

Disparity Analysis of Images

STEPHEN T. BARNARD, MEMBER, IEEE, AND WILLIAM B. THOMPSON, MEMBER, IEEE

Abstract—An algorithm for matching images of real world scenes is presented. The matching is a specification of the geometrical disparity between the images and may be used to partially reconstruct the three-dimensional structure of the scene. Sets of candidate matching points are selected independently in each image. These points are the locations of small, distinct features which are likely to be detectable in both images. An initial network of possible matches between the two sets of candidates is constructed. Each possible match specifies a possible disparity of a candidate point in a selected reference image. An initial estimate of the probability of each possible disparity is made, based on the similarity of subimages surrounding the points. These estimates are iteratively improved by a relaxation labeling technique making use of the local continuity property of disparity that is a consequence of the continuity of real world surfaces. The algorithm is effective for binocular parallax, motion parallax, and object motion. It quickly converges to good estimates of disparity, which reflect the spatial organization of the scene.

Index Terms—Disparity, matching, motion, relaxation labeling, scene analysis, stereo.

I. INTRODUCTION

DIFFERENCES in images of real world scenes may be induced by the relative motion of the camera and the scene, by the relative displacement of two cameras, or by the motion of objects in the scene. The differences are important because they encode information that often allows a partial reconstruction of the three-dimensional structure of the scene from two-dimensional projections. When such differences occur between two images we say that there is a *disparity* between the two images, which we represent as a vector field mapping one image into the other. The determination of disparity has been called the *correspondence* problem [1], [2]. A contingent problem is the *interpretation* of disparity into meaningful statements about the scene, such as specifications of depth, velocity, and shape.

There is much evidence that disparity is important in human vision. Gibson discussed the nature of visual perception in a dynamic environment [3]. He argued that the visual stimulus is inherently dynamic and that the patterns of change in the stimulus are important sources for the perception of the spatial environment. Gibson described the patterns of optical flow that occur when the observer moves. He argued that binocular disparity and retinal motion are highly informative stimulus variables for spatial perception, and that along with other important visual phenomena, such as texture gradient and linear perspective, they interact with the kinesthetic “body

Manuscript received March 5, 1979; revised September 14, 1979. This work was supported in part by the National Science Foundation under Grant MCS-78-20780.

S. T. Barnard was with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455. He is now with SRI International, Menlo Park, CA 94025.

W. B. Thompson is with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455.

senses” in the perception of a stable, upright, three-dimensional world.

Julesz’s experiments with random dot stereograms [4] support the contention that the human visual system is able to process differences between images. A stereogram is a pair of images that are recorded simultaneously by laterally separated sensors. A random dot stereogram is artificially constructed by shifting parts of an image of random dots to the left or right to form a second image. Both images consist of uncorrelated random dots, but there is a disparity relationship between the two which specifies the apparent relative depth of the shifted parts. People with normal stereo vision can easily achieve a binocular fusion of the two images and perceive the apparent relative depth of the various “surfaces.” Even though each separate image of a random dot stereogram contains no depth information, people are able to perceive depth in the pair of images by measuring the disparity between the shifted areas.

Disparity analysis may be broadly defined as the determination of the geometric differences between two or more images of the same or similar scenes. The differences may be the result of binocular parallax, motion parallax, object motion, or some combination of these modes. The goal of the analysis is to assign disparities, which are represented as two-dimensional vectors in the image plane, to a collection of points in one of the images. Disparity analysis is useful for image understanding in several ways. There is information in a disparate pair of images that is difficult to find or even absent in any single image. This point is convincingly made by Julesz’s random dot stereogram experiments. Disparity provides a way, independent of high-level knowledge, of determining the spatial relationships between points and surfaces in a scene. The objects in the scene may be completely unfamiliar, but their observed disparities will conform to precise rules that depend only on the location and velocity of objects in three-dimensional space. Disparity is therefore a very general property of images which may be used in a variety of situations. The measurement of depth and velocity will certainly be useful in many applications, but there is a more fundamental requirement in image understanding. A system for understanding dynamic scenes can use observed disparity to establish conceptual relationships between images that are invariant over several observations. Visual invariants can be used to predict future observations, to eliminate noise in any one observation, and in general to link several observations into one perceptually coherent whole.

II. MATCHING

Matching is a natural way to approach disparity analysis. Assigning disparity classifications to points in a sequence of images is equivalent to finding a matching between sets of

points from each image. Let $S_1 = \langle s_1^x, s_1^y \rangle$ and $S_2 = \langle s_2^x, s_2^y \rangle$ be points in images 1 and 2, respectively. These two points should be matched if and only if they are image plane projections of the same real world surface point. Matching S_1 with S_2 is the same as assigning to S_1 a disparity with respect to image 2 of $D_1^2 = \langle s_1^x - s_2^x, s_1^y - s_2^y \rangle$, where $D_j^k(S)$ is a vector function defined on points of image j , which specifies the disparity of point S in image j with respect to image k .

