadratic metrics have had considerable success in terrain modelling [86]. Recently, computational and analytical evidence for use of the Hausdorff metric [87], which is related to the L_∞ metric, has arisen. Overlap of volume bounded by terrain surfaces has been used with considerable success [88],[82] for urban terrain. In this chapter, we first review methods that are currently in use and then describe potential developments in the future. The classical metrics, widely used and useful as they are, generally do not provide good measures of characteristics that correspond to the goals of the human user/observer. In Section 5.1, we provide information about the procedures in current use. In Section 5.2, we provide information about human-goal-based metrics that could be created in the future.

5.1 CURRENT METRICS FOR QUANTITATIVE ASSESSMENT

5.1.1 L_p Metrics

The L_p metrics (norms) used to measure how well a function f (for example, a 2.5D surface in a model of urban terrain) approximates a function g ("ground truth") have been adopted from classical engineering. These metrics have the advantages of being based on well-understood theory and being applicable to a vast number of situations in classical physical science and engineering. In the L_p metric (see [89]), from which much of the material in this sub-section is taken), the difference d_p between two functions f and g on a domain D is:

$$d_p = \left\{ \int_D |f-g|^p \, \mathbf{d}D \right\}^{1/p} \tag{5.1}$$

when $1 \le p < \infty$ and:

$$d_\infty = \sup_D |f-g| \tag{5.2}$$

when $p = \infty$. The widely used rms metric (quadratic metric) is a discrete version of the L_2 metric.

Consider a flat surface g such as depicted in Figure 5-1. Let f be the same flat surface in most places but have a long thin ridge (for example, a fence or thin wall), as depicted in Figure 5-2. For thin ridges, the L_p metric, $1 \le p < \infty$, of the difference between f and g will be small. Nevertheless, in human perception, f is generally not considered a good approximation of g. One important aspect of human perception is visibility. The visibility between observer-target pairs on a surface f with a long thin ridge is very much different from the unobstructed visibility on the flat surface g. For $1 \le p < \infty$, L_p metrics for the difference between f and

g can thus be small when the human-perceived difference between f and g is large. When g is a flat surface and f is the same surface with additional small-amplitude oscillation, the L_∞ metric of the difference between f and g is small when the human-perceived difference is large. The L_p metric, $1 \leq p \leq \infty$, of the difference can also be large when the human-perceived difference is small, for example, when g is a Heaviside function (one value on one side of a line or curve and some other value on the other side) and f is the same function but with a slight shift in the location of the discontinuity. Heaviside functions represent sides of buildings and cliffs and are common in urban and natural terrain. The L_p metrics and all commonly used classical metrics give equal weight to equal amounts of undershoot and overshoot. However, this equal weighting does not match human goals well because "too high" (overshoot) and "too low" (undershoot) are not opposites of each other when visibility, drainage, communication (radio, optical, etc.) and many other human goals are under consideration.

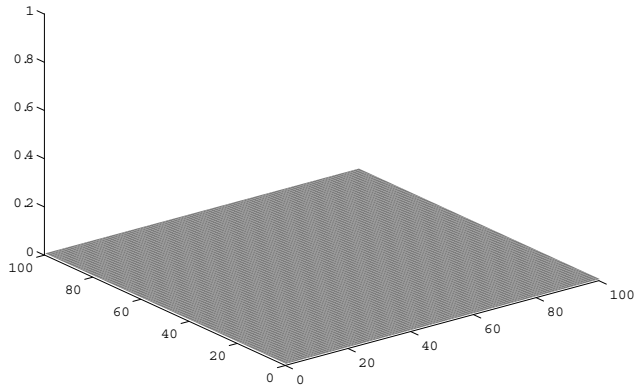

Figure 5-1: Flat Surface.

