Experiments with Edge-Based Stereo Matching

Joshua S. Greenfeld
Department of Civil & Environmental Engineering, New Jersey Institute of Technology, Newark, NJ 07102
Anton F. Schenk
Department of Geodetic Science & Surveying, The Ohio State University, Columbus, OH 43210

ABSTRACT: Edge matching is the most critical step in feature-based stereo vision. Edges, as obtained by various operators on the image, are discontinuities in the light intensity function. In general, they correspond to physical boundaries in the object space. Here we are concerned with two methods of matching edges in digital stereopairs. In one method the edges are segmented and approximated by polygons. Corresponding edges are found by comparing vertices, leg orientation, and approximate location within the image. The second method described is based on the ψ - s representation (to be defined later) of the edge. Thus, the search space for matching is reduced by one dimension. Results obtained with these matching methods are reported.

INTRODUCTION

FOR OVER THREE DECADES, automating the stereo matching process has been one of the most challenging tasks in photogrammetry. In recent years the focus on the solution for this problem has shifted from gray level correlation to feature-based matching. There are two reasons for this transition. First, it has been realized that gray-level correlation did not live up to the promise of an automatic stereo plotter. The dependency on guidance from the human operator and on favorable image conditions could not be resolved by this matching solution. The second reason for the shift is the extensive work on stereo matching done in computer vision. Aside from a handful of correlation-based solutions, feature-based solutions constitute the overwhelming majority of published work on stereo in computer vision.

Feature-based solutions can be divided into point features and edge features. In a point-based solution, distinct points (interest points) are extracted from each image and matched by means of similarity and consistency measures. Edge-based solutions follow a similar procedure. Edges (not points) are extracted from each image by an edge detector and are subsequently matched according to a similarity and consistency criteria. In photogrammetry, most of the suggested feature-based matching systems utilize point features. In computer vision, however, the trend is different. The tendency is toward using the edge-feature solution. Third diversity can be explained by the primary objectives for performing the match. The objective of stereo matching in photogrammetry is to provide accurate three-dimensional (3D) coordinates of object-space points. This is done by measuring corresponding points on the images and processing them through the collinearity equations. Thus, point-features seems to be a very appropriate solution. The objective of stereo matching in computer vision is image understanding. For this purpose, edges are a more appealing solution. Edges convey rich information about size, orientation, and shape which are essential parameters of pattern recognition and image understanding.

This paper deals with edge feature-based matching. Previously, this approach was successfully applied to stereo matching in photogrammetry (Greenfeld, 1987). It was shown that stereo matching can be achieved, even when only very limited constraints are imposed on the image configuration.

Edge representation methods are discussed in the next section. Specifically, we elaborate on polygon approximation and ψ - s representation of edges. In the third section, we describe the edge matching method that has been developed. In the fourth section experiments, examples, and results are discussed. Finally, concluding remarks are presented.

EDGE REPRESENTATION

Stereo matching is the process in which two representatives (edges) of the same object are paired to establish a 3D model of the object. These edges are extracted from images which were recorded from at least two distinct points of view. To reduce the tremendous search space for matchable pairs of edges, corresponding edges can be found either by imposing constraints on the process, such as compliance with epipolar geometry, or by analyzing the characteristics of these edges.

Shape information and structural descriptions of an edge can be obtained in several ways. The edge can be mapped into different parameter spaces such as the spatial domain/frequency domain, Hough space, or the ψ - s space, to name only a few. A description of these mapping spaces and references to previous work can be found in Ballard and Brown (1982). Here we are concerned with two specific matching methods. The first is matching polygon approximations of the edges. The other is matching the ψ - s representation of the edges.

APPROXIMATING EDGES BY A POLYGON

Polygon approximation is the process in which break points are marked on a (digital) curve. These points are used to approximate the continuous curve. They should be selected in such a way that the structural and shape characteristics of the curve will be maintained. In other words, the polygon approximation must convey the same information as the fully extended curve. Polygon approximation is used for several tasks. Among others, it is used for data compression, pattern recognition (object matching), and cartographic line generalization. Different tasks may dictate different requirements from the approximation process. Some may require the ability to reconstruct the original curve perfectly (data compression), while others may require smoothing and simplification of the curve (generalization). For matching purposes, we require that points with maximum local curvature be marked as break points. This requirement follows the observation (Attneave, 1954) that most of the shape information of the curve is concentrated in these points.

