ON THE INTERPOLATION PROBLEM OF AUTOMATED SURFACE RECONSTRUCTION

Raid Al-Tahir **Toni Schenk**

Department of Geodetic Science and Surveying
The Ohio State University, Columbus, Ohio 43210-1247
USA

Commission III

ABSTRACT

Automatic surface reconstruction entails two major problems: determining conjugate points or features (matching) and densifying the matched points in object space (interpolation). The two tasks are usually performed sequentially in a hierarchical approach, without interacting with one another. In order to improve the success rate and the reliability of automated surface reconstruction, particularly in large-scale urban areas, the matching on subsequent levels must take into account the results from densifying and analyzing the surface. In this paper we focus on a surface interpolator that produces as realistic surface representation as possible. The interpolation and surface analysis may give clues about surface discontinuities and occlusions – a vital feedback for the matching process on the next level in the hierarchical approach.

KEY WORDS: Machine Vision, Image Analysis, Surface Reconstruction.

1. INTRODUCTION

The main objective of digital photogrammetry is to collect enough information to model the portion of the real world that has been photographed. Two kinds of information are of major interest to accomplish that goal; surface topography, represented by Digital Elevation Model (DEM), and objects on the surface (natural or man-made) which are characterized as discontinuities in the surface. Besides being an essential intermediate step for object recognition, reconstruction of a portion of the earth's surface is the end product for digital photogrammetry.

Automatic surface reconstruction entails two major problems: determining conjugate points or features in the images (matching), and densifying the matched points in object space (interpolation). The two tasks are usually performed sequentially in a hierarchical approach, without interacting with one another. In order to improve the success rate and the reliability of automated surface reconstruction, particularly in large-scale urban areas, the matching on subsequent levels must take into account the results from densifying and analyzing the surface.

This paper is a part of ongoing research focusing on the process of surface interpolation and analysis. The purpose of this paper is to define the tasks for such a process. The paper reviews previous works that have been done in the related fields. The emphasis is on the applicability of suitable for an automated surface interpolation.

2. OSU SURFACE RECONSTRUCTION SYSTEM

Due to the large amount and variety of information in the aerial images, the success of any image processing operation can not be guaranteed. This is especially the case of large-scale urban scenes because occlusion is more frequent, and the visible surface is less smooth. The only alternative to constrain the processes is to adopt a scale-space approach that proceeds hierarchically from the lowest resolution for a

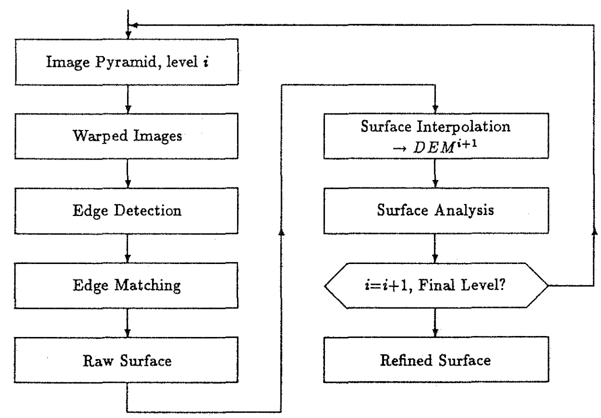

Figure 1: Outline of OSU surface reconstruction system.

stereo pair to the finest. OSU surface reconstruction (Schenk & Toth, 1992) is such hierarchical approach. It consists of several modules that are executed in an iterative fashion(Figure 1). Each level of the process aims at refining the geometry of the images and improving the surface representation.

In the OSU surface reconstruction system, the process starts by having two conjugate images sampled at the lowest level of resolution. The orientation of these images is obtained through edge detection and matching. The results of this step are the orientation parameters, as well as a set of highly reliable matched points. The raw surface is then constructed by computing the 3-D object space coordinate for the set of points. These points are sparsely and irregularly distributed. Thus, a dense surface representation (DEM) must be interpolated for. A DEM, tesselated at the next higher level of resolution, is essential for surface analysis, and for the subsequent cycles. The final step is surface analysis for hypothesis generation and verification concerning potential break lines and surface segmentation.