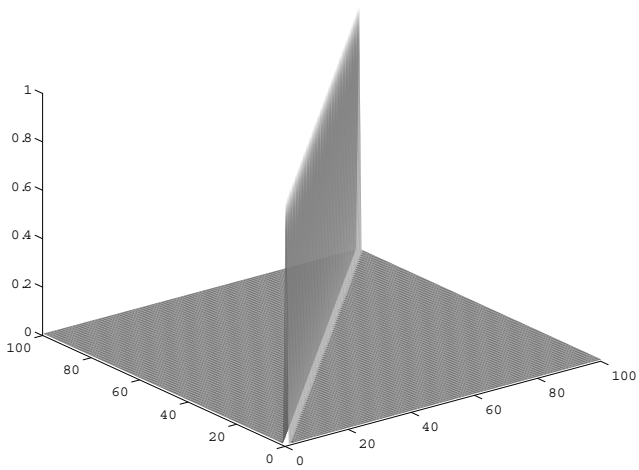

Figure 5-2: Flat Surface with Long Thin Ridge.

5.1.2 Hausdorff Metric

Metrics other than L_p metrics have been used to measure goodness-of-fit in geometric modelling. Petukhov [90] carried out analysis in the Hausdorff metric (the maximum of the minimum distances between the surfaces), but the subject area goes back further to Sendov and the Bulgarian school of approximation [96] To make this precise the maximum distortion $d_H(A,B)$ of two sets A and B is computed

as the larger of the distances of the further-most point of A from the set B and the furthermost point of B from the set A:

$$d_H(A,B) = \max(\max_{x \in A} dist(x,B), \max_{y \in B} dist(y,A))$$

Two surfaces S_1 and S_2 satisfying $d_H(S_1,S_2) < \varepsilon$ means that each point of S_1 lies within an ε - bubble around S_2 and similarly each point of S_2 lies within an ε - bubble around S_1. The relationship of the Hausdorff metric in 3D to the rms metric L_2 and the uniform metric L_∞ is illustrated for 2D cross-sections in Figure 5-3.

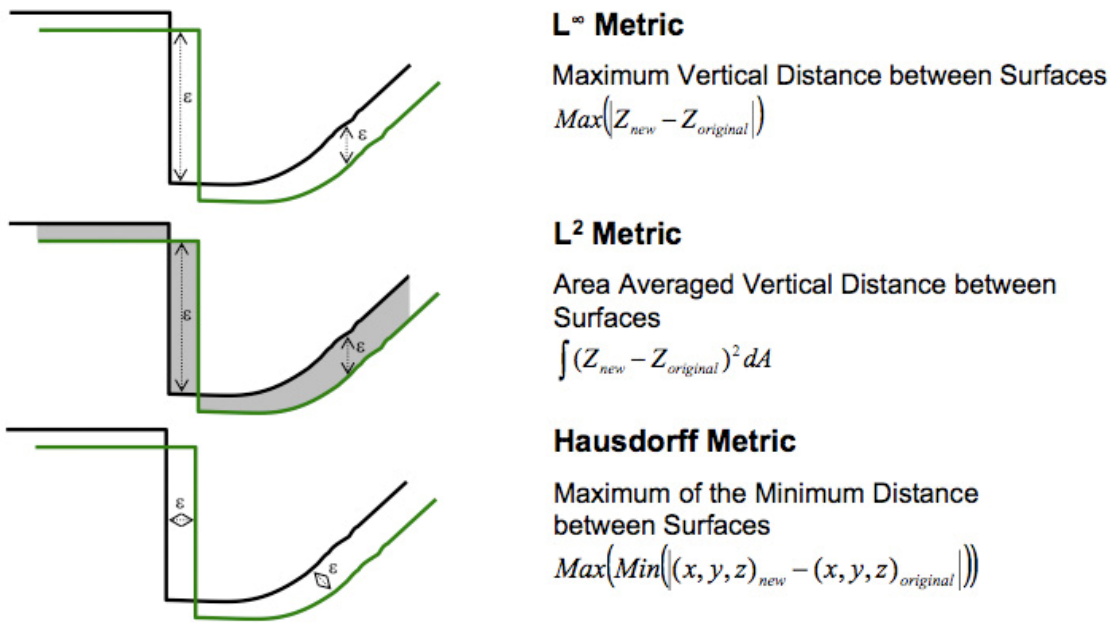

Figure 5-3: ε Distance Measurements between Black and Green Surfaces in the L_∞, L_2, and Hausdorff Metrics, Respectively. For clarity, only cross-sections of the two surfaces are illustrated.