The curvature is defined as the rate of change of the slope as a function of arc length. For $y = f(x)$, the curvature is

$$k = \frac{y''}{\sqrt{(1 + (y')^2)^3}} \tag{1}$$

where y is the curve function, and y', y'' are the first and second derivatives of y, respectively.

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 55, No. 12, December 1989, pp. 1771-1777.

0099-1112/89/5512–1771$02.25/0

This definition of the curvature holds for continuous functions in which the first and the second derivatives exist. A digital curve is defined in the discrete space; thus, curvature is not rigorously defined. There are, however, a few methods to compute an approximation to the curvature at each point along the curve. A straightforward approximation is to use the chain code of the curve as the function y (Ballard and Brown, 1982). The chain-code function is a sequence of integers between 0 to 7 (or 0 to 3) which indicate the 1–8 (or 1–4) direction (neighboring pixel) to the next point of the curve. The derivatives of y, namely, y' and y'', can be computed by the first and the second differences, of the chain code, respectively. Other examples for computing the curvatures along the digital curve are [Teh and Chin, 1988]

(1) A cosine measure $C\,S_{pk}$: i.e.,

$$C\,S_{pk} = \frac{a_{p(-k)} \cdot b_{p(+k)}}{|a_{p(-k)}| \cdot |b_{p(+k)}|} \tag{2}$$

where $a_{p(-k)}$ is the vector from point p to point $(p - k)$ and $b_{p(+k)}$ is the vector from point p to point $(p + k)$; or

(2) A curvature measure CR_{pk}: i.e.,

$$C\,R_{pk} = \frac{1}{k}\sum_{j=-k}^{-1} f_{p-j} - \frac{1}{k}\sum_{j=0}^{k-1} f_{p-j} \tag{3}$$

where f_{p-j} is an integer corresponding to the chain code value of point $(p-k)$.

The parameter k is called the extent of the region of support. Large values of k will result in smoothing of fine, but important, details. A small region support will mark local insignificant fluctuations along the course of the curve as break points. Thus, k controls whether we will detect too many, too few, or just the right amount of break points. One should devise a method which will produce desired break points independently of the size of k.

In order to overcome the region support problem, we suggest the following procedure. First, curvature points are detected using a small value of k (say $k = 6$). The intention of selecting such a region of support is to ensure the detection of all significant maximum local curvatures. A geometric thresholding algorithm is subsequently applied to the overdetermined break point list. Thresholding will eliminate local spikes which have a local maximum curvature; however, they do not represent a significant change in the orientation of the curve.

Representing an Edge by a $\psi - s$ Curve

The $\psi - s$ curve is a representation of a curve in a $\psi - s$ space. This is a two-dimensional rectangular space with one axis (ψ) mapping cumulative orientation changes along the course of the curve, and the other axis (s) is the cumulative curvilinear length of the curve (Ballard and Brown, 1982; O'Neill and Mark, 1987). The orientation of the curve (i.e., the tangent) at any given point can be found be evaluating the first derivative of the function of the curve at that point. A digital curve is a list of pixels which, when connected, form a continuous curve. Edges detected from a digital image are examples of digital curves. As we discussed earlier, a simple and very useful function of a digital curve is the chain code (Freeman, 1974). Using the chain code as the curve function, we obtain

$$\psi_i = \sum_{j=1}^{i} (f_{j+1} - f_j) \tag{4}$$

$$s_i = i \tag{5}$$

where ψ_i and s_i are the $\psi - s$ representation at point i along the curve and f_j is an integer corresponding to the chain code value of point j.

There are several important characteristics of the $\psi - s$ curve which need to be mentioned. First, the degree of the polynomial required to represent a certain shape is reduced by one. For example, a straight line from the (x,y) space, regardless of its orientation, is mapped into a horizontal straight line (a constant). A circle in from the (x,y) space, is mapped to a inclined straight line. The slope of this line is a function of the radius of the circle. Thus, the geometric parameterization of the curve is simplified. A second characteristic of the $\psi - s$ curve is that it is orientation invariant. Thus, the same features in the (x,y) space will be mapped into the same representation in the $\psi - s$ space regardless of their orientation. The importance of invariant representation will be discussed in the next section. The $\psi - s$ curve, however, is scale dependent as its length depends on the size of the original curve.

Matching $\psi - s$ curves is essentially another problem of edge matching. In other words, we are presented with two curves, one from each image, which need to be matched. The only difference is that we have to match simpler geometric features. An obvious choice to perform the matching is to use the polygon approximation method previously described. Instead of using the original edge as an input to the program, the $\psi - s$ curve is being processed.