A matching approach to disparity analysis must solve two problems. First, how are points selected for matching? It is clear that not all points can be matched with equal confidence because some are located in regions which lack identifying detail. Some points may not be matched at all because they may be visible in one image but not in another. To avoid ambiguous matches, it is advantageous to attempt to match only points which are easily distinguished from their neighbors. It is important to select only those points which are projections of distinct, precisely positioned local features on real world surfaces, such as spots, corners, and other small local forms. Interest operators sensitive to local variance, edges, and other properties of subimages can be used to choose potentially matchable points [5]-[9]. An alternative to this strategy is to partially segment each static image independently and then use properties of the segments, such as shape and color, as similarity variables [10], [11]. This is an attempt to match entire surfaces, not points.

The second problem in matching is to determine the basis for deciding which matches are correct. The matched points should have similar properties, of course, because they are both projections of the same surface point, but in many cases there will be ambiguity. Many studies have used cross correlation or mean-square difference as a measure of similarity. Typically, a small region in image 1 surrounding an interesting point is used as a template, and a search is made for the region of maximum similarity in image 2. Two problems with this approach are that the correct match may not be the one of maximum similarity due to noise or distortion (there are ambiguous matches), and the cost of searching a two-dimensional image is high. One way to avoid ambiguity is to increase the sensitivity of the similarity measurement. Levine, O'Handley, and Yagi [7] use an adaptive correlation window, the size of which varies inversely with the variance of the region surrounding the point. Mori, Kidode, and Asada [12] use a Gaussian-weighted correlation window to minimize the errors due to distortion of the extremities. They also vary the window size with ambiguity of the match and use a prediction/correction algorithm, modifying one image to fit the other according to a predicted matching and iteratively using the error of fit to improve the prediction. Stochastic matched filtering can reduce ambiguity by improving similarity detection in the presence of noise [13].

Several strategies have been used to limit search in a cross correlation approach. Studies of stereopsis use a fixed camera model to constrain the search to one dimension [5], [7], [8], [12]. Nevatia uses a series of progressive views to constrain disparity to small values [8]. This also reduces the chance of ambiguous matches and increases the sensitivity of the similarity measurement by minimizing distortion. Another strategy

is to use a coarse search to approximately locate the matching points, followed by a fine search to more accurately locate them [7], [14], [15]. Sequential similarity detection can be used for more efficient matching [16].

Many studies have used heuristics based on real world models to limit search and to resolve or avoid ambiguity. Julesz [4] and Gibson [3] observed that disparity varies continuously across unoccluded surfaces and discontinuously only at occluding edges. Marr and Poggio [17] used this property, which they call the adjacency constraint, in an iterative cooperative matching algorithm which fuses random dot stereograms. Levine *et al.* [7] use it to limit the range of search of proximate points, and Mori *et al.* [12] use it to avoid ambiguity by matching "well-contrasting" regions with high confidence and favoring these disparities for nearby points.

In the studies discussed above we can identify three properties of image pairs which can strongly influence disparity classification. The first, which we call discreteness, is a property of individual points. Discreteness is a measurement of the individual distinctness of a point, and is important for selecting good candidates for matching. The interest operators described above are good examples of discreteness measures. The second property, similarity, is a measurement of how closely two points resemble one another. Such measures are usually simple functions of small areas surrounding the points. The third property, consistency, is a measurement of how well a particular match (that is, a particular disparity classification) conforms to nearby matches. The three-dimensional spatial continuity of real world surfaces constrains the two-dimensional spatial distribution of disparity in the image plane. Disparity is discontinuous only at occluding edges. The continuity of disparity over most of the image can be used to avoid false matches based on similarity alone, by suppressing matches in the absence of supporting local evidence.

III. A LOCALLY PARALLEL MODEL FOR MATCHING

This section describes a computational model for analyzing disparity in a variety of real world situations. Object motion, stereo, and motion parallax modes of disparity are treated uniformly. Because the model has the locally parallel, globally sequential structure of a relaxation labeling algorithm, the notation in this section follows Rosenfeld, Hummel, and Zucker [18]. The general approach is similar to matching models proposed by Julesz [4] and Marr and Poggio [17] for the fusion of random dot stereograms and to a matching model proposed by Ullman [2] for retinal motion.

The theory behind the locally parallel model is that the discreteness, similarity, and consistency properties can interact to enhance the overall performance and rigor of a matching algorithm. The use of multiple properties reduces the chance of error by minimizing the dependence on any one property. The consistency property allows the most obvious classifications to improve the analysis of the more difficult ones. The discreteness property is used to minimize expensive searching. Sets of candidate matching points are selected from each image, and searching is done between these two relatively sparse sets instead of between the two gray-valued images. The emphasis throughout the formulation of the model has

been on simplicity. We will demonstrate the effectiveness of this matching approach in even a minimal system.

The first step is to find the points in the two images which will be candidates for matching. Matchable points should locate small discrete local features such as spots and corners which are likely to be detectable in both images. They should be the centers of highly variable areas, and furthermore the variance should be high in all directions. (If the variance is high in general but low along one direction, as would be the case with a straight line, for example, the point will not be easily distinguishable from its neighbors along that direction). A very simple interest operator described by Moravec [19] is effective in selecting these matchable points. The sums of the squares of the differences of pixels in four directions (horizontal, vertical, and the two diagonals) are computed over a small area (5×5 areas are used in all examples presented here). The initial value of the interest operator for the center point is the minimum of these variances. A point will have a high initial value only if there is a high variance in all four directions. The final values of the interest operator are obtained by suppressing (i.e., setting to zero) the initial values of all but the local maxima. Any point with a nonzero final interest value must be located in an area with a high variance in all four directions, and the initial value of the point must be greater than that of any of the point's neighbors.