A new cycle starts with sampling the original stereo pair at the subsequent level of resolution, and warping the left and right images with respect to the interpolated surface. The whole process is repeated until the final refined surface is reached. At each level, images are rectified, the matching accuracy and reliability are improved, and a better surface representation is obtained. At the last level, the matching vectors vanish, the warped images become orthophotos, and the true surface is reconstructed.

From this overview, it is clear that one of the objectives of surface interpolation is to construct as a realistic surface representation as possible. This task is crucial for the success of matching on subsequent levels. The search for a match is performed by centering a correlation window over a point of a zero-crossing contour in one image. On the other image, the search window is placed and shaped according to the expected depth range in that area (Schenk & Toth, 1991).

The other goal of the surface interpolation is to provide information for the surface analysis. It is important that the interpolator does not introduce new characteristics to the surface other than what is derived from the observations. Creating new maxima or minima in the surface is an example for undesired side effects of interpolation. Additionally, the surface interpolator should not smear essential surface shape characteristics. Such a situation may occur when a smooth surface is interpolated over observations on break lines.

3. SURFACE INTERPOLATION

The problem of surface fitting consists of taking a region containing a list of function values, and finding a function on this region that agrees with the data to some extent and behaves reasonably between data points (Lancaster & Salkauskas, 1986). The accuracy that can be obtained from a fitting process depends on the density and the distribution of the reference points, and the method. Data points are arranged in various distribution patterns and densities. Accordingly, surface fitting methods designed for one case differ from those designed for dealing with other distribution patterns.

There are several criteria for classifying surface fitting methods. The first criterion is the closeness of fit of the resulting representation to the original data. Thereby, a fitting method can be either an interpolation or an approximation. Interpolation methods fit a surface that passes through all data points. Approximation methods construct a surface that passes near data points and minimizes, at the same time, the difference between the observed and the interpolated values.

Another criterion is the extent of support of the surface fitting method; a method is classified as a global or a local one. In the global approach, the resulting surface representation incorporates all data points to derive the unknown coefficients of the function. By doing so, some of the local details submerge in the overall surface, and editing one point affects all distinct points. With local methods, the value of the constructed surface at a point considers only data at relatively nearby points. Thus, the resulting surface emphasizes the small-scale trends in the data (Watson, 1992). Many global schemes can be made local by partitioning the original domain into subdomains.

Yet another criterion for classifying interpolation methods is their mathematical models. Surface interpolation methods are divided into three main classes; weighted average methods, interpolation by polynomials, and interpolation by splines.

3.1 Weighted average methods

These methods use a direct summation of the data at each interpolation point. The value of the surface at a non-data point is obtained as a weighted average of all data points. The weight is inversely proportional to the distance r_i. Shepard's method may serve as an example. Here, the value of a point is evaluated as

$$f(x,y) = \begin{cases} \sum_{i=1}^{N} F_i/r_i^{\mu} \div \sum_{i=1}^{N} 1/r_i^{\mu}, & \text{when } r_i \neq 0, \\ F_i, & \text{when } r_i = 0. \end{cases} \quad (1)$$

Weighted average methods are suitable for interpolating a surface from arbitrarily distributed data. However, one drawback is the large amount of calculations, especially for many data points. To overcome this problem, the method is modified into a local version. A smaller subset of data is selected for each non-data point based on a fixed number of points, or a fixed area. The problem now is to define proper parameters (e.g. the variable μ in equation (1)).

3.2 Interpolation by polynomials

A polynomial p is a function defined in one dimension for all real numbers x by

$$p(x) = a_0 + a_1x + \cdots + a_{N-1}x^{N-1} + a_Nx^N, \quad (2)$$

where N is a non-negative integer and $a_0, \ldots, a_N$ are fixed real numbers. Generally, fitting a surface by polynomials proceeds in two steps. The first one is the determination of the coefficients of the polynomial based on the set of data points and the criteria controlling the fit of the polynomial function. Then, using the computed parameters, the second phase evaluates the polynomial to obtain values of the fitted surface at given locations.