EDGE MATCHING

Matching is the most difficult task of the stereo process. Any edge (or its representation, i.e., a polygon) on one image has to be compared and evaluated against all the edges on the other image. The matching is basically a selection process in which edges are paired according to some measures of similarity and consistency. First, an initial list of matches is established by singling out similarities in the characteristics of the edges. Next, a consistency check takes place to eliminate erroneous matches. Thus, it is not enough to find edges which "look alike," but they must also agree with the pattern of the surrounding matches.

Similarity measures between polygon approximations can be done by comparing their vertices and/or their legs (sides). In earlier work (Medioni and Nevatia, 1985; Boyer and Kak, 1986) the orientation, the end point coordinates, the mid-point coordinates, and the length of the polygon legs have been used for stereo matching. In addition to these geometric primitives (properties), radiometric primitives may be used for similarity assessments. An example of a radiometric primitive is the edge's strength. This primitive will be explained later. In both Medioni and Nevatia (1985) and Boyer and Kak (1986), an epipolar geometry is assumed to prevail; thus, comparing location in terms of coordinates (especially y) may yield satisfactory results. If one does not restrict the matching to an epipolar geometry, coordinate comparison must be avoided.

The consistency check can be done by geometric or statistical tests. The most common approach is to devise a probability analysis for a correct match, given that other matches are correct as well (Medioni and Nevatia, 1985). An example of a geometric consistency check is an affine transformation to map one image into the other. The transformation parameters are computed from all the matches, and those which display large discrepancies in the mapping are discarded.

Similarity Parameters

Our approach to polygon matching is based on vertices rather than on polygon legs. This results from our attempt to develop a more general solution which does not require an epipolar geometry. The epipolar geometry is solved as an integral part of the matching process, but is not required *a priori*. Polygon legs are not a good matching primitive because they are sensitive to the segmentation process. Corresponding edges usually have a slightly different shape due to the perspective projection, among

other reasons. This may result in some extra break points in one image. Thus, the leg's length, for instance, may vary significantly only because of the segmentation process.

To establish the initial list of matches, the following similarity primitives have been used:

- Angle at the vertices.
- Orientation of the vertices.
- Strength.
- Zero crossing sign.

These primitives are illustrated in Figure 1. The angle of the vertex is obtained from the difference in the azimuths of the adjacent legs. The orientation of the vertex can be defined as the azimuth of its bisector which corresponds to the average azimuth of the adjacent legs. A reasonable approximation to this primitive is the orientation of the polygon leg preceding the vertex. The third primitive; strength, is usually a measure of magnitude of the edge's gradient. The LoG (Marr and Hildreth, 1980) edge detector, which we used for a variety of reasons (Greenfeld, 1987), does not provide a gradient value directly. Instead, the strength measure we used was the steepness of the zero crossings. The steepness measure 's' is shown in Figure 1b. A large value for 's' means a strong edge or an abrupt change in image and vice versa. The zero crossings provide us with the fourth similarity measure. The vertex can be positive or negative (for example the vertex in Figure 1a is negative) depending on the sign of the convolved image at the vertex.

Consistency Check

The initial list of pairs of similar vertices contain many incorrect and ambiguous matches. The goal of the consistency check is to eliminate these incorrect matches and resolve ambiguities. We developed several voting strategies to achieve this goal. The voting schemes are used to study the geometric pattern and the inter-relationships between correct matches. A particular configuration which achieves the highest score (most frequent) among all the matches is assumed to be the correct one. All matches that do not agree with this configuration are assumed to be incorrect. The voting is performed sequentially, that is, matches which failed in a voting scheme are eliminated from the matching process and do not participate in the next voting.

The first consistency check is to compute the vertical disparity (y parallax) of matches, assuming zero rotation angles between the images (photographs). A space resection (i.e., relative orientation) is computed using the coordinates of the vertices, while the rotation angles are set to zero. A histogram of the residuals (P_y) is established for some interval p. The size of the interval can be relatively large (several mm in image coordinate system) as this voting is used only to eliminate the low occurences of certain parallaxes on both ends of the histogram. The effect of the verticality assumption (no rotations), on a non-vertical stereo model, will be that the highest parallax frequency will not necessarily correspond to zero but to some bias value. This does not pose any problem because we are not looking for small values for the parallaxes but rather for high occurrence of a certain parallax.

