Towards an Autonomous System for Orienting Digital Stereopairs

Toni Schenk, Jin-Cheng Li, and *Charles Toth*
Department of Geodetic Science and Surveying, The Ohio State University, Columbus, OH 43210

ABSTRACT: We present a method to automatically compute the orientation of digital stereopairs. The system that we developed is based on feature-based, hierarchical matching combined with precise point determination. We use the LoG operator to determine zero crossings which we then match in the $\psi - s$ domain. No assumptions are required because the $\psi - s$ representation is rotation and scale invariant. In order to obtain maximum possible accuracy for corresponding points, a precise point matching scheme is employed. With the approximations obtained from matching zero crossings, corresponding points with sub-pixel accuracy are determined by using the Förstner interest operator combined with area correlation. The orientation parameters are computed by a rigorous bundle adjustment with blunder detection. The detailed description of our approach is followed by an example. Finally, we summarize results obtained with four different stereopairs.

INTRODUCTION

WITH THE EMERGENCE OF DIGITAL PHOTOGRAMMETRY, expectations are raised that photogrammetric processes may be automated. A fundamental process is the orientation of a stereopair. Not much has been reported about determining the orientation parameters automatically. It is virtually always assumed that this step is performed by a human operator at an earlier stage. In machine vision applications, it is the basic assumption that the cameras are mounted in parallel and are perpendicular to the base, rendering images of epipolar geometry. That is, scan lines are thought to be epipolar lines.

Most of the stereovision algorithms are based on epipolar geometry and are quite sensitive to this condition. Initially, the cameras of a moving robot may be perfectly aligned, but this condition is likely to change in time. We argue that the camera positions should be periodically surveyed and, if necessary, readjusted. Obviously, this can best be performed with an automated orientation. By that, we mean the determination of the camera parameters with respect to an object space.

We do not differentiate between relative and absolute orientation as we are taking a more general view and determine the camera parameters with respect to an object space. This requires eliminating the datum defect by introducing some common information between image and object space. If the object space is identical to the ground control system, the common information is usually the control points. If the object space is an arbitrary coordinate system, usually called the model system, then a variety of possibilities exist to eliminate the datum defect. We leave it to the particular application on how common information is introduced.

The problem of automatic orientation can easily be formulated. Select a sufficient number of points in one image and find the corresponding points in the other image, followed by an adjustment with the orientation elements as parameters. The crucial step is obviously to find the corresponding (conjugate) points. This task is accomplished by image matching methods, such as gray level correlation or feature matching. As long as good approximations (a few pixels) are available and the gray levels yield enough signal within the correlation windows, traditional correlation methods work well.

In this paper we present a general solution that does not require approximations or is not otherwise limited by introducing assumptions and constraints. Previously, we reported about some aspects of our approach (e.g., various matching procedures in Greenfeld and Schenk (1989) or about preliminary results in Greenfeld (1987). The purpose of this paper is to present the current status in a coherent fashion, and to report about results and experiences gained over the last three years.

In the next section we provide the reader with some background information concerning the overall framework, and concepts used. In the third section we explain how initial approximations are found by employing the scale space concept and edge matching. Next, the problem of how to determine points with sub-pixel accuracy is addressed. For this task we use interest points and gray level correlation. Finally, we report about results obtained with several digitized aerial stereopairs, followed by conclusions and future research.

BACKGROUND

The automatic orientation is the first module of an automated mapping system whose conceptual schema is depicted in Figure 1. The major building blocks reflect the computer vision paradigm with three distinct levels. On every level specialized modules process image information and pass it to the next level where it is analyzed and integrated. Often, the result of individual modules may not be robust, but their combination is. That is, the synergism of multiple information arising from different image modules improves the robustness.

Figure 1 should not be confused with the architecture of a real system, say a vision work station, where much more care regarding efficient communication between modules and implementation of hardware and software components is required. The concept serves more as a master plan to coordinate our experiments which are being performed on all levels.

In the context of the conceptual mapping system, the orientation module serves three purposes. Obviously, the main purpose is to establish the relationship between image space and object space. After having determined the orientation parameters, the images are then resampled to epipolar geometry. Another important result is a first approximation of the surface. We determine many (hundreds) of well distributed corresponding points whose object coordinates can be used to interpolate a digital elevation model (DEM). This approximate DEM plays an important role in the surface reconstruction module (see Schenk *et al.*, 1990).

Figure 2 provides an overview of the orientation module. The first part is concerned with determining approximation, followed by computing corresponding points as precisely as pos-

*Currently with Holcomb Research Institute, Butler University, Indianapolis, IN 46208.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 57, No. 8, August 1991, pp. 1057–1064.

0099-1112/91/5708–1057$03.00/0

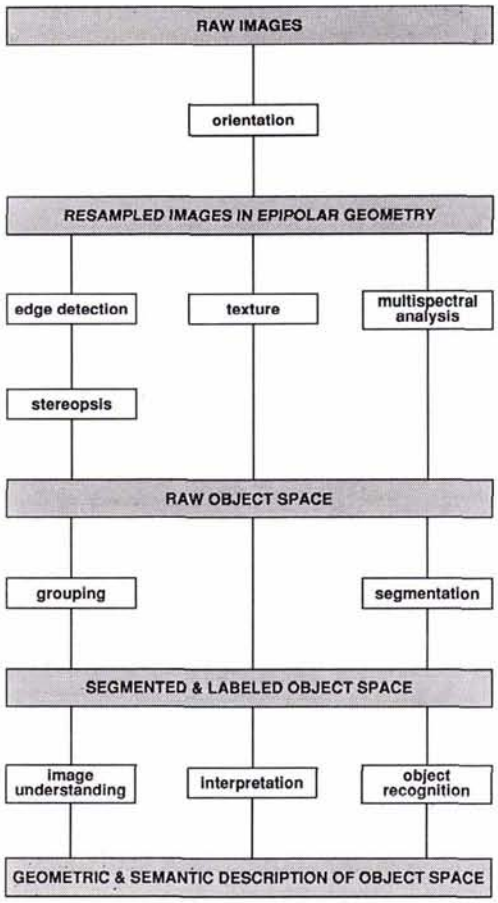

FIG. 1. Major building blocks of conceptual mapping system.