The interest operator is applied independently to each image. Points with locally maximal but very small interest values are rejected by thresholding. The selection of the threshold is not critical. In the examples shown in this paper it was set to give a reasonable number of points, expressed as a percentage of the total number of pixels in each image.

After the two sets of candidate points are found the next step is to construct a set of possible matches. Ideally, we would like to match each candidate point from image 1 with exactly one candidate from image 2 such that the two points are image plane projections of the same real world point. However, we can realistically expect to find valid matches for only some of the points of image 1 because the interest operator does not perform perfectly, and because some points may be occluded, shadowed, or not visible in image 2 for some reason. An initial set of possible matches is constructed by pairing each candidate point from image 1 with every candidate from image 2 within some maximum distance of the (x, y) location of the point in image 1. This distance r is the maximum detectable disparity in the x or y direction. The set of possible matches is organized as a collection of "nodes" $\{a_i\}$, one node for each candidate point from image 1. Associated with each node a_i is a tuple (x_i, y_i) which is the location of the point in image 1, and a set of labels L_i which represents possible disparities that may be assigned to the point. Each label in L_i is either a disparity vector (l_x, l_y) , where l_x and l_y are integers in $[-r, r]$, or it is a distinguished label l^* denoting "undefined disparity." A node a_i has undefined disparity if point (x_i, y_i) in image 1 does not correspond to any candidate point in image 2. Every label set L_i must initially contain the element l^* . The point (x_i, y_i) in image 1 is tentatively matched to a point at (x', y') in image 2 by entering a label $l = (x_i - x', y_i - y')$ into L_i . Note that not every vector with integral coordinates

in the square of side $2r + 1$ need be represented in the label set of all possible matches for a node, but that the undefined disparity is always represented.

For every node a_i we want to associate with every label $l = (l_x, l_y)$ in L_i a number $p_i(l)$ which we can interpret as an estimate of the probability that point (x_i, y_i) in image 1 has disparity l . This requires that $p_i(l)$ be in $[0, 1]$ and $\sum_l p_i(l) = 1$. These probability estimates will be successively improved by applying the consistency property. If relatively many nearby points have a high probability of having disparity l , then $p_i(l)$ will increase. Otherwise, $p_i(l)$ will decrease.

The initial probabilities $p_i^0(l)$ are based on the sum of the squares of the differences between a small window from image 1 centered on (x_i, y_i) and a window from image 2 centered on $(x_i + l_x, y_i + l_y)$. (The window sizes in all examples presented here are 5×5 .) Let this sum be $s_i(l)$ for $l \neq l^*$. When $s_i(l)$ is small $p_i^0(l)$ should be large, and *vice versa*.

Let

$$w_i(l) = \frac{1}{1 + c * s_i(l)}, \quad l \neq l^* \quad (1)$$

be the "weight" associated with the label l of node a_i for some positive constant c . ($c = 10$ for all examples presented in this paper.) A disparity label which associates highly similar pairs of regions will have a large weight value. Note that $w_i(l)$ is in the interval $[0, 1]$ and is inversely related to $s_i(l)$, but in general the sum of the weights is not 1 and $w_i(l^*)$ is undefined, so we cannot use these weights as our probability estimates. Nevertheless, $w_i(l)$ has the proper qualitative behavior for probability estimates.

We first estimate the initial probability of the undefined disparity, $p_i^0(l^*)$, by observing that in many cases the label of maximum weight is the correct one. If no label has a high weight then there is probably no valid match. It may often be the case that the correct label is not the one of maximum weight, but if there is a significant correlation between maximum weight and ground truth the relationship will prove an adequate initial estimate of the probability that each node is matchable.

Using this relationship, let

$$p_i^0(l^*) = 1 - \max_{l \neq l^*} (w_i(l)) \quad (2)$$

be the initial estimate of the probability that the point (x_i, y_i) in image 1 corresponds to no point in image 2.

We can apply Bayes' rule to obtain an initial estimate of the probability that a_i should be label l for labels other than l^* . Let

$$p_i^0(l) = p_i(l|i) * (1 - p_i^0(l^*)), \quad l \neq l^* \quad (3)$$

where $p_i(l|i)$ is the conditional probability that a_i has label l given that a_i is matchable, and $(1 - p_i^0(l^*))$ is the probability that a_i is matchable. We can estimate $p_i(l|i)$ with

$$p_i(l|i) = \frac{w_i(l)}{\sum_{l' \neq l^*} w_i(l')} \quad (4)$$

Equations (2), (3), and (4) can be used to calculate the initial probabilities for every label of every node.

The initial probabilities, which depend only on the similarity of neighborhoods of candidate matching points, can be improved by using the consistency property. We want a rule for updating $p_i^k(l)$ which has the following property. The new probability $p_i^{k+1}(l)$ should tend to increase when nodes with highly probable labels consistent with l are found near node a_i . Labels are considered consistent if they represent nearly the same disparities,

$$\|l - l'\| \leq \Theta$$

for an appropriate threshold. For the examples used in this paper the condition for consistency is

$$\|l - l'\| = \max(|l_x - l'_x|, |l_y - l'_y|) \leq 1.$$

A node a_j may be considered near a_i if

$$\max(|x_i - x_j|, |y_i - y_j|) \leq R.$$

That is, the points corresponding to a_i and a_j in image 1 are no more than R rows or columns apart. ($R = 15$ in all examples presented here.)