The second consistency check is to vote on the most probable azimuth and distance between pairs of matched vertices. The image coordinates (or the pixel coordinates) of the vertices are used to compute the azimuth and distance from the left vertex to the right one (see Figure 3.). Because the images are displaced by the model base, one can expect this azimuth (assuming y axis is "north") to be about 90°, and the distance roughly the same as the base. As before, a histogram is established to determine the most probable azimuth and then the most probable distance. The histogram interval for the distance must be large because differences in the distance are not necessarily an indication for an incorrect match. They can be caused by relief displacement. However, unreasonable distances, such as less than one half of the base, can be discarded.

Finally, the relative orientation between the images is computed by means of a least-squares adjustment. Large residuals are assumed to indicate incorrect matches. The elimination of large residuals (gross errors) can be done manually or automatically using robust estimation or another gross error detector. Experience shows that the number of matches is relatively high; thus, a statistical analysis of the residuals is based on sound ground. A refinement of this process is to extract interest points around the vertices, match the interest points, and use them in the relative orientation computation (Greenfeld, 1988).

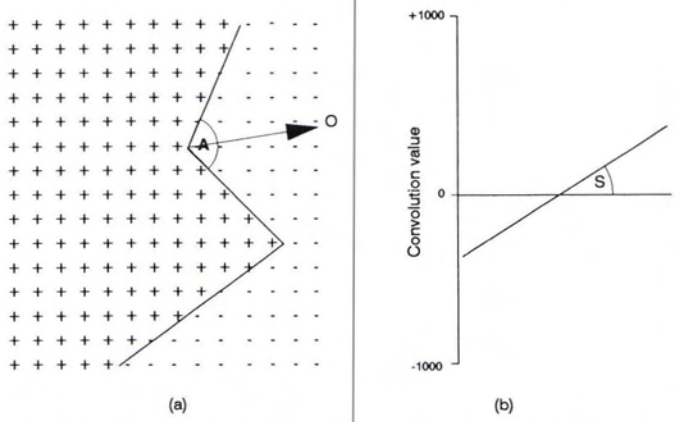

FIG. 1. Similarity primitives of an edge which was detected by the LoG operator. (a) Angle (A), orientation (O), and sign of a vertix. (b) A cross section at an edge showing the zero crossing and its strength or steepness.

EXPERIMENTS AND RESULTS

The edge matching methods discussed earlier have been implemented and tested on several images. In this section we report on some implementation details and results. We do not describe a complete matching system but rather concentrate on the process of matching edges. The complete matching system includes the edge detection process and an intensive coarse to fine strategy to gradually densify the matches. An overall description of the system can be found in Greenfeld (1987, 1988).

Figure 2 shows the stereo image which we used to demonstrate the matching process. It also shows the particular edge resolution which we used as an example. This edge was extracted from a very course description of the image which depicts only its most pronounced objects. The edge was obtained by a LoG operator (Marr and Hildreth, 1980) with $\sigma = 12.4$ (or $w = 35$). The reason for using this coarse resolution is that it is easier to demonstrate and follow the matching process.

Polygonal Approximation

Figure 3 shows the results of the polygon approximation of the edges from Figure 2. The approximation process followed the procedure described earlier. Namely, break points with maximum local curvature were detected and a geometric thresholding process was applied to eliminate those break points which are insignificant. The break points were detected by applying a cosine curvature measure with a $k = 6$ region of support. This choice results in many break points that mark the windings of the edge in detail. A tolerance band (Boyer and KaK, 1986) method was subsequently applied to eliminate break points that lie within two pixels from the line connecting their adjacent break points.