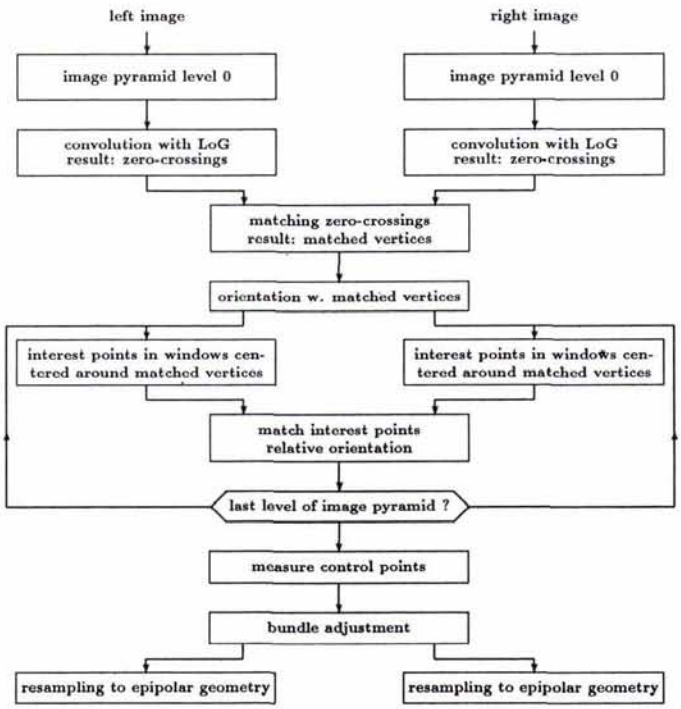

FIG. 2. Overview of orientation module.

sible. Clearly, the first task is more challenging as we only make weak assumptions about the geometrical configuration of the stereopair.

Newer theories about the impressive stereo vision capability of the human visual system suggest that edges are extracted and matched at various scales. Results obtained from small scales constrain the matching process on larger scales. Representing and processing an image at different scales and integrating the results on subsequent finer resolutions (larger scales) has gained much interest in computer vision.[1] Our approach follows this paradigm. The representation of an image at different scales is obtained by smoothing it with filters of varying spatial extent. Smoothing and detecting edges is combined with the Laplacian of Gaussian (LoG) operator.

We begin with building an image pyramid from original images by filtering it with a Gaussian operator. Matching edges obtained on the coarsest level and finding corresponding points are probably the most crucial steps of the entire orientation module. Thus, we explain the procedure at some length in the next section.

The other tasks are concerned with tracking the initial positions through the image pyramid. In windows centered around the matched vertices, interest points are determined and matched to sub-pixel accuracy. Their matched positions serve as new window centers on the next level in the image pyramid. Note that, up to the identification and measuring of control points, all steps are performed automatically. We use Intergraph's Interpro 3055 workstation to compute, to display results, and to measure control points.

COMPUTING INITIAL APPROXIMATION

MULTISCALE TECHNIQUE

Objects represented in the image space vary enormously in size and extent. In order to identify and qualitatively describe events in the object space, it is necessary to evaluate and combine the image at different scales, a procedure known as the multispace technique. Smoothing the original image with a low pass filter of varying size results in images at various scales. The filter size is a parameter of the scale space of the image function. The Gaussian filter has some interesting properties that make it unique for this smoothing process. Furthermore, because object space events are implicit in the image, function primitives are used to help make such events explicit. Originally proposed and described in Marr (1979), the zero-crossings of the Laplacian of Gaussian (LoG) operator combine the Gaussian smoothing filter with the Laplacian intensity change operator into a very popular method of describing the image function in the scale space.

In the discrete case of the multiscale approach, the sizes of the smoothing filter, chosen *a priori*, define the scale levels. The choice is somewhat arbitrary. We exploit the scale-space technique for finding corresponding points on a coarse level and tracing them through the scale space.

We have implemented the multiscale concept by first smoothing the original image with Gaussian filters of various sizes. The corresponding images are also called an *image pyramid*, a term we have adopted throughout this paper. The second step then involves convolving the image pyramid with the LoG operator. Obviously, the same result would be obtained by convolving the original image with a LoG with larger w. It should be noted that the representation of the discrete scale space as an image pyramid is strictly an implementation question.

Most of our digitized aerial photographs have a resolution of 4096 by 4096 pixels. At the top of the pyramid the image is usually represented by 512 by 512 pixels. As indicated in Figures 2 and 3, edges are matched and corresponding points are de-

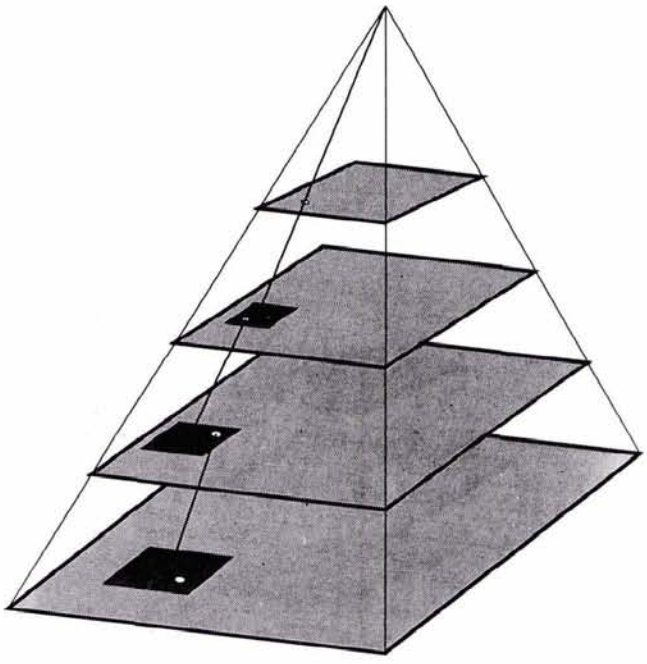

FIG. 3. Image pyramid as a representation of the discrete scale space. The figure also illustrates the concept of tracing corresponding points through the pyramid by computing new corresponding locations within the windows on every level.

termined on the coarsest level (matched vertices of segmented zero-crossings). The positions of matched vertices serve as centers of windows on the next level (1024 by 1024 pixel resolution). Here, interest points are determined and matched which in turn are centers of windows on the next level. In this fashion, the original matched vertices are traced through the pyramid and the positions of corresponding points are refined on every level (Figure 3 illustrates that concept).