Observe from this figure that, in the uniform metric (L_∞), the differences of the surfaces are measured only in the vertical direction. In this case, if a footprint of a building is perturbed, no matter how small, the error becomes the full height of the building. The rms metric computes an averaged value of these same vertical errors between the two surfaces, i.e., the square of the vertical errors is integrated over the region of interest. On the other hand, the Hausdorff metric estimates the worst distortion between the curves, but not biased to any direction. There are many variations of Hausdorff metrics, including averaged L_p versions of the closest distances between the surfaces for which an approximation theory has been developed [92]. A generalization, the Hausdorff-Gromov metric, allows for incorporation of isometric distortions [93] to be applied to either surface before computing the Hausdorff error which may be useful for SFM estimation. A related, but different distortion metric has been developed by Guibas and his collaborators [94]. This metric was named Earth Mover's Distance since it may be interpreted as the work required transporting mass from one place to another.

To illustrate with terrain mapping the effects of different metrics, we refer to a report by Thies et al. [95] where these three metrics are used as the distortion metrics in best approximating an urban terrain with a fixed budget and computing the corresponding regions of line of sight or "viewsheds". The computational experiment used a section of a high quality DTED terrain map of Baltimore as ground truth and computed the best approximate surfaces of the terrain by selection of the optimal 5,000 vertex heights to reduce the

respective metrics. Viewsheds for these three test surfaces, each produced by minimizing the corresponding metric, were computed at 20 uniformly spaced positions proceeding north along a main road. This urban terrain and route are illustrated in the cartoon of Figure 5-3. The line of sight region was then determined for each of the 20 locations along the road at a common height of 3 meters above the road surface with the results recorded in Figure 5-4. These results simulate a convoy proceeding along the road.

Figure 5-4: Baltimore Height Data Used to Estimate Distortions of Viewshed Using Different Metrics. The orange line depicts the path of a convoy traveling north on a main road.

In Figure 5-5, the resulting viewsheds are rendered for the rms (L_2), uniform (L_∞) and Hausdorff metrics. In these images, the areas coloured yellow are false positives, that is, points seen on the metric-approximated surface but are not visible from the convoy's route on the true surface. The areas coloured red are false negatives, that is, points that are not seen on the approximate surface but that are seen on the true surface. Grey denotes the region that can be seen on both the approximate and the true surface; black signifies the terrain that cannot be seen on either surface.

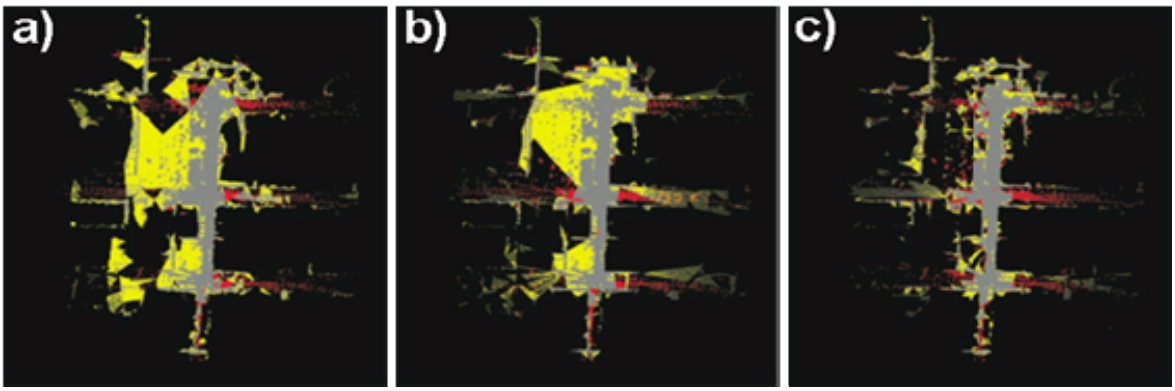

Figure 5-5: Correct and Erroneous Portions of Viewsheds for Convoy Route in Baltimore – a) L_2 Metric; b) L_∞ Metric; and c) Hausdorff Metric. See text for explanation.