The degree to which the label l' of a_j reinforces $p_i(l)$ should be related to the estimated likelihood that l' is correct. To compute the new probabilities $p_i^{k+1}(l)$ for all l in L_i we examine each node in an area surrounding a_i , but not including a_i . Let

$$q_i^k(l) = \sum_{\substack{j \ni a_i \\ \text{near } a_i \\ j \neq i}} \left[\sum_{\substack{l' \ni a_j \\ \|l - l'\| \leq \Theta}} p_j^k(l') \right], \quad l \neq l^*. \quad (5)$$

In all cases $q_i^k(l) \geq 0$. This quantity is zero if and only if no nodes surrounding a_i have possible matches with disparity labels similar to l . It will be large when there are several nodes with highly probable matches surrounding a_i which have disparity labels similar to l . In general, $q_i(l)$ varies according to the consistency of l with the current estimate of the disparities in the local neighborhood of a_i . All nodes surrounding a_i are treated uniformly.

We can use the following rule to update the label probabilities of node a_i using q_i . Let

$$\hat{p}_i^{k+1}(l) = p_i^k(l) * (A + B * q_i^k(l)), \quad l \neq l^* \quad (6)$$

and

$$\hat{p}_i^{k+1}(l^*) = p_i^k(l^*). \quad (7)$$

We must normalize the \hat{p} 's to obtain the new probabilities,

$$p_i^{k+1}(l) = \frac{\hat{p}_i^{k+1}(l)}{\sum_{l' \text{ in } L_i} \hat{p}_i^{k+1}(l')}. \quad (8)$$

Parameters A and B are positive constants which influence the convergence characteristics of the model ($A = 0.3$ and $B = 3$ in all examples presented here). The role of A is to delay the total suppression of unlikely labels. Even if $q_i(l)$ is zero, the positive A ensures that the new probability does not become zero. This is desirable because information may propagate to a_i from other nodes which will eventually cause

$p_i(l)$ to increase. The role of B is to determine the rate of convergence. The larger B is relative to A the faster will be the convergence of the disparity assignments. A and B may be interpreted as damping and gain parameters. There is an effective constant correlation for label pairs which are similar, and a negative correlation for all others. (For a more general discussion of updating rules in relaxation labeling algorithms see [18].)

The probability of the label l^* (undefined disparity) is affected only by the normalization step (8). If the net contribution to $\hat{p}_i(l)$, $l \neq l^*$, is such that

$$\sum_{l \neq l^*} \hat{p}_i^{k+1}(l) < \sum_{l \neq l^*} p_i^k(l), \quad (9)$$

then the probability that a_i has disparity l^* increases, but if this is not the case the probability decreases or perhaps remains the same.

The complete procedure to estimate the most likely disparities for each potentially matchable point in image 1 can be summarized as follows. First, a network of nodes corresponding to possible matches is constructed. Each possible match is assigned an initial likelihood using (1), (2), (3), and (4). These likelihood estimates are iteratively refined using (5), (6), (7), and (8). We have observed that after a few iterations most of the possible matches have very low probability. To increase the efficiency of the algorithm it is effective to purge these from the set of possible matches after each iteration. (In the examples presented here we purge a tentative match when its probability falls below 0.01.) If a node a_i loses all its possible matches, then (x_i, y_i) is classified as "unmatchable." That is, $p_i(l^*) = 1$. This procedure may be repeated until the network reaches a steady state, but in practice we arbitrarily stop it at ten iterations. Those nodes having a disparity with an estimated likelihood of 0.7 or greater are considered to be matched. Some nodes may remain ambiguous, with several potential matches retaining nonzero probabilities.

IV. RESULTS

An interesting and important property of this matching algorithm is that it works for any mode of disparity and does not require precise information about camera orientation, position, and distortion. Disparity in multiple images of the same scene may be due to translation or rotation of the sensor or due to motion of objects in the scene. At least three specific cases are commonly found in real world situations. The disparity is strictly temporal when a single, stationary camera records a sequence of images. Temporal disparity specifies the motion of objects in the image plane. The disparity is stereoptic when two horizontally displaced cameras simultaneously record the same scene. Often the focal axes are also rotated about a fixation point which has an arbitrary disparity of $(0, 0)$. (If the focal axes are parallel the fixation point is at infinity.) Another salient mode is forward sensor motion. While the motion is commonly along the focal axis, camera rotations and/or off-axis motion may also occur. The computational model described in the previous section was tested on examples of each of these cases. A precise photogrammetric model could translate these results into quantitative

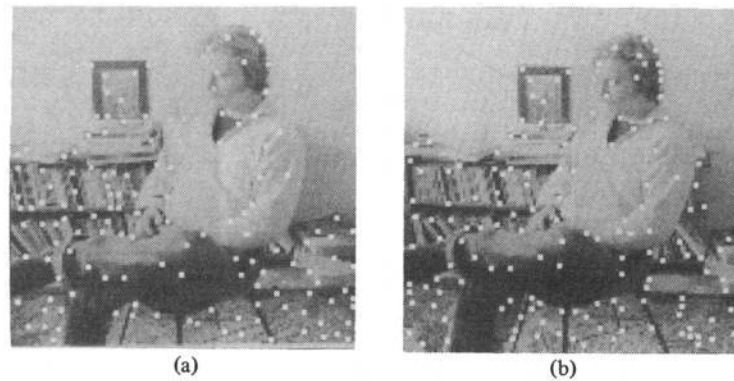


Fig. 1. Stereogram.