With a standard deviation of 1 pixel for matched vertices, the maximum error will be ± 3 pixels. Thus, the centers of windows on the next level may be off by ± 6 pixels at most. In order to assure 80 percent overlap between two corresponding windows, their size should be 37 by 37 pixels.

EDGE DETECTION

Edges correspond to abrupt light intensity changes in the image. The subject of determining edges is intensively discussed in computer vision, and many operators with different properties have been proposed. For reasons explained above, we use the LoG operator in most of our experiments.

We briefly summarize. The Gaussian filter,

$$G(x,y) = \frac{1}{2\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{1}$$

is ideal because it optimizes the two conflicting constraints of being limited in the spatial and frequency domain. The second derivative of the smoothed image indicates abrupt intensity changes. The two operations, that is smoothing and taking the second derivative, can be combined. Taking the second derivative of the Gaussian $G(x,y)$, we obtain in Equation 2 the definition of the LoG operator; i.e.,

$$\nabla^2 G(x,y) = \left[\frac{x^2 + y^2}{\sigma^2} - 2\right] e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{2}$$

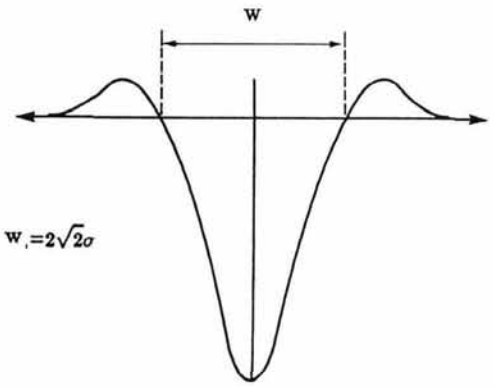

FIG. 4. Cross section through LoG operator. The width of the central lobe, w, determines the degree of smoothing.

Figure 4 shows a cross-section of the LoG operator. From Equation 2 we conclude that the width of the central lobe, w, is related to σ, the parameter of the Gaussian filter, by $w = 2\sqrt{2}\,\sigma$.

Convolving the image $I(x,y)$ with the LoG, we obtain the convolved image

$$C(x,y) = \nabla^2 G(x,y) * I(x,y) \tag{3}$$

Edges are found at locations where $C(x,y) = 0$, called *zero crossings*. The average distance between neighboring zero crossings is approximately w.

We now address the question of the initial w_0 with which the top level of the image pyramid is to be convolved. The selection is motivated by the desire to obtain just enough zero crossings that can be unambiguously matched. Suppose the coarsest scale corresponds to an image resolution of 512 by 512 pixels. In our implementation we have chosen a truncation radius of the LoG operator of $1.8w$. Coefficients outside the radius are smaller than $1/2048^{th}$ of the maximum coefficient in the center. With $w = 15$ pixels, we may expect $(512 - 2 \times 27)/15 \approx 30$ zero crossings, taking the border effects into account. Figure 9 confirms these statistical considerations. Assuming that one-third of zero crossings can be matched, we will obtain over 100 corresponding points.

Because of the discrete nature of data representation, the determination of zero crossings is not without problems. We have developed a method which conceptually corresponds to intersecting the convolution surface with a horizontal plane at convolution value zero as shown in Agouris *et al.* (1989). This assures not only closed contours but also correct connections at T-junctions. Starting with a seed pixel, the neighborhood is expanded until a sign change occurs. Pixels on the boundary are flagged as zero crossings. We use the extended Freeman chain-code representation to avoid the discontinuity from code 7 to code 0.

MATCHING EDGES

In most feature-based methods the matching is performed in identical scan lines, assuming epipolar geometry. Matched pixels of neighboring scan lines are then checked for continuity. However, digital stereopairs are not in epipolar geometry, that is, corresponding points are not in scan lines. Thus, we attempt to match entire zero crossings as proposed, for example, in Schenk and Hoffmann (1986).

A closer examination of Figure 9 reveals that zero crossings in both images are similar in shape, but may be differently connected or may be only present in one image. Therefore, the

problem of finding corresponding edges must be formulated quite general. For a set of edges $L = L_1,\ldots,L_m$ in one image find in set $R = R_1,\ldots,R_n$ of the other image matching edge segments $S = \{S_j S_k\}$ with $S_j \in L$ and $S_k \in R$. It is important to note that neither the beginning nor the end of edges can be assumed to be identical.

The matching criteria are similar shape, similar convolution gradients, and certain consistency constraints in terms of location in both images. What are reasonable primitives suitable for inexact shape matching, independent of rotation and scale differences in both images?

To that end, we transform the zero crossings into the $\psi-s$ domain. It is not possible to repeat the entire development here, but we will attempt to touch on salient points. The reader is referred to more detailed descriptions provided in Li and Schenk (1990a).

Representation of Edges in $\psi - s$ Domain. The $\psi-s$ representation of a line is a function $\alpha = \psi(s)$ where the length s is the parameter of the tangent ψ. The function α is known as the $\psi-s$ curve (see Ballard and Brown, 1982). Some $\psi-s$ representations use the curvature. Using the tangent proved much more robust in our application.

In the chain-code representation of the edges, s corresponds to the number of pixels. The curvature is the difference between neighboring chain codes, and the tangent amounts to the sum of differences.