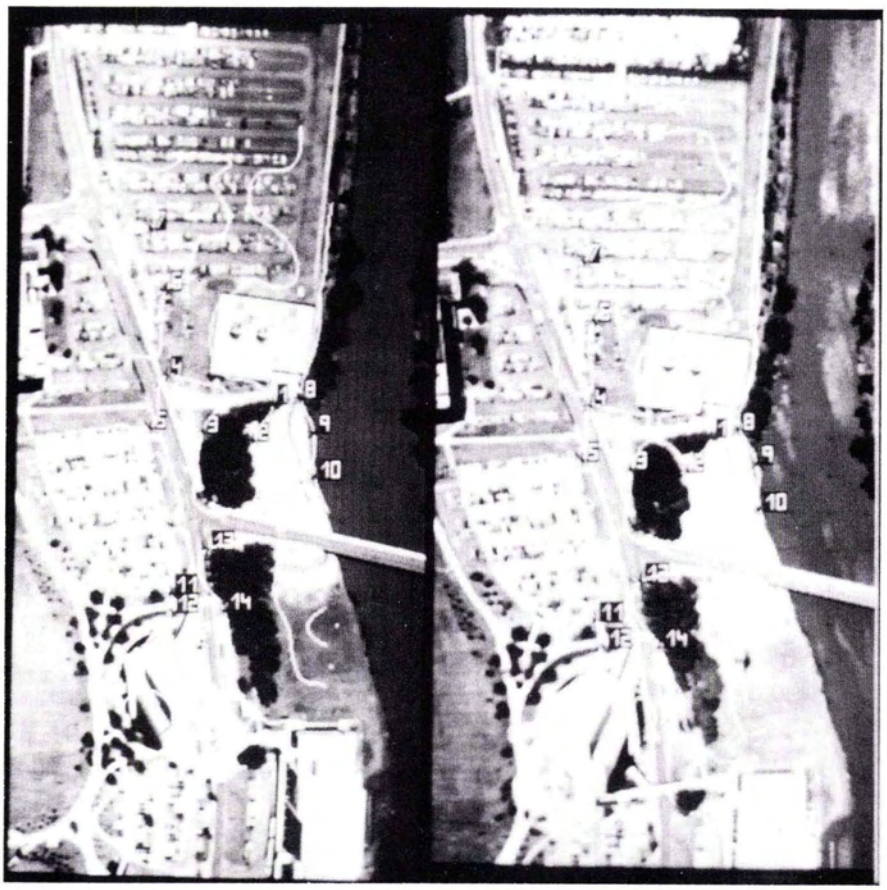

FIG. 2. A stereo pair of aerial images. The edges were detected with LoG (w = 35). The numbers indicate matched vertices.

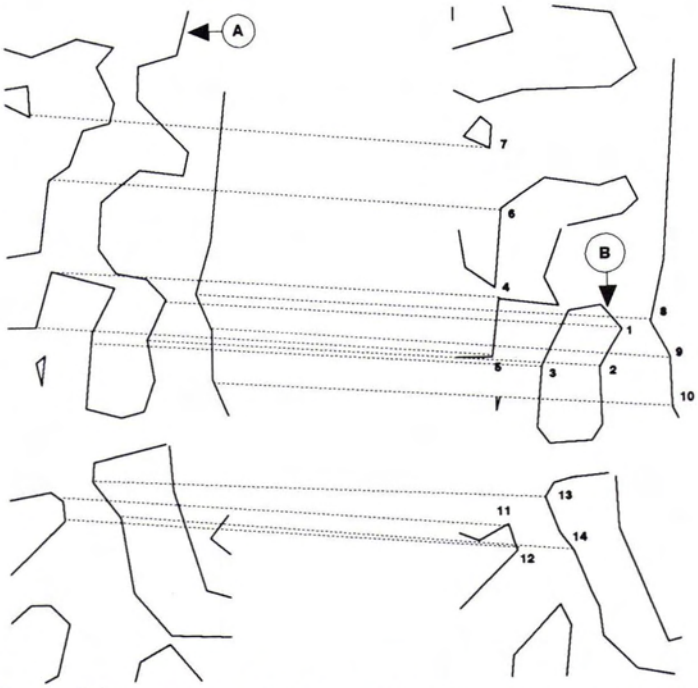

FIG. 3. Polygon approximation of the edges from Figure 2. The lines connecting corresponding vertices illustrate the azimuth and distance consistency checks.

Next, a list of initial matching candidates was established. The list was based on the complying with the four similarity measures discussed in the previous section. The tolerance for similar angle and orientation of the vertices was ± 15°. The strength tolerance was ±20 percent and, naturally, the sign of the zero crossing had to match exactly. The tolerance values were chosen to accommodate variations between the left and right images due to geometric and radiometric distortions. Lenient tolerances carry a price of increasing the number of incorrect matches in the initial matching list. However, this problem is resolved during the consistency check. Generally, consistency is a more profound measure of correctness of a match than similarity. In our example, 90 pairs of vertices have been found as initially potential matches.