Piecewise polynomials are the local version for surface fitting with polynomials. This approach works well with irregularly spaced data. The general procedure for surface fitting with piecewise polynomials consists of the following operations:

1. partitioning the surface into patches of triangular or rectangular shape, the vertices of which are the reference points.
2. fitting locally a leveled, tilted, or second-degree plane at each patch, using one or more terms of the polynomial.
3. solving the unknown parameters of the polynomial. To enforce continuity (and smoothness) along the joining sides of neighboring patches, partial derivatives must have been estimated at each reference point.

Least squares fitting by polynomials performs well if many points are available and the surface has fairly simple form (Hayes, 1987). On the other hand, interpolation by polynomials with scattered data causes serious difficulties, one of which is a singular system of equations due to data distribution (e.g. data lie on a line). Another problem is an ill-conditioned normal equation system as is the case of consecutive intervals that contain no data. Yet another problem in using polynomials is their tendency to oscillate, resulting in a considerably undulating surface.

3.3 Interpolation by spline functions

A spline is a piecewise polynomial function defined on contiguous segments. In defining a spline function, the continuity and smoothness between two segments are constrained at the interior knots by demanding the existence of certain derivatives. For example, a spline of degree n has n-1 derivatives at the knots, denoted by C^{n-1}.

Bicubic splines, which have continuous second derivatives (i.e. C^2), are commonly used for surface fitting. The solution is obtained by a least-squares approach or the tensor product of orthogonal functions. With increasing number of data points, problems with computing efficiency and accuracy may occur. B-splines are also frequently used for surface fitting. They are characterized by their finite support, which is the interval over which the function is non-zero. Limiting the support of a spline changes the normal equation into a band form. Thereafter, the amount of computations is reduced by a factor of (number of knots/4)^2 (Hayes, 1987).

Bicubic splines and B-splines work best in the case of gridded or uniformly-distributed dense data (Hayes, 1987). However, rank-deficiency in the system of equations becomes a serious problem when applying these approaches to scattered data. Because of data distribution, data points may not lie in the support region of splines. Another situation rises when the data are clustered in one region creating a set of linear equations of marginal differences, thereby producing near singularity.

Nodal basis-functions are another sub-group of methods for surface fitting with splines. The general procedure in this approach consists of defining a set of basis functions and the corresponding data points. Each basis function is centered over a data point (node). The interpolation spline function then is a linear combination of the basis functions. The advantage in using such an approach is that knowledge about spline locations (knots) is not required. Another advantage is that values at the nodes of a regular grid are found directly instead of the two step approach mentioned earlier (Briggs, 1974).

Thin plate splines are derived from the nodal basis-functions. These splines are also called "minimum curvature splines" since they are obtained by minimizing the total curvature of cubic spline s

$$\int\int \left(\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2}\right)^2 dx\, dy. \tag{3}$$

The same form can be obtained by solving the small deflection equation of an infinite plate that deforms by bending it only. The displacement u due to a force f_i acting at N points is represented by the differential equation (Briggs, 1974)

$$\begin{aligned}\frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} &= f_i, \text{ at observation position,} \\ &= 0 \text{ otherwise.}\end{aligned} \tag{4}$$

Adopting the physical analogy, depth data is represented by a set of vertical pins scattered within the region; the height of an individual pin is related to the elevation of the point. Fitting a surface is then analogous to constraining a thin (elastic) plate to pass over the tips of the pins (Figure 2).

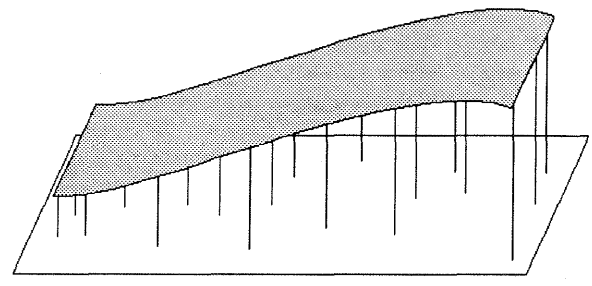

Figure 2: Fitting thin plate over pins.