The following properties of the $\psi-$s representation prove advantageous in matching edges (compare also Figure 5):

- The representation is invariant with respect to the position in the original x,y domain. More importantly, it is rotation invariant.
- Degree of original line is reduced by one. Straight lines are represented as horizontal straight lines, circles as straight lines with the slope being proportional to the curvature.
- Vertices of straight lines cause discontinuities in the $\psi-s$ curve.
- Symmetry of $\psi-s$ curve if the sequence of the original curve is reversed.

The most important advantage of the $\psi-s$ domain is certainly the fact that it reduces the matching to a one-dimensional problem, making an otherwise almost intractable problem solvable.

Matching Edges in $\psi-s$ Domain. The edges represented in the $\psi-s$ domain are now approximated by polygons. Vertices indicate changes in curvature. This piecewise linear approximation corresponds in the original x,y domain to a piecewise *circular* approximation where vertices would be found at positions of large curvature. Changes in curvature convey more information and are more meaningful contrasted with large curvatures.

Edges of similar shape are characterized by similar vertices. That is, the angles as well as the orientation are similar. Let P_{ij} be vertex j of edge L_i in one image. The first stage is concerned with finding all vertices in the other image with similar angles and orientations, resulting in a set of candidate matches $M = M_{io},\ldots, M_{in}$ for every vertex.

The next steps in our matching scheme are governed by the following considerations:

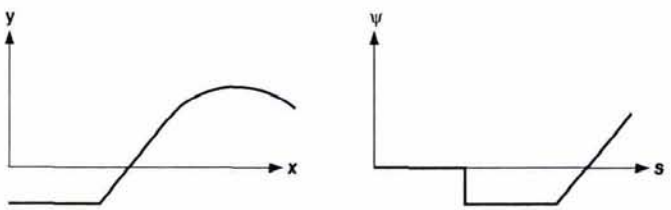

FIG. 5. Lines and curves are shown on the left in the x-y coordinate system and on the right as the $\psi - s$ representation.

(1) For every vertex P_{ij} there exists only one true match M_{io}, assuming the surface is continous.
(2) For the same reason, the x- and y-parallaxes of true matches must be within limits.
(3) In case of several plausible matches, the one with maximum line consistency is chosen.

The second constraint is implemented by a two-dimensional histogram of the x- and y-parallaxes. True matches are clustered around a point, the average model parallax. The size of the cluster in the x-direction is a function of the topography (elevation differences) while the size in the y-direction depends largely on the orientation parameters. This simple but rather effective approach eliminates the majority of wrong matches but preserves all true matches.

The third constraint deals with global consistency. Suppose we have several plausible matching candidates for vertex P_{ij}. The neighboring vertices with their matching candidates are now examined. We would select those matching candidates which belong to the same edge and are somewhat consistent in their sequence.

This method of inexact matching is further complicated by the fact that corresponding edges are not necessarily segmented in the same way in both images. Vertices may be at different locations and their number may be different too. The result of the matching procedure is a list of matched vertices. Figure 10 depicts the matched zero crossings.

In summary, our matching method is independent of rotation differences between the two images. Our experiences indicate that this method may be quite useful for solving matching problems in other applications, such as registering digital images to a map or to other images.

POINT DETERMINATION TO SUB-PIXEL ACCURACY

Theoretically, it is possible to use the matched zero crossings as observed entities for the orientation. In that case, the collinearity model would have to be extended to cope with lines instead of points. A bundle ray, for example, would correspond to a surface defined by the projection center and the zero crossing (sequence of pixels in the digital image). The reader should bear in mind that individual pixels of matched zero crossings do *not* correspond because the images are not registered in epipolar geometry.

In our system, the matched zero crossings serve as approximate locations for determining corresponding points as accurately as possible. From the matching procedure described in the previous section we have a set of matched points which correspond to maximum changes in curvature of zero crossings. These points define centers of windows within which interest points are computed monocularly. The next step is concerned with determining which of the interest points are in fact corresponding (interest point matching). Finally, sub-pixel accuracy is computed for the matched interest points.

The procedure requires that the windows refer to the same object patch. How close are the matched zero-crossings points to corresponding points? It is well known that zero crossings are not accurately located, particularly at positions of significant curvature. Here, the basic assumption of infinitely extended straight edges is violated most. Also, the larger the central lobe of the LoG operator, the larger the deviations from object contours. However, the displacements are similar in both images, because the image function does not change a great deal, except perhaps in areas of occlusions. Our experiments confirm that the majority of matched zero-crossings points are within a few pixels of the corresponding location. In fact, when performing a relative orientation with matched vertices, all points that exceed a y-parallax of 2 pixels are rejected. This assures that image patches centered around matched vertices sufficiently overlap (see also Figure 3).

Determining and Matching Interest Points

We use the interest operator proposed by Förstner (1986). The operator determines corner points in zero crossings in the following fashion:

$$\begin{bmatrix} \Sigma f_x^2 & \Sigma f_x f_y \\ \Sigma f_x f_y & \Sigma f_y^2 \end{bmatrix} = \begin{bmatrix} \Sigma f_x^2 x & \Sigma f_x f_y y \\ \Sigma f_x f_y y & \Sigma f_y^2 y \end{bmatrix} \quad (4)$$

with f_x, f_y the first partial derivatives of the image function $f(x,y)$. The solution is constrained by the request that the error ellipses should be close to a circle and should be as small as possible. The reader is referred to Förstner (1986) and Li and Schenk (1990a) for more details.

The interest operator computes several points in windows of 37 by 37 pixels. We turn now to the problem of determining corresponding interest points.

At the outset we have two sets of interest points, $\{L_1, L_2, \ldots, L_n\}$ and $\{R_1, R_2, \ldots, R_m\}$, corresponding to the windows in the left and right image, respectively (see also Figure 6). Every matched pair must satisfy the following conditions:

- The x- and y-parallaxes cannot exceed certain threshold values. That is $x_{L_i} - x_{R_j} < x_{max}$ and $y_{L_i} - y_{R_j} < y_{max}$.
- For points satisfying the above requirements, the cross-correlation factor of gray levels of a sub-window, size 5 by 5 pixels, is computed. The pair with the highest correlation factor is accepted as the matching pair.