5.1.3 Quality Measures Based on Overlap of Volume

Metrics such as the root mean square (rms) metric cause problems when applied to 3D terrain with complex structures [82] due not merely to lack of rationale for application of the rms metric in this situation but also to a point matching problem (which points on the models to be compared should indeed be compared with each other). In the literature, there exist a number of alternative quality measures. These have been derived for different tasks and can partially be converted in each other. Selection criteria for the quality measures are [88]:

1) The values of the quality measures should be *reliably computable* with *moderate technical effort*. Special cases that can occur with empty model sets should be avoided.

2) *Evaluations in 2D and 3D* should be possible without any change of methodology.

3) The quality measures should have a *limited range*. Measures with a range between zero and one can be interpreted immediately as percentages. They allow comparison of the results for different model assessments.

4) The quality measures should *be easily interpretable*. Concerning conformance with a specification, it should be possible to give a statement such as “10% of the urban terrain has not been acquired”.

5) The values of the quality measures should be *independent of the volume*. The volume to be rastered is defined by a bounding box for each data set. Since this cuboid is often aligned with the coordinate axes, the volume depends on the coordinate system. A complement to the data set or a rotation of the data set can change the volume.

6) The quality measures should be locally and global identical and their values locally and global computable. This allows for instance the comparison of a local measure value for a single building with the corresponding value for the entire model, which can be seen as an average value. It should be possible to carry out the interpretation of the results hierarchically. The global 2D and 3D measures can first be taken into consideration. If their values are insufficient for the fulfilling the need, local measure values can be used.

7) The quality measures should be invariant with respect to conjoint translations and rotations of the test and reference model.

The quality measures used in the following are based on a test model set T and a reference model set R. Quality measures in the literature that satisfy the criteria specified above include:

The *quality rate*

$$\rho_q = \frac{|T \cap R|}{|T \cup R|} \qquad \rho_q \in [0, 1] \tag{5.3}$$

The value of the quality rate is independent of the assignment of the test and reference data set (symmetry). Its optimum value is 1.

The *detection rate*

$$\beta = 1 - \frac{|R \setminus T|}{|R|} \qquad \beta \in [0, 1] \tag{5.4}$$

measures the proportion of buildings or building parts correctly detected. Its optimum value is 1.

The *branch factor*

$$\rho_b = \frac{|T \setminus R|}{|T \cap R|} \tag{5.5}$$

measures the proportion of objects falsely detected. Its optimum is 0.

The *miss factor*

$$\rho_m = \frac{|R \setminus T|}{|T \cap R|} \tag{5.6}$$

measures the proportion of objects not acquired. Its optimum is 0.

Measures (5.5) and (5.6) do not fulfil Criterion 3, but are nevertheless easily interpreted. These measures can be computed for whole scenes in 2D or 3D and can also be computed for individual buildings. The quality rate (5.3) can be computed with weights, for example, those resulting from a distance transformation.

To illustrate the applicability of these measures, Figure 5-6 shows two building models obtained by two different acquisition methods, cf. [88]. The assignment of test and reference is irrelevant for this example, since no information about the accuracies is a priori known. For the comparison of the test and reference data sets intersection and union sets have to be determined. In principle these calculations can be done vectorially. However, in the case of volume determinations, these calculations and the corresponding data structures are rather complex. A voxel-based approach, that is, a spatial enumeration in cells or voxels, was chosen to circumvent these problems and to suppress the effects of different topologies. The precision of the quality measures depends on the spatial resolution defined by the sizes of the volume cells. This approach has the advantage of allowing the evaluation in 2D (position) or in 3D (position and height) by using the same quantitative quality measures and methods.

