

# 3-D Surface Description from Binocular Stereo

Steven D. Cochran, *Member, IEEE*, and Gérard Medioni, *Senior Member, IEEE*

**Abstract**—We present a stereo vision system that attempts to achieve robustness with respect to scene characteristics, from textured outdoor scenes to environments composed of highly regular man-made objects. Unlike most stereo approaches, it integrates “area-based” and “feature-based” primitives. This allows it to take advantage of the unique attributes of each of these techniques. The area-based processing provides a dense disparity map, and the feature-based processing provides an accurate location of discontinuities. We are able to generate a disparity map that is sufficiently accurate to allow us to detect depth and surface orientation discontinuities, provided that the resolution is fine enough. We use an area-based cross correlation along with an ordering constraint and a weak surface smoothness assumption to produce an initial disparity map. Unlike other approaches, however, a match is accepted only if both views agree on a correlation peak and if this peak is strong enough. This disparity map is only a blurred version of the true one because of the smoothing introduced by the cross correlation. The problem is most acute at depth discontinuities but can be reduced by introducing the edge information: The disparity map is smoothed (subject to the constraint that the disparity at edgels is fixed) and the unsupported points removed. It is important to note that this method gives an active role to edgels parallel to the epipolar lines, whereas they are discarded in most feature-based systems. We have obtained very good results on complex scenes in different domains and have been able to locate visible surfaces, penumbral and off-the-edge areas, and depth and orientation discontinuities in the images.

**Index Terms**—Area-based matching, binocular stereo, feature-based matching, surface segmentation.

## I. INTRODUCTION

**H**UMANS ARE able to surmise depth in 2-D “monocular” images and to perceive it through the stereoscopic fusion of a pair of images. The process used by the human visual system to achieve this, however, is not well understood. Much research has been devoted to the automatic abstraction of information about objects in images in order to produce autonomous systems that are able to “perceive” and operate on their environments and, in some cases, to gain insight about the operation of the human visual system. In 1982, Barnard and Fischler [1] defined six steps necessary for stereo analysis: image acquisition, camera modeling, feature acquisition, image matching, depth reconstruction, and interpolation.

Manuscript received May 12, 1991; revised November 11, 1991. This work was supported by the Defense Advanced Research Projects Agency (DARPA) under contract F33615-87-C-1436, monitored by the Air Force Wright Aeronautical Laboratories, and by a TRW Doctoral Fellowship. Recommended for acceptance by Associate Editor W. E. L. Grimson.

S. D. Cochran is with the Digital Mapping Laboratory, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213.

G. Médioni is with the Institute for Robotics and Intelligent Systems, Departments of Computer Science and Electrical Engineering, University of Southern California, Los Angeles, CA 90089-0273.

IEEE Log Number 9202386.

Of these steps, image matching is widely considered to be the most difficult to solve and is clearly dependent on the choice of feature primitives. Given two views of a scene, a correspondence must be established between those points that are visible in both scenes.

When the matched features are low level and dense, such as the intensity at each pixel, and the matching strategy is applied locally throughout the image at each pixel, we call the matching strategy “area-based.” When a few interesting (usually more abstract) features are first selected from the image and the matching strategy is applied to this subset, we use the term “feature-based.” Some systems (like ours) use a hybrid approach with multiple features that are both sparse and dense.

There are several problems that make the process of matching the selected features more difficult:

- **Photometric Variation:** The light reflected from the scene and recorded by the camera depends on the position of that camera relative to the scene, as well as noise and nonlinearities in the camera itself. Thus, when a camera is moved to a new position, or when two cameras view a scene from two viewpoints, the intensity at the corresponding points may be different.
- **Occlusion:** Occlusion is due to the occurrence of a depth discontinuity that causes an obstructed view of part of the scene (behind the occluding edge) that is “observed” by only one of the cameras. The border of the image can also act as a sort of occluding edge in this sense.
- **Repetitive Texture:** When the texture is repeated, such as the bricks in a brick wall, multiple possible correspondences exist. This problem is exacerbated when coupled with photometric variation and can lead to wrong matches that, due to these distortions, appear to be better matches within a local area.
- **Lack of Texture:** In real-world scenes, most objects and surfaces are textured. Although texture often does not give rise to useful, matchable abstract features, such as segments, it forms a basis for the real strength of an area-based match by providing a statistically matchable pattern. However, part of the scene may be without texture and, in that area of the scene, no area-based or feature-based match is possible.