The result is a set of matched interest points. In our example of Figure 6 we find L_1 matched with R_3, L_3 with R_2, and L_4 with R_5. No acceptable match was found for L_2, L_5, R_1, R_4, R_6.

Determining Corresponding Points to Sub-Pixel Accuracy

Finding the precise location of corresponding points is achieved in two steps. First corresponding pixels are found at locations of maximum correlation. Then sub-pixel accuracy is computed by interpolating correlation values in a 3 by 3 window.

Suppose we are concerned with the first matching pair L_1, R_3. The corresponding position to L_1, R_3^c, may be off from the position of R_3 by a few pixels. The correlation coefficients for an area of 7 by 7 pixels, with R_3 at the center, is computed, based on a 5 by 5 correlation window. Figure 7 illustrates the concept for pixel $-3, 2$, assuming a local coordinate system with R_3 as origin. The maximum correlation coefficient within the 7 by 7 window is considered to be the corresponding pixel R_3^c to L_1.

Finally, the sub-pixel location R_3^{sub} is found within pixel R_3^c by determining the maximum of the surface constructed by the correlation coefficients cc in a 3 by 3 window around R_3^c. Equation 5 is used to approximate cc: i.e.,

$$cc = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2. \quad (5)$$

The coefficients $a_0 - a_5$ are obtained by a least-squares adjustment of the second order two-dimensional orthogonal polynomial. In the interest of brevity, we skip the detailed derivations and refer the interested reader to Li and Schenk (1990b) and Haralick (1984). The sub-pixel location x_{sub}, y_{sub} is obtained from Equations 6 and 7: i.e.,

$$x_{sub} = \frac{a_2 a_4 - 2a_1 a_5}{4a_3 a_5 - a_4^2} \quad (6)$$

$$y_{sub} = \frac{2a_3 x_{sub} + a_1}{a_4} \quad (7)$$

All corresponding points are now checked for global consistency in a relative orientation with blunder detection. All accepted points are transferred to the next level of the image pyramid where they serve as initial approximations. The procedure is repeated until the final level with the highest resolution is reached.

Finally, we would like to comment on the use of interest points. Because they are *not* corresponding points *per se*, one could argue not to use them. After all, corresponding points are found by gray level correlation in an area of 37 by 37 pixels centered around the matched vertices of zero crossings. For one thing, matched interest points are better approximations than are vertices. But more importantly, interest points signal locations of significant changes in gray levels. In turn, more reliable correlation coefficients can be expected.

RESULTS

In this section we present experimental results. All the stereopairs we have been using are digitized aerial photographs. The geometrical relationship between aerial photographs and their digitized versions, the pixel image, is established by a projective transformation. The eight transformation parameters are determined by a least-squares adjustment with reference points known in both images. We used fiducial marks as reference points and, in model "Campus," reseau marks. Depending on the aerial camera used, the fiducial marks are printed at specific locations, and their shape is unique, too. Therefore, it is not difficult to develop an automatic procedure to detect and subsequently compute the precise location of the fiducial marks.

The stereopair, shown in Figure 8, is part of a block covering the campus area of The Ohio State University. The photoscale is approximately 1:4000. We used our reseau camera RMK-AR from Zeiss, equipped with a wide-angle cone. The diapositives

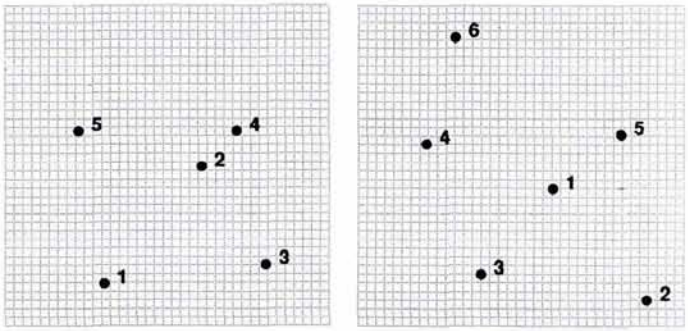

FIG. 6. Windows of 37 by 37 pixels, centered around matched vertices. The interest operator detected five points in the left window and six in the right window.

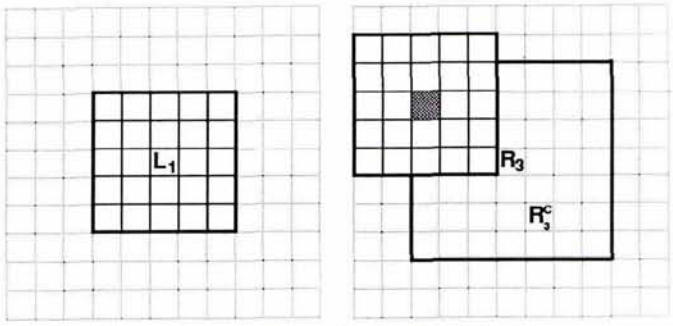

FIG. 7. Matching interest points to sub-pixel accuracy.

were scanned by The Intergraph Corporation, using their new scanner with a pixel-size resolution of 30 micrometres. To be consistent with the other models, however, a resolution of only 60 micrometres was used. This corresponds to 4096 by 4096 pixels.

We measured 25 well distributed reseau points (crosses) on Intergraph's Interpro 3055, and determined the interior orientation of the pixel image with a projective transformation to an accuracy of one third of a pixel.

Figure 9 shows the zero crossings obtained with a LoG, width $w = 15$ pixels. The matched zero crossings are displayed in Figure 10. The matching process, described in the section on edge detection, rendered 179 vertices which were checked in a relative orientation for consistency with the collinearity model. With a threshold value of 1 pixel, 69 points were rejected. The standard deviation of the remaining 110 points is one pixel or half a millimetre. Note that this procedure is only performed on the top level of the image pyramid (512 by 512 pixels).