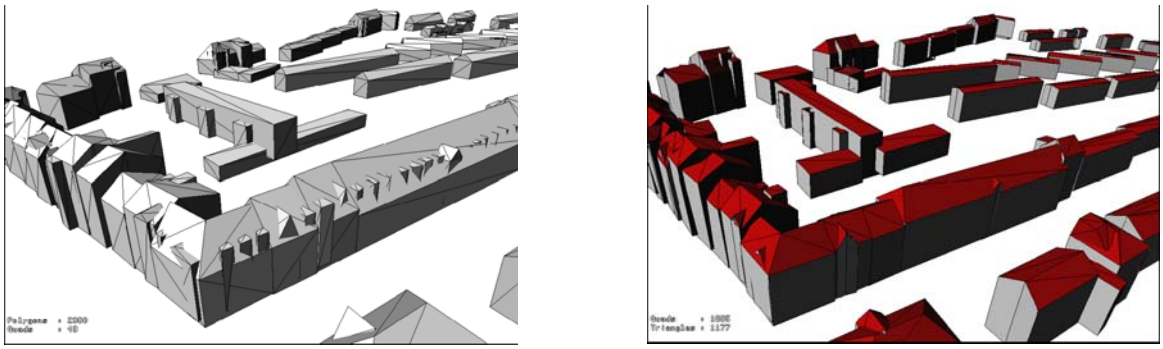

Figure 5-6: Building Models Obtained by Two Different Acquisition Methods.

Both data sets were rastered with a cell size of 0.5 m. After this screening, the differences, unions and intersections of the volumes were calculated and the connected components determined. Figure 5-7 shows an example of a screening and a difference set.

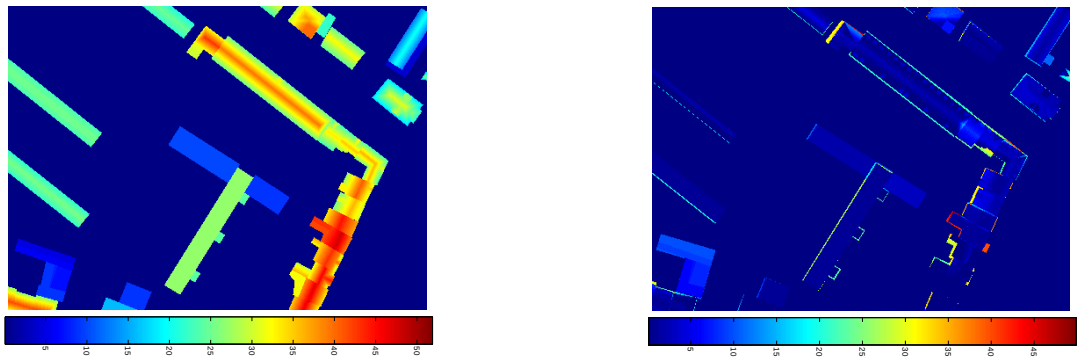

Figure 5-7: (Left) Detail of the Screened Test Data Set T and (Right) Difference Set T\R.

The quality measures listed above in this section, computed for this example for both 2D ("footprint") and 3D, are listed in Table 5-1.

Table 5-1: Quality Measures (Globally Calculated) for the Two Models of Figure 5-6.

Quantity	2D	3D	Range	Optimal Value
Quality Rate	0.87	0.75	[0,1]	1
Miss Factor	0.10	0.24	≥0	0
Branch Factor	0.05	0.10	≥ 0	0
Detection Rate	0.96	0.91	[0,1]	1

The evaluation can be carried out on a per-unit basis. Figure 5-8 shows the results for the quality rate computed for buildings or building parts.