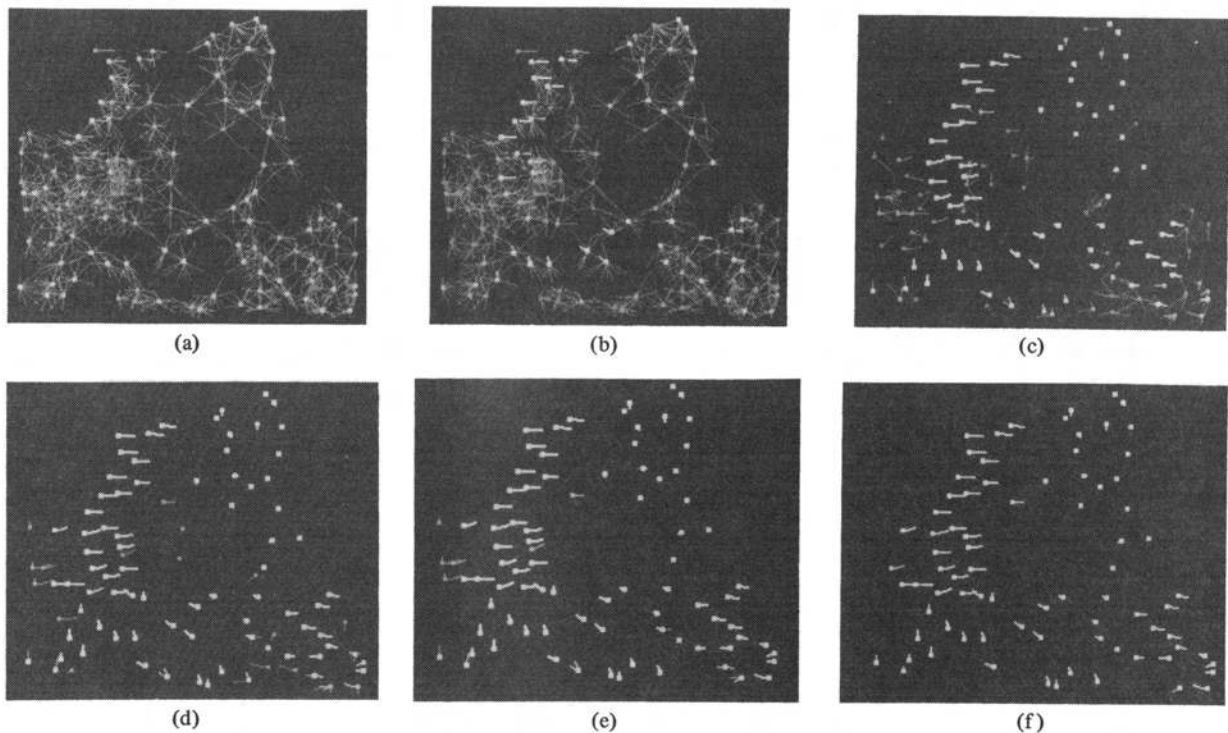


Fig. 2. (a) Initial probability assignments for Fig. 1. (b)–(f) Iterations 2, 4, 6, 8, and 10.

measurements of velocity and depth, but we shall only discuss their qualitative significance.

The first case is a stereogram (Fig. 1). The candidate matching points selected by the interest operator are superimposed on the images. Fig. 2 shows the performance of the model over ten iterations, displaying every second iteration in addition to the initial state. Each point represents a node and each line a possible match. Each line is in the direction of the corresponding disparity vector, and the length is proportional to the length of the vector. A match of small disparity appears only as a point because the line is too small to be visible. The brightness of each point is proportional to the probability that the point is matchable, $(1 - p_i^k(I^*))$, and the brightness of each line is proportional to the probability of the disparity assignment, $p_i^k(I)$. Initially, there are many possible matches with relatively low probability, but by iteration 4 [Fig. 2(c)] almost

all of them have been discarded. On iteration 10 [Fig. 2(f)] only those classifications with probability greater than 0.7 are shown. The cameras are fixated on the person's head, which has a disparity of $\langle 0, 0 \rangle$. More distant points have disparities with positive x components and nearer points have disparities with negative x components. Observe that from the person's head to his left knee and then to his left foot disparity varies smoothly, following the smooth transition from far to near and back to far [Fig. 2(f)]. Between the person and the background, however, there is a step change in depth at the occluding edge, and here the assigned disparities exhibit an abrupt transition from near to far. The nonzero y components of some of the closer points is due to motion of the subject between exposures (only one camera was used to record this stereogram).