These initial matches have been tested for consistency following the voting procedures described earlier. Histograms were established to obtain the most frequent y-parallax, azimuth, and distances. These histograms are presented in Figure 4. The main parameter which controls the frequency results in a histogram is the size of the class interval. For the y-parallax histogram, a large class interval is in order because of the zero rotation approximation. This, one cannot expect all the correct matches to correspond to a narrow class interval. In our matching system the class interval for this voting is 30 pixels. The central interval ranges between −15 and +15. For the azimuth voting, the class interval is 10° and for the distance it is ten pixels. In the event of two adjacent intervals with almost the same frequency, matches that fall within both intervals are accepted and carried on to the next consistency check. The evaluation of the distance voting

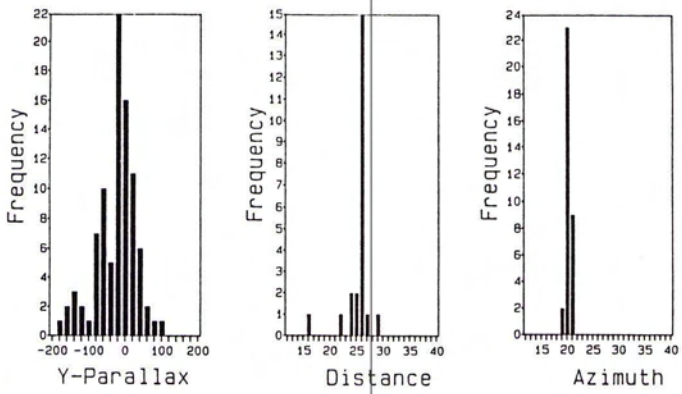

FIG. 4. Frequency histograms to select the most probable pattern for correct matches.

is somewhat different as not only the most frequent class is accepted, but also its neighboring intervals. As we discussed earlier, this is because distance between correct matches may vary as a function of elevation. From the initial 90 potential matches, 34 were selected by the y-parallax voting, 23 by the azimuth voting, and 18 by the distance voting. Four more points were eliminated by computing an affine transformation between the corresponding vertices from the left and the right images.

The resulting matches are shown in Figure 2. The small number of matches is due to the coarse edge representation and the coarse approximation of the edge. The same image with edges detected by the LoG with $\sigma = 5$ produced 62 matches. The matching system was tested on other images as well. Figure 5 shows the results of matches found from edges detected with $\sigma = 5$. In this example 67 matches were found. A visual inspection didn't reveal any incorrect matches. A statistical inspection can be done by computing a relative orientation from these matches and analyzing the residuals. Examples of such statistical analysis can be found in Greenfeld (1987, 1988).

$\psi - s$ METHOD

The $\psi - s$ method was tested on some limited number of edges rather than on all the detected edges. The main purpose of these tests was to assess the suitability of this method for edge matching. Some edge segments which seemed to be very similar could be matched only partially with the polygon approximation method. Thus, it seemed to be appropriate to search for other matching methods to complement the one previous described. Tests with the $\psi - s$ method yielded extremely good results so that in the future we intend to integrate the $\psi - s$ method into our matching system.

Figure 6a shows an edge from the left image (see Figure 3 for the polygon approximation of edge "A"). The corresponding edge from the right image is depicted in Figure 7a ("B" in Figure 3). The $\psi - s$ representation of these edges is shown in Figures 6b and 7b, respectively. The two edges are very similar; however, due to segmentation differences the polygon approximation method did not match the entire edge form Figure 7a with the one in Figure 6a. Comparing the $\psi - s$ representations of these edges shows a complete match in very simple geometric parameters. All one needs to do in order to match Figures 6b

FIG. 5. Another example of a stereo pair of aerial images. The edges were detected with LoG ($w = 15$). The numbers indicate matched vertices.