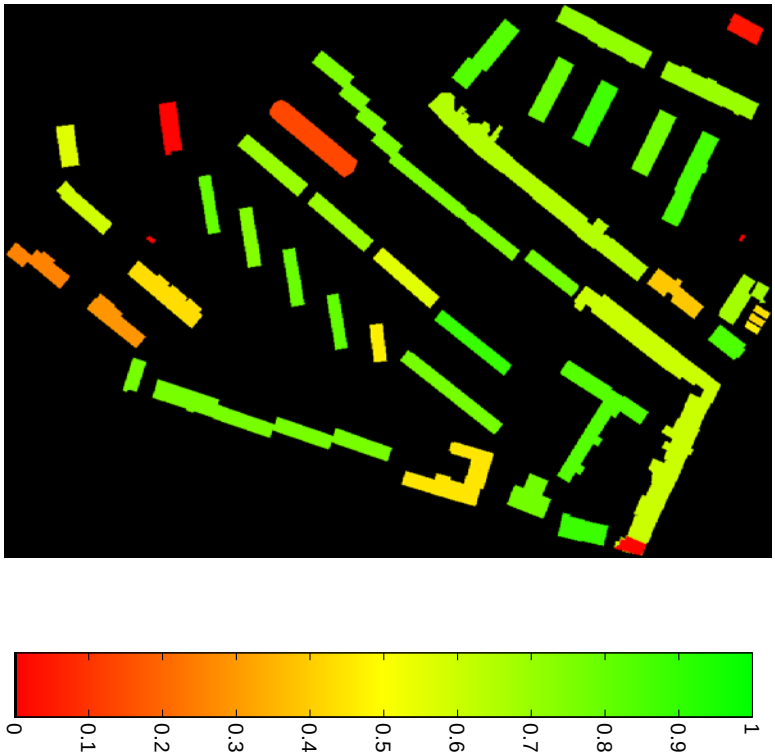

Figure 5-8: Diagnostics by Visualisation of the Local Quality Rates, Computed in 3D.

5.2 FUTURE HUMAN-GOAL-BASED METRICS FOR QUANTITATIVE ASSESSMENT

The metrics described above provide information about accuracy and other properties of models. However, these metrics are only incompletely linked with human goals. This is particularly true of the widely used L_p metrics.

If a metric for the difference between two functions is intended to be part of a human decision-making process, then this metric should be directly linked to the human goals in this process. Unfortunately, metrics that fully express what is important in human goals have not yet been developed. There are many different human goals, including accuracy of visibility, that could be bases for measuring the difference between two surfaces. In [89], from which much of the material in this section is taken, options for defining metrics that express accuracy of visibility and more general human goals are described.

The extent to which the visibility regions (line-of-sight regions) of two surfaces coincide is an issue related to a human goal that is important in many situations. We define here a visibility function and two metrics based on visibility. Let there be given a height function $\varphi(x,y)$ for (x,y) in some 2D domain D. Let there also be given two 3D domains, O and T, at the points of which "observers" and "targets", respectively, are located. For example, when observers and targets are humans, unmanned ground vehicles and unmanned aerial vehicles that are always on or above φ and always at or below a certain height H, the domains O and T would both be:

$$\{ (x, y, z) \mid (x, y) \in D, \varphi(x, y) \leq z \leq H \} \tag{5.7}$$

A target at a point $(\hat{x},\hat{y},\hat{z})$ in T is visible to an observer at a point (x,y,z) in O if the (open) line segment from (x,y,z) to $(\hat{x},\hat{y},\hat{z})$ is always above the surface φ, that is:

$$z + t(\hat{z} - z) > \varphi(x + t(\hat{x} - x), y + t(\hat{y} - y)), \quad 0 < t < 1 \tag{5.8}$$

In this case, we say that there is "line of sight" from (x,y,z) to $(\hat{x},\hat{y},\hat{z})$.

Define a point-to-point visibility function:

$$\begin{aligned} \mathrm{vis}_{\varphi}(x,y,z;\hat{x},\hat{y},\hat{z}) &= 1 \text{ if point } (\hat{x},\hat{y},\hat{z}) \text{ is visible from point } (x,y,z) \\ &= 0 \text{ if point } (\hat{x},\hat{y},\hat{z}) \text{ is not visible from point } (x,y,z) \end{aligned} \tag{5.9}$$

Based on this point-to-point visibility function, we can define a visibility metric V for the surface φ and for an observer at (x,y,z) in O to be the integral of vis_{φ} over T, that is:

$$\mathrm{V}_{\varphi;T}(x,y,z) = \iiint_T \mathrm{vis}_{\varphi}(x,y,z;\hat{x},\hat{y},\hat{z})\, \mathrm{d}\hat{x}\, \mathrm{d}\hat{y}\, \mathrm{d}\hat{z} \tag{5.10}$$