The next example illustrates temporal disparity. Fig. 3

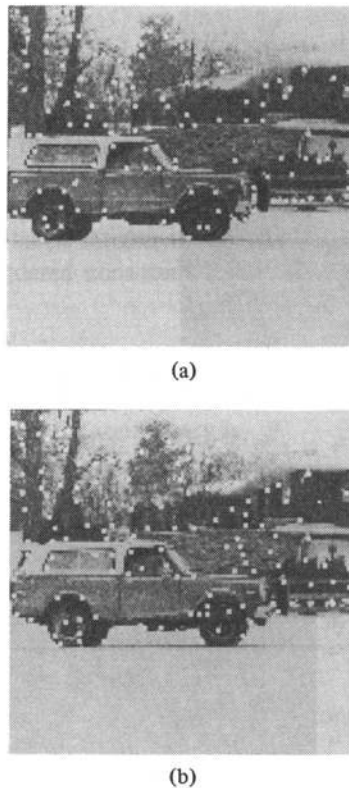


Fig. 3. Object motion.

shows a sequence in which the truck moves approximately seven pixels to the right while the camera, and hence the background, remains stationary. The truck is easily distinguishable as a cluster of points with disparity of about $\langle 7, 0 \rangle$ and the background as a cluster of points with disparity of about $\langle 0, 0 \rangle$ [Fig. 4(f)]. An error occurs at a point slightly above the roof of the truck. The point is actually part of the background but its estimated disparity is characteristic of the truck surface.

The third example simulates a view from an airplane flying over a city (Fig. 5). Actually, the subject was a scale model of downtown Minneapolis which was photographed with an ordinary camera and tripod. As the camera "flies" over the "city" it rotates downward in the frontal plane to fixate on a point near the center of the image (Fig. 6). The final disparity vectors diverge from the fixation point which has disparity $\langle 0, 0 \rangle$ [Fig. 7(f)]. Two distinctive features in Fig. 7(f) are the cluster of points on the large building in the near foreground, which has a large disparity because it is very close, and the cluster of points in the far background (upper left corner), which has mainly vertical disparity because of the rotation of the camera. The algorithm is susceptible to an aliasing effect when it encounters a dense cluster of similar points, such as would occur in a high frequency periodic subimage. An example of this may be seen at the middle of the right border of Fig. 7(f), where a number of points have been misclassified. One way to avoid this problem would be to enlarge the area in which interest values must be locally maximal to produce feature points, thereby constraining the density of feature points to a relatively small value.

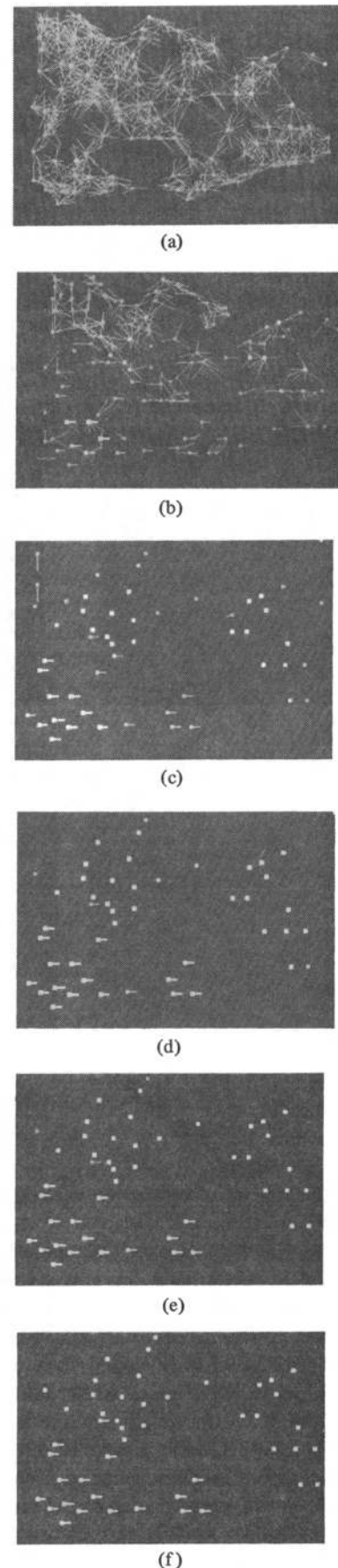


Fig. 4. (a) Initial probability assignments for Fig. 3. (b)-(f) Iterations 2, 4, 6, 8, and 10.