One method for solving the differential equation is by finite differences or finite elements. Following this approach, the discrete interpolation becomes a repeated passage of a set of simple masks, such as the following mask for elements within a grid:

$$\begin{bmatrix} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & 20 & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{bmatrix} \tag{5}$$

3.4 Surface interpolation by regularization.

A problem is well-posed if a solution exists, is unique, and depends continuously on the initial data. It must also be well conditioned to ensure numerical stability (robust against noise) (Poggio et al., 1985). Shorter than these conditions, the problem is considered ill-posed. Reconstruction of the visible three-dimensional surfaces from two-dimensional images is an ill-posed problem because some information is lost during the imaging process (projecting 3-D into 2-D) (Poggio et al., 1985). Other reasons are the noise and erroneous, inconsistent, and sparse measurements (Terzopoulos, 1985).

Regularization is the frame within which an ill-posed problem is changed into a well-posed one (Poggio et al., 1985). The class of possible solutions is restricted by introducing suitable a priori knowledge, which in the case of surface interpolation is the continuity of the surface. The problem is then reformulated, based on the variational principle, so as to minimize an energy function E constructed from two functionals. The first one measures the smoothness of the solution S, while the second one, D, provides a measure of the closeness of the solution to the observations. The two measures are combined to form the energy function $E = S + D$. Applied to the surface reconstruction problem, the energy function can be written as

$$\int\int [f''(x,y)]^2 dx dy + \lambda \sum [f(x_i, y_i) - d_i]^2. \tag{6}$$

In practice, the function in the integration is either a thin-plate spline $(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2)$, a membrane $(f_{xx}^2 + f_{yy}^2)$, or a combination of both. The variable λ is the regularization parameter which controls the influence of the two functionals. If λ is very large, the first term in the integral heavily affects the solution, turning it into interpolation (close to data). On the other hand, if λ is small, the solution emphasizes the smoothness of the surface.

4. DISCONTINUITY DETECTION

There are only a few methods which try to detect discontinuities in the surface. Grimson and Pavlidis propose detecting discontinuities before interpolating the surface to overcome the problem of oscillations in the fitted surface (Grimson & Pavlidis, 1985). The main idea for this approach is to fit locally a simple surface (plane) to the data and examine the distribution of the residual error. If it appears to be "random", then the hypothesis of no discontinuity is accepted. If there is a systematic trend, then a discontinuity of a certain type is hypothesized. Discontinuities are subdivided into various types, each of which is characterized by a certain combination of change in magnitude and sign of the residual. Once a discontinuity is detected, the surface is broken down into smaller regions, and the surface reconstructor is passed over each of them.

The second approach, proposed by Terzopoulos (Terzopoulos, 1985), is related to the energy function of a thin plate. The thin plate surface over-shoots constraints near the discontinuity causing a sign change of the bending moments at surface inflections. Depth discontinuities are detected and localized by examining the bending moments in the interpolated surface. Changing control parameters within the energy function allows the surface to crease and fracture at the detected discontinuities and reduce the total energy.

Another approach we investigated for detecting discontinuities is based on the concept of a "line process" introduced in (Geman & Geman, 1984. A line process is a set of variables located at the lines which connect the original lattice (pixels or grid cells) (Figure 3). The purpose of a line process is to decouple adjacent pixels and reduce the total energy if the values of these pixels are different. In such a case, the variable of the line process associated with these pixels is set to one, otherwise it is set to zero.

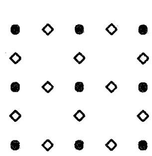

Figure 3: Dual lattice of depth (•) and line (◇) elements.

Eventually, breaking the surface into small pieces around each data point will result in the lowest energy state. To avoid this, a penalty α should be paid (in terms of energy) when a break line is introduced. Thus, a break line will only be introduced when paying the penalty is less expensive than not having the break line at all. The penalty function takes the form $P = \alpha l_i$, where l_i is the line process. This function is added to the original energy function, changing the problem into minimizing

$$E = S + D + P. \tag{7}$$

The result is a combination of a continuous function for the surface and a discrete one for the lines. This combination allows surface reconstruction and discontinuity detection at the same time. However, E is a non-convex function that has many local minima.