In Figure 11 we show a window of 37 by 37 pixels from the 2048 by 2048 image pyramid, left and right image. The windows are centered around matched vertices. Also, the interest points are shown as detected with the Förstner interest operator. Matched interest points on this level of the pyramid serve as new window positions on the 4096 by 4096 pixel resolution

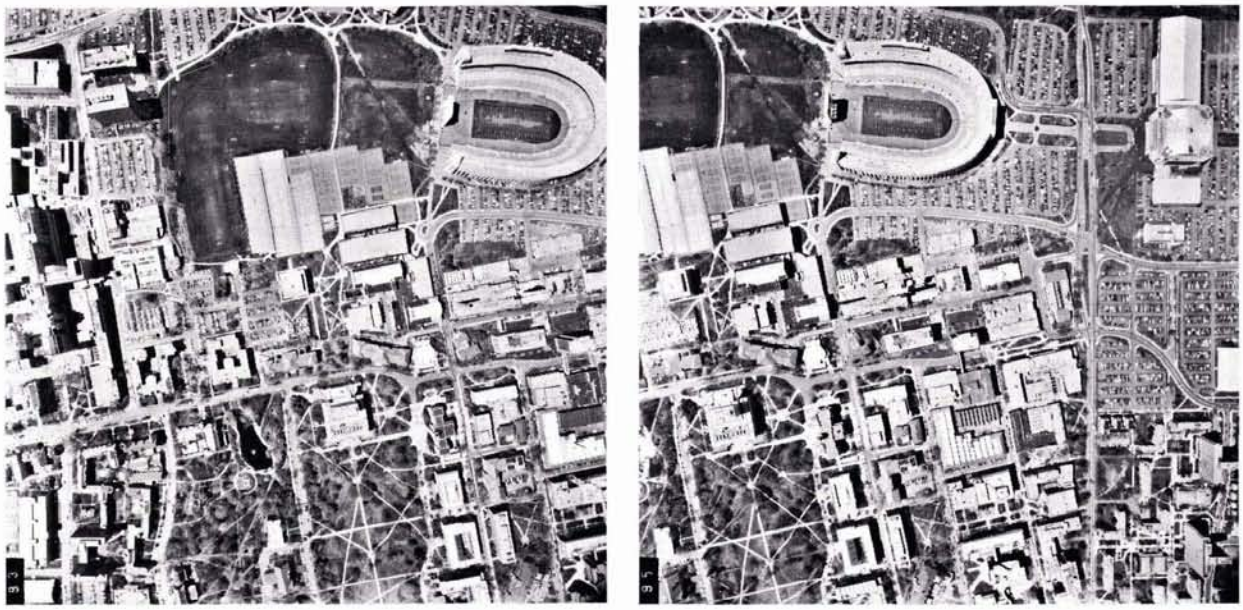

FIG. 8. Model "Campus." Photoscale approximately 1:4000 (reduced to approximately 1:11,400).

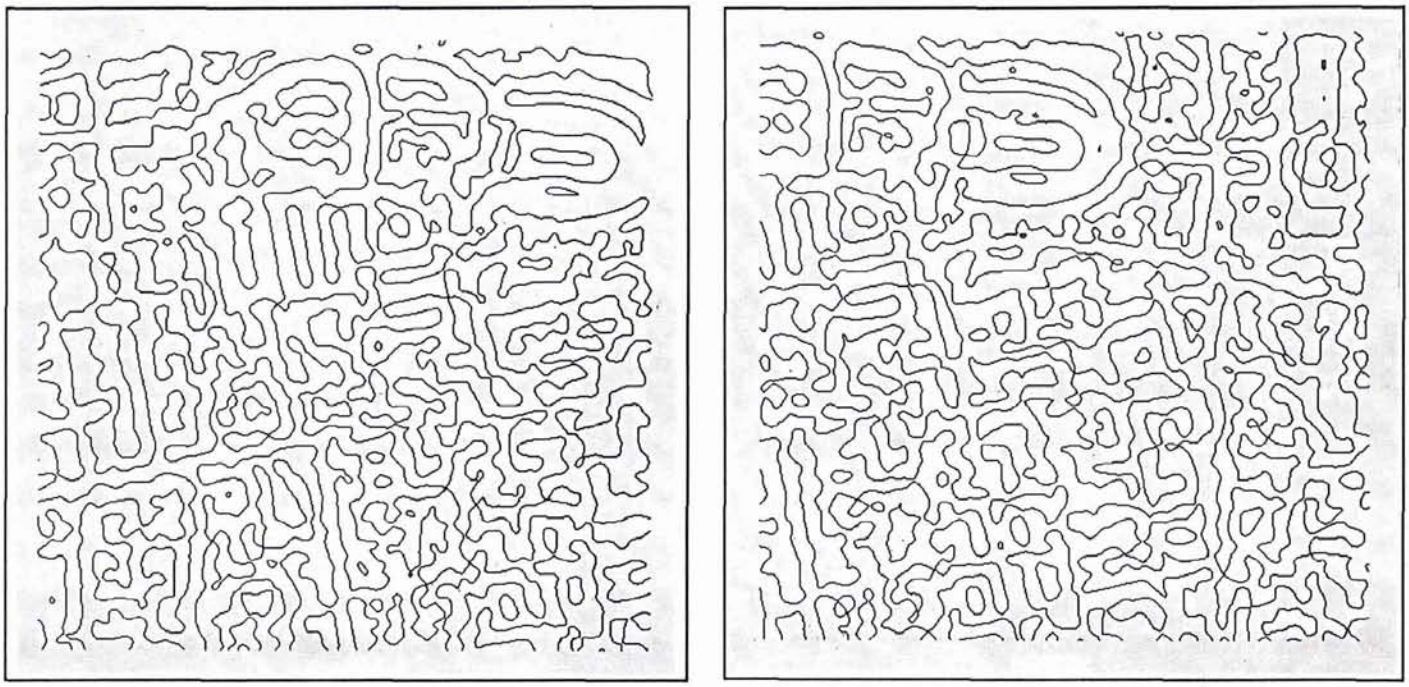

FIG. 9. Zero crossings obtained with a LoG operator, $w = 15$ pixels on the coarsest level of the image pyramid (512- by 512-pixels resolution).