Let there be given two height functions, a model $f(x,y)$ and "ground truth" ("real terrain") $g(x,y)$. For an observer at (x,y,z), we define the difference-of-visibility or "DV" metric between functions f and g to be:

$$\mathbf{DV}_{f,g;T}(x,y,z) = \iiint_T \left| \begin{matrix} \mathbf{vis}_f(x,y,z;\hat{x},\hat{y},\hat{z}) \\ -\mathbf{vis}_g(x,y,z;\hat{x},\hat{y},\hat{z}) \end{matrix} \right| \mathbf{d}\hat{x}\, \mathbf{d}\hat{y}\, \mathbf{d}\hat{z} \tag{5.11}$$

If the DV metric is small (large), visibility on f does (does not) closely approximate visibility on g. In practice, the ground truth g that appears in vis_g in the DV metric is rarely known completely. Only data points on and perhaps some other information (derivatives at some points, monotonicity, drainage patterns, etc.) about g are known. Hence, an exact DV metric will typically have to be replaced by an approximate DV metric.

There are many extensions. The point-to-point visibility function vis_{φ} assumes that all visible targets are equally visible, no matter how distant they are from the observer. With telescopic lenses, this assumption is reasonable in many circumstances. However, when telescopic lenses are not available (for example, in low-cost sensors) and when visibility decreases with distance (for example, due to haze or smog in the atmosphere), it is appropriate to introduce in the DV metric a weighting function w that decreases as the distance of the target from the observer increases, perhaps at different rates in different regions and directions. In this case, the point-to-point visibility function $\mathrm{vis}_{\varphi}(x,y,z;\hat{x},\hat{y},\hat{z})$ for a surface φ can be defined to be:

$$\begin{aligned} \mathrm{vis}_{\varphi}(x,y,z;\hat{x},\hat{y},\hat{z}) &= w(x,y,z;\hat{x},\hat{y},\hat{z}) \text{ if point } (\hat{x},\hat{y},\hat{z}) \text{ is visible from point } (x,y,z) \\ &= 0 \text{ if point } (\hat{x},\hat{y},\hat{z}) \text{ is not visible from point } (x,y,z) \end{aligned} \tag{5.12}$$

When a telescopic lens is not available but the atmosphere is clear, one could choose:

$$w(x,y,z;\hat{x},\hat{y},\hat{z}) = \frac{1}{(x-\hat{x})^2 + (y-\hat{y})^2 + (z-\hat{z})^2} \tag{5.13}$$

to express the fact that the visually recorded area of the target as seen by the observer is inversely proportional to the square of the distance from the observer to the target.

When the observer position is variable, it may be best to use "global" metrics. One can define a global visibility metric GV for a function φ to be the integral of V over O, that is:

$$\mathrm{GV}_{\varphi;O,T} = \iiint_O \left[\iiint_T \mathrm{vis}_\varphi(x,y,z;\hat{x},\hat{y},\hat{z})\, d\hat{x}\, d\hat{y}\, d\hat{z} \right] \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{5.14}$$

One can define a global difference-of-visibility metric GDV between functions f and g to be the integral of DV over O, that is:

$$\mathrm{GDV}_{f,g;O,T} = \iiint_O \left[\iiint_T \left| \begin{matrix} \mathrm{vis}_f(x,y,z;\hat{x},\hat{y},\hat{z}) \\ -\mathrm{vis}_g(x,y,z;\hat{x},\hat{y},\hat{z}) \end{matrix} \right| \mathrm{d}\hat{x}\, \mathrm{d}\hat{y}\, \mathrm{d}\hat{z} \right] \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{5.15}$$

As defined in (5.11) and (5.15), the DV and GDV metrics weight all observer and target points equally. In many cases, users may wish to weight some regions in O and T differently from other regions to reflect the importance of accurate modelling in those regions. This can be done by introducing weight functions in the integrands of the DV and GDV metrics. Analogous adjustments can, of course, be made in the integrands of the visibility metric V of (5.10) and the global visibility metric GV of (5.14).