One proposal to solve the non-convex function is to adopt a deterministic approach. The line process P is merged with the interpolation function S (Blake & Zisserman, 1987). The modified function is expressed in one dimension as

$$g(u_i - u_{i-1}) = \lambda^2(u_i - u_{i-1})^2(1 - l_i) + \alpha l_i. \tag{8}$$

The resulting function controls the interaction between neighboring grid cells. Such a function prefers continuity in the surface, but allows occasional discontinuities if that makes for a simpler overall description – a theme called "weak continuity constraints".

The modified configuration is then solved by the graduated non-convexity algorithm. The non-convex function E is gradually approximated by a convex one through a family of p intermediate functions. The parameter p represents a sequence of numbers ranging from one to zero. The function $E^{(1)}$ is a crude approximation to the non-convex function. However, as p goes to zero, $E^{(p)}$ becomes closer to the original non-convex one. The neighbour interaction function is also modified into a function of λ, α, and p.

5. EXPERIMENTS AND CONCLUSION

For experimental purposes, we designed synthetic data representing a set of irregular blocks in a small region. Depth information is arranged in a fashion that mimics the pattern of the results of the matching process in the surface reconstruction system. Thus, depth values were provided for some points on, and near by, the edges of the blocks and the edge of the region as shown in figure 4. Figure 5 is a 3-D representation of these points. The location and value of a data point is represented by a peak, while no data points are set to zero.

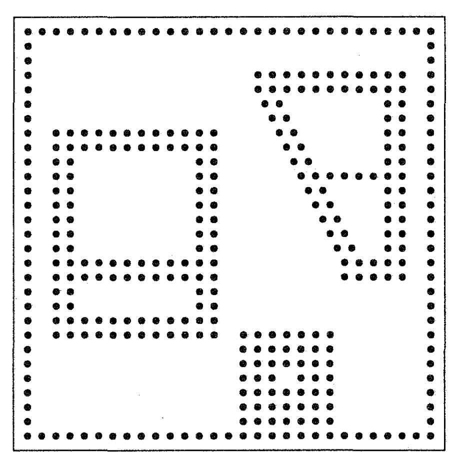

Figure 4: Distribution of synthetic data points.

We evaluated the interpolation methods according to the following criteria:

1. Interpolated surface must be plausible compared to the visible surface in the real world.
2. The interpolation method must not jeopardize clues for surface analysis.
3. The method should be able to utilize a priori information on break lines.

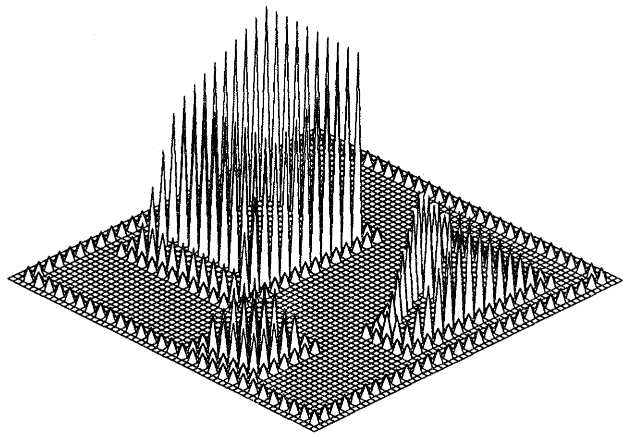

Figure 5: 3-D representation of synthetic data points.

4. The method must be suitable for automation. No human interaction should be necessary to correct parameters.

5. Reasonable demand on computer resources, i.e. time, memory, and storage.

Matching aerial images typically renders a large number of data points, especially at the finer resolutions. Therefore, we have excluded all methods of least square fitting by polynomials or splines because of computational considerations. These methods would lead to a huge system of equations (in the worst case is one equation per point). In addition, having sparse data increases the risk of deficiency in the normal equation. Fitting a surface by piecewise polynomials, furnished with proper triangulation algorithm, stands a better chance for more efficient and realistic surface interpolation. However, the user must identify the set of break lines prior to the interpolation. Otherwise, a peculiar surface representation would be obtained.