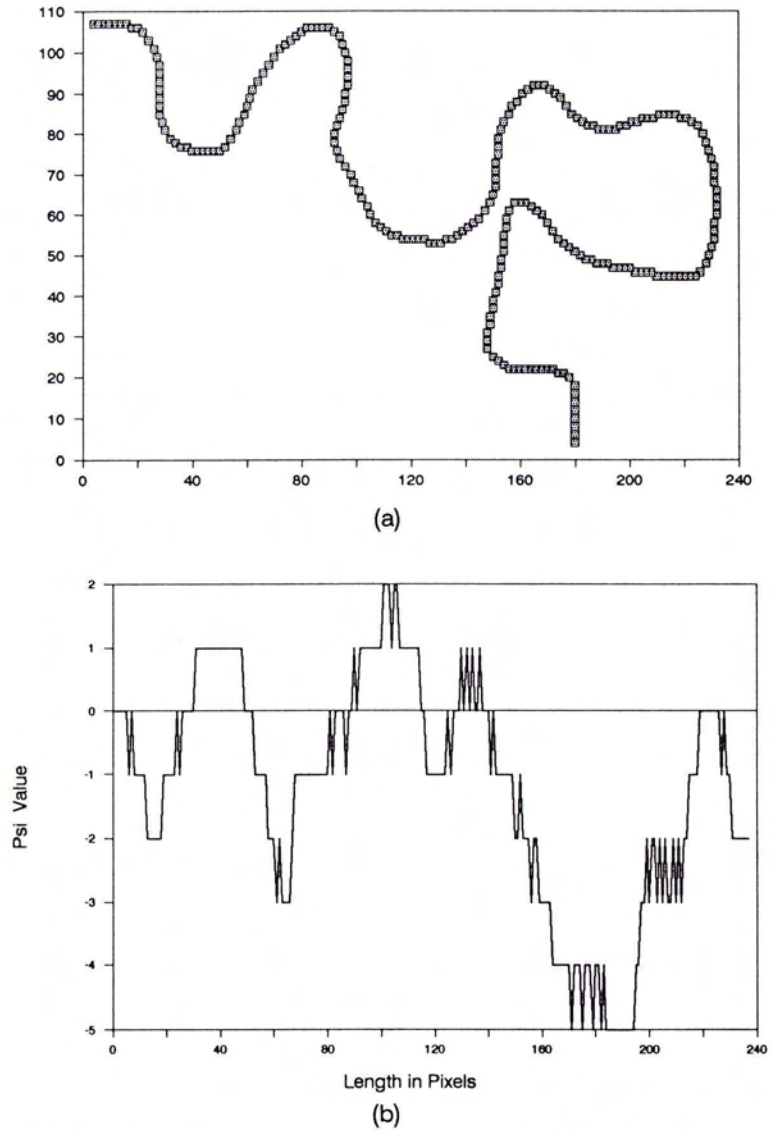

FIG. 6. A $\psi - s$ representation of an edge from the left image. (a) The original edge. (b) The $\psi - s$ representation.

and 7b is to compare some straight line segments following the same pattern. A smoothing process for eliminating the spikes and the staircase types of noise makes the comparison even simpler.

One should realize that for extended and complicated shapesp the $\psi - s$ method offers a simple and robust matching procedure. However, for a simple edge such as a line which is nearly straight, this method will fail. The reason for this is because straight lines, regardless of the their orientation, have the same $\psi - s$ representation. Thus, a north-south and an east-west edge can be mistakenly matched as both are represented by a horizontal straight line.

CONCLUSIONS

Feature-based stereo matching solutions are gaining ground in the photogrammetric community as the appropriate solution for the stereo matching problem. A stereo matching technique based on edge features rather than on point features was presented. This choice is due to the inherited characteristics of an edge which conveys rich information that distinguishes one edge from another.

The actual primitives of the matching process were described, as well as the similarity and consistency voting strategy, to accomplish the match. The spatial accuracies of the matches have not been discussed here as the prime objective of this paper was to test the feasibility of using edges to match an areal stereo pair without human operator intervention. The larger context of edge-based matching methods, in a comprehensive matching solution, will be discussed in Greenfeld (1989).

We presented two methods for describing edges. One is based on polygonal approximation of the edge while the other is based on a $\psi - s$ representation. The weaknesses and the strengths of each method were discussed and demonstrated on various images and different edges. It seems that these two methods complement each other into a matching solution which does not require *a priori* imposed constraints such as epipolar geometry.