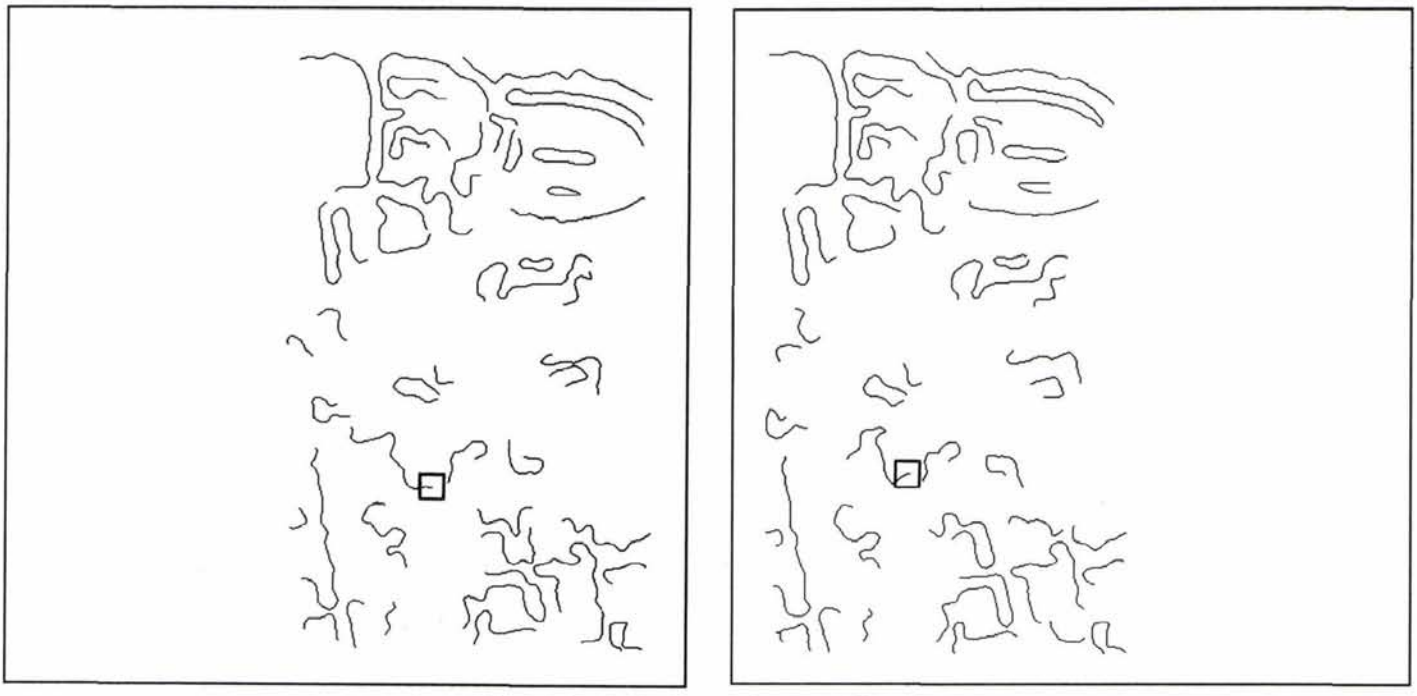

FIG. 10. Matched zero crossings from Figure 9.

level. The matched points on all levels are always checked for blunders.

Finally, the exterior orientation of the two images is determined by a rigorous bundle adjustment. The program copes with any number of corresponding points. The reduced normal equation matrix contains only the exterior orientation parameters. The simple, but robust, blunder detection module takes advantage of the large redundancy. The exterior orientation parameters are relatively stable, even in the presence of blunders. In a first pass, the orientation parameters are determined. Now, every point is computed by spatial intersection and the residuals are checked with a tolerance that is a function of the standard deviation obtained by the adjustment. Points exceeding this tolerance are flagged as blunders and do not take part in the following adjustment. However, blunders are checked again with the new orientation parameters.

We identified and measured the control points manually on the Interpro 3055 with an estimated accuracy of one third of a pixel, or 20 micrometres. For the campus mapping project, mostly natural points were selected for control points. The accuracy is approximately 2 cm. The adjustment is performed with photo coordinates obtained from the pixel image by using the inverse projective transformation. As shown in Table 1, the standard deviation is 0.5 micrometres for x and 4.2 micrometres for y.

As a final quality control test, we placed the same model on the Zeiss PC1 analytical plotter, performed an absolute orientation, and visited all the points computed automatically. An experienced operator measured the difference in z between the points and the ground. That is, the x-parallax, which cannot be determined in the bundle adjustment, was recovered. The RMS computed from the measured differences is 5.1 micrometres, corresponding quite nicely to the standard deviation in the y-direction from Table 1.

CONCLUSIONS

We have presented the automatic orientation, which is the first module of a conceptual system to produce maps automatically, and have elucidated some principles of our approach.

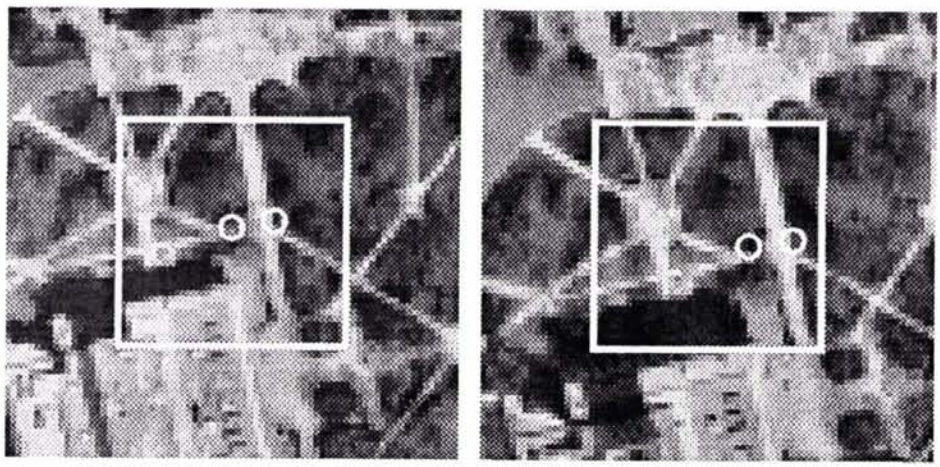

FIG. 11. Windows of 37 by 37 pixels (white squares) centered around the matched vertex pair highlighted in Figure 9. The windows are from the image pyramid, resolution 2048 by 2048 pixels. The two matched interest points serve as centers for windows on the next level of the image pyramid.