The methods of weighted average are better suited for handling sparse data. Besides, they do not introduce new global extrema in the surface. On the other hand, there is no established automatic strategy for defining the data subset for a point. Another concern is the fact that no a priori information about break lines can be included. Therefore, the value of a point is computed based on data across break lines, creating undesired artifacts. Figure 6 shows the result of applying the weighted average method on the test data. The interpolated surface cannot be considered realistic.

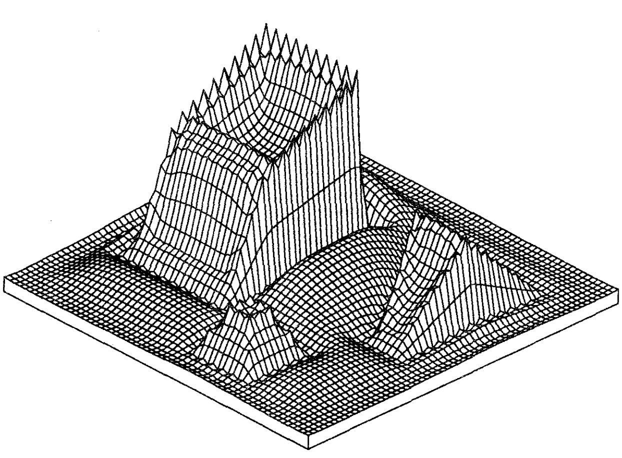

Figure 6: Surface interpolation by weighted average method.

None of these methods provides explicit information for surface analysis. This quite different for fitting a surface by a thin plate (or membrane). Adopting the analogy of a physical model allows exploring the mechanics of such model. Mechanical concepts, such as stress and bending moments of a plate provide the means for detecting break lines. Both models of thin plate and membrane are capable of achieving surface interpolation and break lines detection. Judging from figures 7 and 8, the membrane produces a more realistic surface that the thin plate model.

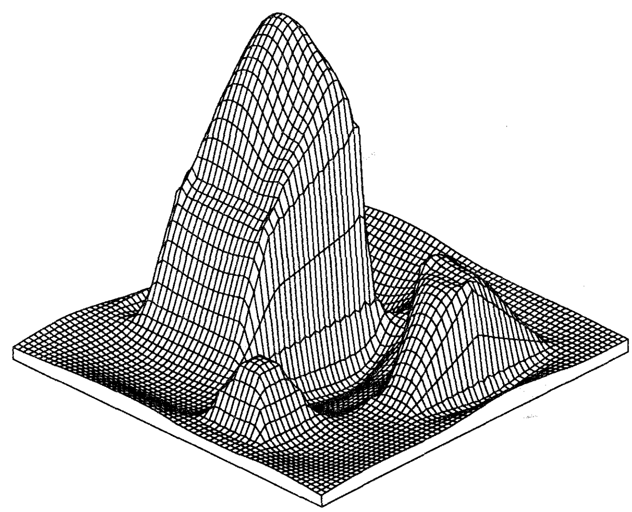

Figure 7: Surface interpolation by thin plate splines.

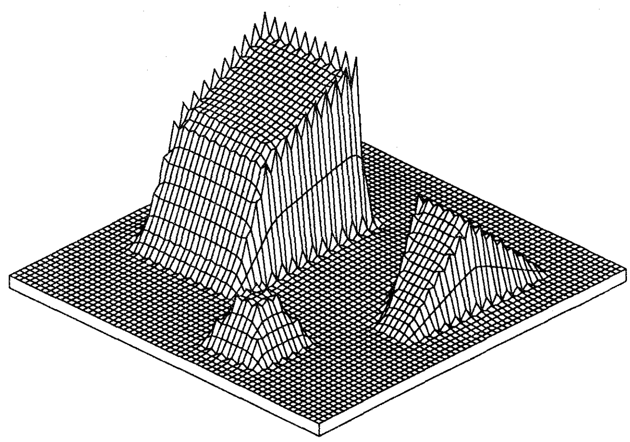

Figure 8: Surface interpolation by a membrane.