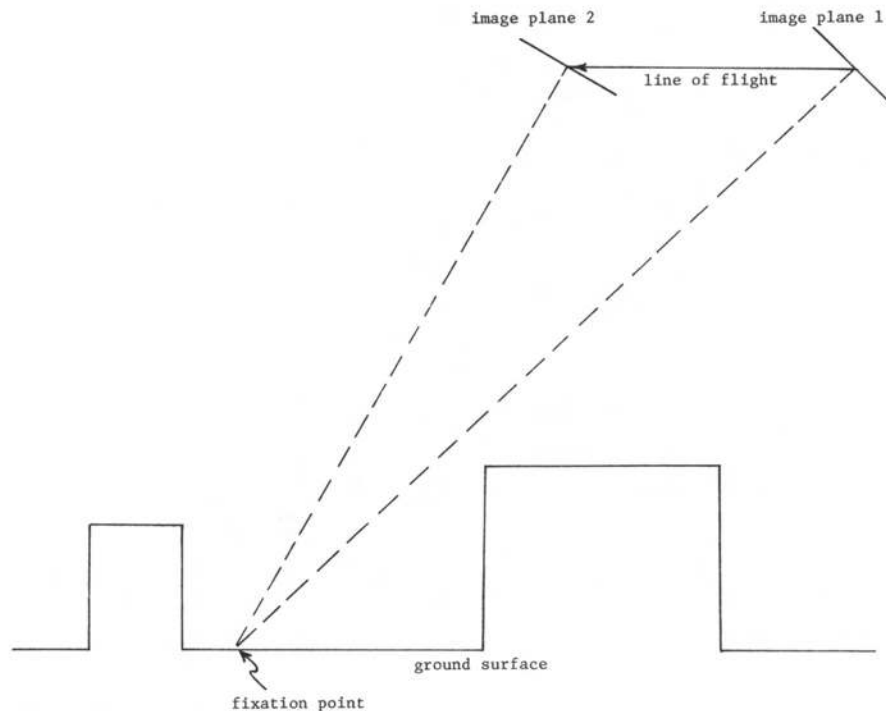
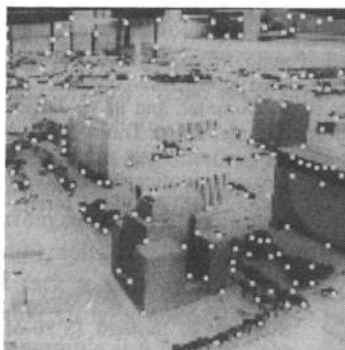
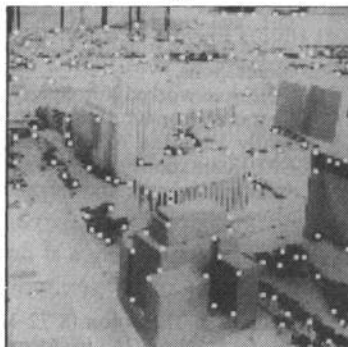


Fig. 5. Schematic of third example.



(a)



(b)

Fig. 6. Motion parallax.

V. SUMMARY AND CONCLUSIONS

Disparity relationships between images convey important information for spatial perception. The conventional cross correlation approach to matching for disparity analysis is prone to errors caused by distortion in the imaging process and the unavoidable occurrence of ambiguous matches. Increasing the sophistication of similarity detection can reduce

the problem of ambiguity, but the necessary adaptive windowing or extensive preprocessing can also reduce the efficiency and reliability of the matching process. Geometric distortion may prevent accurate similarity measurement by cross correlation. Constraining the search with a precisely known camera geometry restricts the process to special cases.

Another way to attack the ambiguity problem is to use information other than similarity to resolve ambiguous matches. The continuity of real world surfaces, a very general property which is independent of camera geometry, constrains disparity to values which are consistent with the local matching context. The consistency property can be used to iteratively refine a probabilistic estimate of disparity classification which allows ambiguous matches. The initial estimate may be obtained from a simple measure of similarity. Expensive searching is minimized by applying a simple interest operator to both images to find two sets of discrete feature points, resulting in a greatly reduced search space.

In practice the procedure is effective in a wide variety of cases, converging quickly to an unambiguous disparity classification which accurately reflects the large-scale spatial characteristics of real world scenes. It is robust in that it is not very sensitive to noise and distortion and does not require careful registration of images. It does not require complicated, carefully tuned updating functions, nor very precise initial probability estimates. The success of the method comes in large part from its ability to integrate effectively different sources of information into a simple procedure.

ACKNOWLEDGMENT

We would like to thank R. Hummel, C. Lemche, H. Nagel, S. Sahni, and A. Yonas for useful suggestions and criticism. We would also like to thank W. Franta and the staff of the

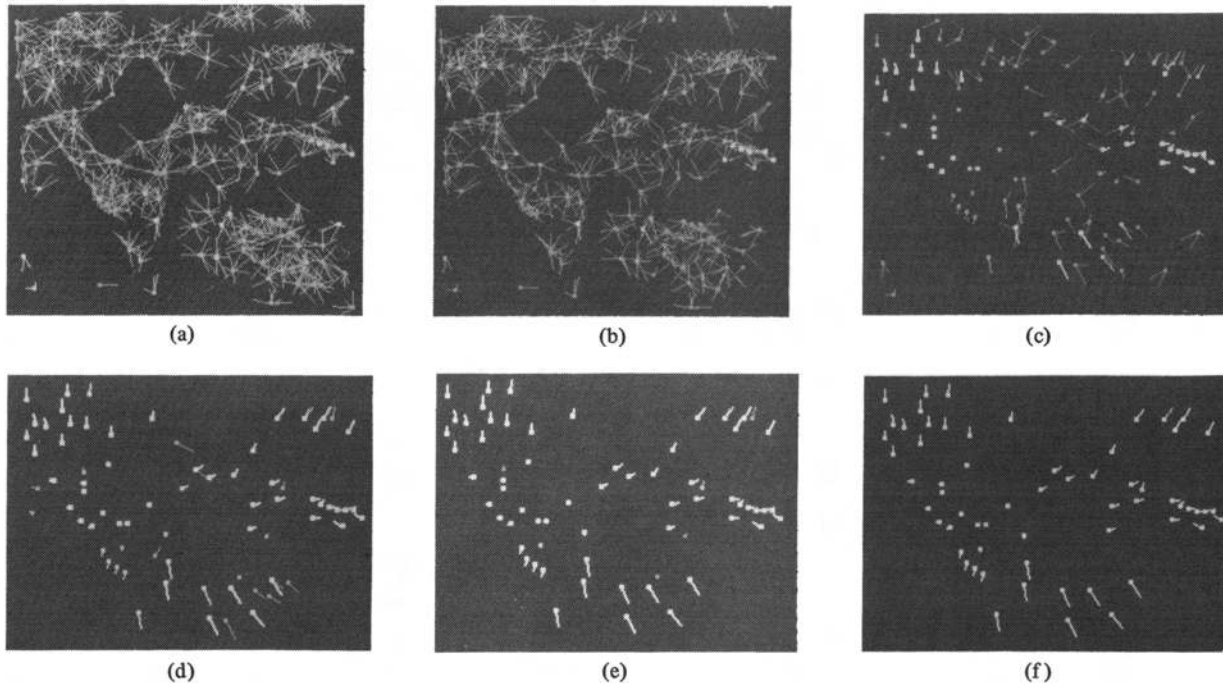
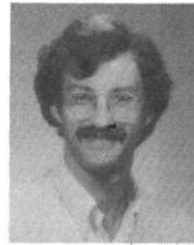