The difference-of-visibility metrics DV and GDV treat false negatives (f predicts no visibility when g has visibility) and false positives (f predicts visibility when g does not have visibility) equally and provide sums of these two types of error. However, there are many circumstances in which one wishes to measure only one of these types of error, not both together. For example, for concealing unsightly civilian assets such as garbage dumps and for concealing many military assets, false negatives are serious while false positives are less so. On the other hand, for emplacing optical communication nodes under friendly conditions, false negatives are generally not serious. They merely increase costs marginally by reducing the options in the planning stage. However, false positives are serious because they result in constructing nodes in locations where the system cannot function and thereby increase costs dramatically. Difference-of-visibility functionals that measure only false negative error or only false positive error are easily constructed. For example, a difference-of-visibility metric for false negative error is:

$$\iiint_T \max\left\{ \begin{matrix} \mathbf{vis}_g(x,y,z;x,y,z) \\ -\mathbf{vis}_f(x,y,z;x,y,z) \end{matrix}, 0 \right\} \mathbf{d}\hat{x}\, \mathbf{d}\hat{y}\, \mathbf{d}\hat{z} \tag{5.16}$$

and a difference-of-visibility metric for false positive error is:

$$\iiint_T \max\left\{ \begin{matrix} \mathbf{vis}_f(x,y,z;x,y,z) \\ -\mathbf{vis}_g(x,y,z;x,y,z) \end{matrix}, 0 \right\} \mathbf{d}\hat{x}\, \mathbf{d}\hat{y}\, \mathbf{d}\hat{z} \tag{5.17}$$

The sum of these two errors is the total difference-of-visibility error DV.

We have worked with functions $f(x, y)$ and $g(x, y)$ that represent univalent height fields or "terrain skins". The extension of the techniques described here to fully 3D terrain (with, for example, space underneath bridges, interiors of buildings, subway tunnels, etc.) is computationally intensive but conceptually straightforward.

The emphasis here has been on visibility as a metric because:

1) Measuring visibility is a common and important military and human goal; and
2) Quantitative metrics for visibility can be defined in ways, such as those described above, that human beings recognize are directly connected with the human goal.

However, the approach here is not limited to visibility and extends to many other situations. Discovering the metrics by which human beings judge similarities and differences in all areas of human interest, the metrics by which they extract features and the metrics by which they understand perceptual cues will require long-term interdisciplinary research by mathematicians, statisticians, cognitive scientists and domain experts. While classical mathematical metrics may on occasion be found to be appropriate, it is likely that most of the metrics related to human goals will be quite different from the metrics that have been commonly used in the past.

The DV and GDV metrics are computationally intensive. Whether one can find computational procedures for calculating these metrics that have acceptably low computational cost is an open question. The DV and GDV metrics for determining the difference in visibility between exact terrain and an approximate model will generally have to be computed in some approximate manner. In cases where the computational cost of the metric is too large, the question of approximate ersatz metrics will arise. This question in turn leads to a question of metametrics, that is, "metrics of metrics" that measure how well one metric approximates another metric. In the past, smoothness has been virtually the only pattern used in approximation theory. However, there are equally strong patterns – ones not related to smoothness – in many classes of non-smooth functions. Human beings instinctively recognize cities and areas of cities by these patterns. These patterns are potential guideposts on the way to developing a fully non-linear approximation theory with non-smooth metrics of non-smooth functions on non-smooth manifolds that is applicable to terrain and many other modern areas of interest.

Besides being ways of determining how accurate the visibility of a given model f is, the DV and GDV metrics and other human-goal-based metrics are potential bases for generation of optimal models f, that is, of functions f that minimize the metrics over an appropriate function space or manifold. This minimization will be challenging, since the metrics need not be smooth. For example, the DV and GDV metrics are only C^0 continuous with respect to f even when f varies smoothly. Utilizing the structure of terrain in whatever algorithm is chosen will likely be essential for computational efficiency.