Figure 7 represents the interpolated test data by a thin plate. The problem of over-shooting between data points is clearly noticeable. Figure 8 shows the interpolation by a membrane. Here, the problem is interpolating between high frequency features. This is avoided by using the weak continuity constraints. Interpolation by a weak membrane is shown in figure 9. The discontinuities are now detected during the surface interpolation. Figure 10 shows the detected break lines superimposed on the surface.

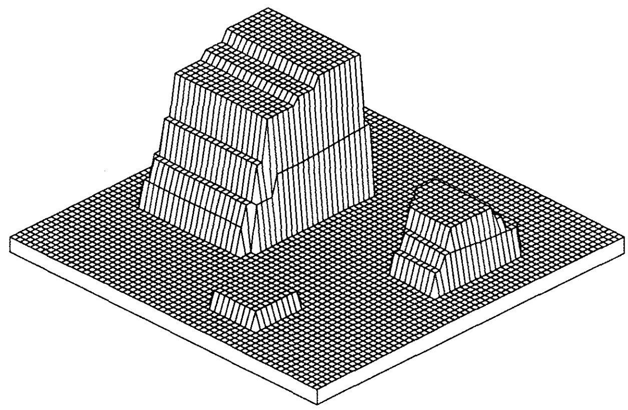

Figure 9: Surface interpolation by a weak membrane.

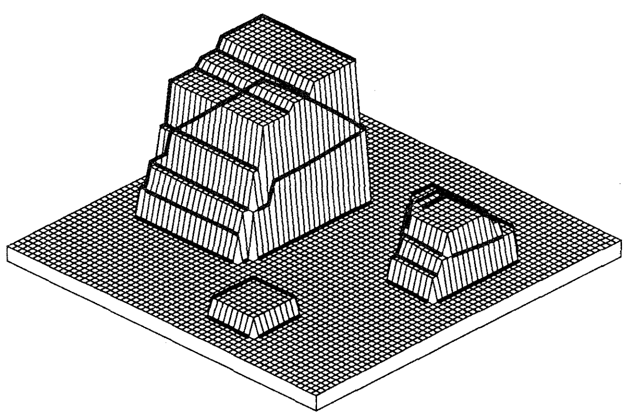

Figure 10: Detected break lines in the surface.

Ongoing research is addressing the following issues:

- Increasing the degree of automation of surface interpolation.
- Subpixel accuracy in the determination of a break line.
- Defining the means to convey discontinuity information to other modules and levels.
- Integration of other cues for discontinuity, such as the residuals between successive levels of surface representation.

References

[1] Blake, A. and Zisserman, A. 1987. Visual Reconstruction. MIT Press. MA.

[2] Briggs, I. 1974. Machine contouring using minimum curvature. Geophysics. 39 (1): 39-48.

[3] Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans on Pattern Analysis and Machine Intelligence. PAMI-6 (6).

[4] Grimson, W. 1981. From images to surface: A computational study of the human early visual system. MIT Press. MA.

[5] Grimson, W. and Pavlidis, T. 1985. Discontinuity detection for visual surface reconstruction. Computer Vision, Graphics, and Image Processing. 30: 316-330.

[6] Hayes, J. 1987. Fitting surfaces to data. The Mathematics of Surfaces II. Martin, R. R. (ed.). Claredon Press. Oxford.

[7] Lancaster, P. and Salkauskas, K. 1986. Curve and surface fitting: An introduction. Academic Press. London.

[8] Poggio, T., Torre, V. and Koch, C. 1985. Computational vision and regularization theory. Nature. 317: 314-319.

[9] Schenk, A. and Toth, C. 1991. Reconstructing visible surfaces. Industrial Vision Metrology. S. El-Hakim (ed.). Proc. SPIE 1526.

[10] Schenk, A. and Toth, C. 1992. Conceptual issues of softcopy photogrammetric workstations. PE&RS. 58 (1).

[11] Schumaker, L. 1976. Fitting surfaces to scattered data. Approximation Theory II. Lorentz et al. (eds.). Academic Press. New York.

[12] Terzopoulos, D. 1985. Computing visible-surface representation. Artificial Intelligence Laboratory Memo. No 800. M.I.T.

[13] Watson, D. 1992. Contouring: A Guide to the Analysis and Display of Spatial Data. Pergamon Press.