Fig. 7. (a) Initial probability assignments for Fig. 6. (b)-(f) Iterations 2, 4, 6, 8, and 10.

Special Interactive Computing Laboratory, University of Minnesota, for their helpful assistance and for the use of their excellent facilities.

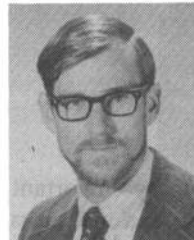
REFERENCES

- [1] R. O. Duda and P. E. Hart, *Pattern Recognition and Scene Analysis*. New York: Wiley, 1973.
- [2] S. Ullman, "The interpretation of visual motion," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, May 1977.
- [3] J. J. Gibson, *The Perception of the Visual World*. Cambridge, MA: Riverside, 1950.
- [4] B. Julesz, *Foundations of Cyclopean Perception*. Chicago, IL: Univ. Chicago Press, 1971.
- [5] M. J. Hannah, "Computer matching of areas in stereo images," Stanford A.I. Memo. 239, July 1974.
- [6] L. H. Quam, "Computer comparison of pictures," Stanford A.I. Memo, AIM-144, May 1971.
- [7] M. D. Levine, D. A. O'Handley, and G. M. Yagi, "Computer determination of depth maps," *Comput. Graphics Image Processing*, vol. 2, pp. 131-150, 1973.
- [8] R. Nevatia, "Depth measurement by motion stereo," *Comput. Graphics Image Processing*, vol. 5, pp. 203-214, 1976.
- [9] D. J. Kahl, A. Rosenfeld, and A. Danker, "Some experiments in point pattern matching," Univ. Maryland, Tech. Rep. TR-690, Sept. 1978.
- [10] K. Price and D. R. Reddy, "Change detection and analysis in multispectral images," in *Proc. Int. Joint Conf. Artificial Intell.*, pp. 619-625, 1977.
- [11] H. Nagel, "Formation of an object concept by analysis of systematic time variations in the optically perceptible environment," *Comput. Graphics Image Processing*, vol. 7, pp. 149-194, 1978.
- [12] K. Mori, M. Kidode, and H. Asada, "An iterative prediction and correction method for automatic stereocomparison," *Comput. Graphics Image Processing*, vol. 2, pp. 393-401, 1973.
- [13] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.
- [14] D. B. Gennery, "A stereo vision system for an autonomous vehicle," in *Proc. 5th Int. Joint Conf. Artificial Intell.*, Cambridge, MA, Aug. 1977, pp. 576-582.
- [15] S. L. Tanimoto, "A comparison of some image searching methods," in *Proc. IEEE Comput. Soc. Conf. Pattern Recognition Image Processing*, Chicago, IL, June 1978, pp. 280-286.
- [16] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast image registration," *IEEE Trans. Comput.*, vol. C-21, pp. 179-186, Feb. 1972.
- [17] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283-287, Oct. 15, 1976.
- [18] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, June 1976.
- [19] H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proc. 5th Int. Joint Conf. Artificial Intell.*, Cambridge, MA, Aug. 1977, p. 584.



Stephen T. Barnard (S'78-M'79) received the B.S. degree in mechanical engineering from Case Western Reserve University, Cleveland, OH, and the M.S. and Ph.D. degrees in computer science from the University of Minnesota, Minneapolis.

He has worked for Sperry Univac, the 3M Corporation, and Honeywell, Inc. He is currently with the Artificial Intelligence Center, SRI International, Menlo Park, CA, where he is working on image understanding and automated visual inspection.



William B. Thompson (S'72-M'75) was born in Santa Monica, CA, in August 1948. He received the Sc.B. degree in physics from Brown University, Providence, RI in 1970, and the M.S. and Ph.D. degrees in computer science from the University of Southern California, Los Angeles, in 1972 and 1975, respectively.

From 1972 to 1975 he was a member of the Image Processing Institute at the University of Southern California. Since September 1975, he has been an Assistant Professor in the Department of Computer Science, University of Minnesota, Minneapolis. His interests are in artificial intelligence and scene analysis.

Dr. Thompson is a member of Sigma Xi, Eta Kappa Nu, and the Association for Computing Machinery.