TABLE 1. RESULTS OF AUTOMATIC ORIENTATION PERFORMED ON FOUR DIFFERENT STEREO MODELS. ALL MODELS HAVE BEEN DIGITIZED WITH A RESOLUTION OF 4096 BY 4096 PIXELS.

Model	Number of Points	Blunders	σ [μm] x	y	Control Points	Photo Scale
Campus	112	12	0.5	3.5	6	1 : 4,000
Hunter-Ligget	112	15	0.2	3.3	4	1 : 80,000
Munich	88	11	1.4	3.9	6	1 : 3,800
Daun-Mehren	121	13	1.5	3.7	6	1 : 12,000

The basic notion is to determine approximations by exploiting scale space methods and matching entire edges, in our case zero crossings. It is important to note that the method does not depend on the LoG operator. That is, edges obtained with other operators can be used instead of zero crossings. In our experiences, matching entire edges as opposed to more traditional point matching methods substantially increases the robustness of the solution. On the other hand, it is more difficult to implement. In fact, edge matching presented the most difficult task, and there is still room for improvement. Currently, we are exploring the potential of graph matching for the task of edge matching.

While matched edges are more reliable than matched points, they are less accurate. Clearly, the highest accuracy is obtained by correlating gray values of small image patches, assuming that the image patches cover the same surface patch, are not too grossly distorted, and yield a reasonable signal. We have combined the advantages of edge matching with gray level correlation in a hierarchical approach. The results, which we have obtained from several stereopairs of varying scale and ground coverage, are encouraging and evidence seems to be accumulating in the right direction. We see great promise in this approach and what it will bring to bear on future work.

Future experiments will involve convergent photography and close-range applications to test the limits of edge matching. We expect that our method will work as long as the edges in both images are still similar in shape. If this does not hold, for example in cases of extreme base-height ratio or large convergent angles, objects (and their boundaries which are detected as edges) are imaged very differently. Matching, then, could no longer rely on similar edges, but would have to include image segmentation, their classification, and matching.

Identifying and measuring control points as necessary for the orientation with respect to the ground control system is presently performed manually. Some of our research efforts are aimed towards finding control points automatically by using associative memory techniques and neural networks. From admittedly limited experiences, it appears that control points whose shape is similar to a pattern stored in a library can be located, even in a noisy background as shown in Al-Tahir *et al.* (1990).

ACKNOWLEDGMENTS

Funding for this research was provided in part by the NASA Center for the Commercial Development of Space Component of the Center for Mapping at The Ohio State University. The authors would like to thank the following: The Intergraph Corporation for partial funding of this research and for digitizing several models of the campus mapping project; The U.S. Army Engineer Topographic Laboratories, Fort Belvoir, for making the model "Hunter-Ligget" available; and Carl Zeiss, Inc. for providing the models "Munich" and "Daun-Mehren."

REFERENCES

Agouris, P., T. Schenk, and A. Stefanidis, 1989. Zero-Crossings for Edge Detection. *Proc. ASPRS-ACSM Fall Convention*, pp. 91–99.

Al-Tahir, R., Ch. Toth, and T. Schenk, 1990. Automatic Ground Control Point Recognition with Parallel Associative Memory. *Proc. ACSM-ASPRS Fall Convention*, pp. B-10–B-18.

Ballard, D.H., and Ch. M. Brown, 1982. *Computer Vision*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Greenfeld, J., 1987. *A Stereo Vision Approach to Automatic Stereo Matching in Photogrammetry*. Rep. No. 381, Department of Geodetic Science and Surveying, The Ohio State University.

Greenfeld, J., and T. Schenk, 1989. Experiments with Edge-Based Stereo Matching. *Photogrammetric Engineering & Remote Sensing*, Vol. 55, No. 12, pp. 1771–1777.

Li, J.C., and T. Schenk, 1990a. *Aerial Image Matching Using ψ-s Representation*. Techn. Notes in Photogrammetry No. 4, Dept. of Geodetic Science and Surveying, The Ohio State University.

———, 1990b. *Determination of Relative Orientation of Digital Stereopairs*. Techn. Notes in Photogrammetry No. 5, Dept. of Geodetic Science and Surveying, The Ohio State University.

Förstner, W., 1986. A Feature Based Correspondence Algorithm for Image Matching. *Proc. Int. Soc. for Photogr. and Remote Sensing*, Symp.Comm. III, Vol.19.

Haralick, R.M., 1984. Digital Step Edges from Zero-Crossings of Second Directional Derivatives. *IEEE, Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-6, No. 1, pp. 58–68.

Marr, D., and T. Poggio, 1979. A computational theory of human stereo vision. *Proc. R. Soc. Lond.* B 204, pp. 301–328.

Schenk, T., and O. Hoffmann, 1986. Stereo Matching Using Line Segments of Zero-Crossings. *Proc. Int. Soc. for Photogr. and Remote Sensing*, Symp.Comm. III, Vol.19, pp. 362–368.

Schenk, T., J.C. Li, and Ch. Toth, 1990. Hierarchical Approach to Reconstruct Surfaces by Using Iteratively Rectified Images. *Proc. Int. Soc. for Photogr. and Remote Sensing*, Symp.Comm.V, Vol.28, Part 5/1, pp. 464–470.

(Received 11 December 1990; accepted 9 January 1991)