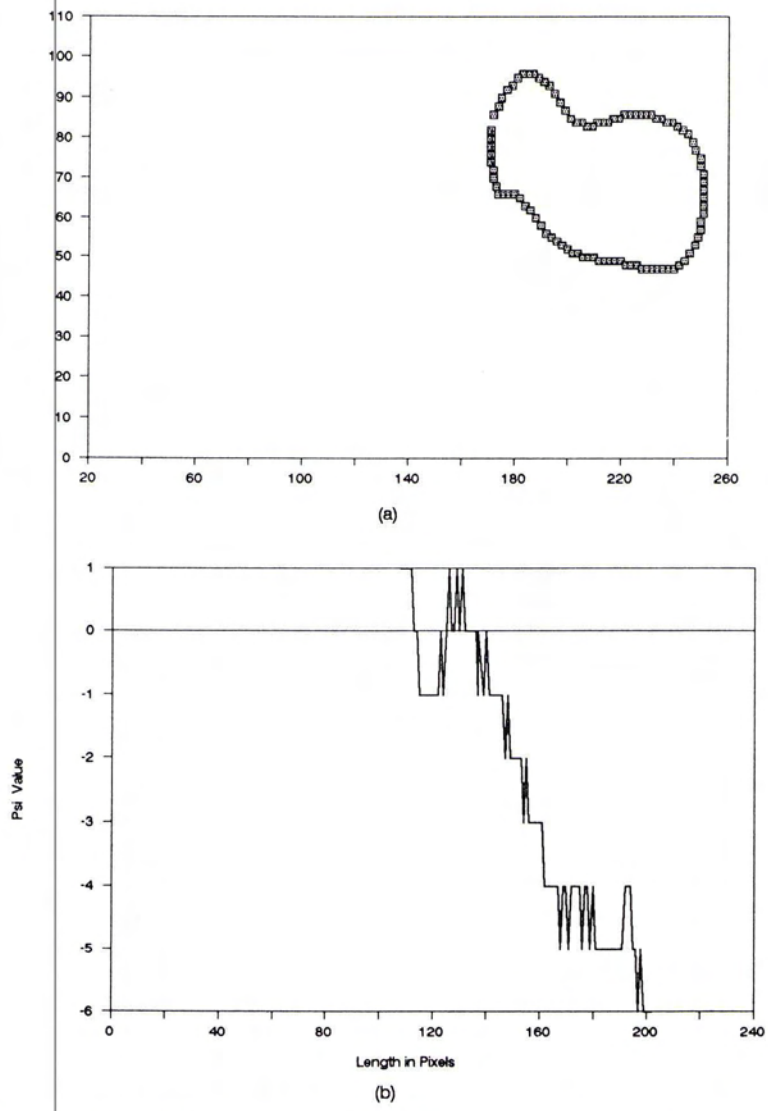

FIG. 7. A $\psi - s$ representation of an edge from the right image. (a) The original edge. (b) The $\psi - s$ representation.

REFERENCES

Attneave F., 1954. Some Information Aspects of Visual Perception, *Psychological Review*, 61, 183–193.

Ballard, D. H., and C. M. Brown, 1982. *Computer Vision*, Prantice Hall Inc.

Boyer, K. L., and A. V. Kak, 1986. *Symbolic Stereo from Structural Descriptions*, Report TR-EE-86-12 Purdue University, West Lafayette Indiana.

Freeman H., 1974. Computer Processing of Line Drawing, *Computer Surveys* 6, 1, 57–98.

Greenfeld, J. S., 1987. *A Stereo Vision Approach to Automatic Stereo Matching in Photogrammetry*, Report No. 381, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio (Also, PhD dissertation, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio).

———, 1988. Stereo Matching Without an Operator's Intervention, *Proceedings of 1988 ACSM-ASPRS Annual Convention*, Vol. 3, 13–18 March, St. Louis, Missouri, pp. 56–65.

———, 1989. An Operator Based Matching System, *Photogrammetric Engineering and Remote Sensing* (Submitted).

Grimson, W. E. L., 1985. Computational Experiments with a Feature Based Stereo Algorithm, *IEEE PAMI* 7, No. 1, 17–34.

Marr, D., 1982. *Vision*, W. H. Freeman and Company, San Francisco, California.

Marr, D., and E. C. Hildreth, 1980. A Theory of Edge Detection, *Phil. Trans. Royal Society of London*, B 207, 187–217.

Medioni, G., and R. Nevatia, 1985. Segment-Based Stereo Matching, *Computer Vision Graphics, and Image Processing*, 31, 2–18.

O'Neill, M. P., and D. M. Mark, 1987. The Psi-s Plot: A Useful Representation for Digital Cartographic Lines, *Proceedings of Auto-Carto 8*, Baltimore, Maryland.

Paderes, P. C., Jr., E. M. Mikhail, and W. Forstner, 1984. Rectification of Single and Multiple Frames of Satellite Scanned Imagery Using Points and Edges as Control, *NASA Symp. on Math. Pattern Recognition and Image Analysis*, June.

Schenk, A. F., and O. Hofmann, 1986. Stereo Matching Using Line Segments of Zero Crossings, *Proceedings, Inter. Society of Photogrammetry*, Comm. III Symp., Rovaniemi, Finland.

Teh, C. H., and R. T. Chin, 1988. A Scale Independent Dominant Point Detection Algorithm, *Proceedings of CVPR Conference*, Ann Arbor, Michigan, pp. 229–234.

Zoraster, S., D. Davis, and M. Hughus, 1984. *Manual and Automated Line Generalization and Feature Displacement*, ETL Report No. 101-43-16.

(Received 27 April 1989; accepted 20 July 1989; revised 8 August 1989)