In the remainder of this paper, we demonstrate the usefulness of combining information from area-based and feature-based approaches. Beginning with a registered image pyramid, we generate a dense depth map that is sufficiently accurate for the detection and localization of depth and surface orientation discontinuities. The approach that we detail in the following sections works well, although many individual elements could be improved.

We start by giving a brief summary of related systems and then present our approach in detail and illustrate the step-by-step processing on a stereo image pair of the Renault part, which is an image pair widely used by researchers in stereo vision. Cross correlation provides an initial disparity estimate, which is refined using a two-view paradigm, an ordering constraint, and the elimination of very small patches. A local interpolation is then performed using a piecewise smoothness assumption to provide the area-based estimate. Edgels are used to accurately locate depth discontinuities and remove the "blurring" inherent in the correlation. Finally, a smoothed disparity surface is obtained from which we can estimate the location of orientation discontinuities. We then present reconstructions of the original scene along with the extracted discontinuities. We show the results of the processing for seven other image pairs from various domains.

## II. RELATED WORK

Our goal is not to provide a complete survey of the state of the art in computational stereo; for these, we refer the reader to Barnard and Fischler [1] or, more recently, to Dhond and Aggarwal [2]. In the following, we briefly mention the ideas that have most strongly influenced our work.

### B. Area-Based Methods

Area-based methods have been applied successfully to the analysis of aerial terrain images, where the surface varies smoothly and continuously. They offer the advantage of directly generating a dense disparity map but are sensitive to noise and breakdown where there is a lack of texture or where depth discontinuities occur.

Mori *et al.* [3] used an iterative prediction and correction method to improve their area-based processing of aerial photographs by varying the size of the window, and they then used edgels to verify their prediction. Their program was only demonstrated for aerial imagery, and they had no method of handling occluded areas. This adaptive cross correlation was reintroduced as part of a coarse-to-fine hierarchical control structure by Quam [4] with the same limitations.

Hannah [5], [6] also showed improvement in area-based cross correlation by using dense features abstracted from the intensity data and introduced heuristics for inferring the distinctions between occlusions, correspondence errors, and off-the-edge overlaps. She has implemented a complete system [7] for stereo processing with little or no operator intervention. This system operates in several passes over the data. First, a set of well-scattered, reliable matches are obtained by locating interest points based on variance and edge strength and then utilizing an unconstrained hierarchical match algorithm. Next, a camera calibration is performed, and an epipolar-constrained hierarchical matching algorithm is used to match the interest points. Those points that are evaluated to have the most reliable matches are used as "anchors" for a final matching of all of the interest points. An interpolation may also be performed to construct a regular grid of points if desired.

### B. Feature-Based Methods

The feature-based approaches match more abstract features, rather than matching texture regions in the two images, since such features are less sensitive to noise. Feature-based analysis provides more precise positioning (for the feature) in the individual images, and it can attain correspondingly higher accuracy for its correspondences in 3-D (see Arnold [8], [9]). The most commonly used features are points along the edges of intensity discontinuities. These points, which are termed edgels for edge elements, are useful because they represent the points at which high-confidence, unambiguous matches may most likely be made. However, the feature-based methods provide only sparse matches and require interpolation as well as some method for modeling occlusion. In addition, the feature-based process may be confused by large local change in disparity, and it is very difficult to incorporate the smoothness assumption into the matching strategy since it is most likely to be violated at the edges.

Marr and Poggio [10] proposed a computational model of human stereo vision, using zero crossings in the Laplacian of the Gaussian of the intensity, as a matching feature. They suggest that three constraints should be satisfied in choosing global correspondence: compatibility, uniqueness, and continuity. The latter is similar to the figural continuity constraint proposed by Mayhew and Frisby [11]. Grimson [12] implemented an improved version of this model, which gives good results when there is a sufficiently dense set of features. Arnold [8] and Baker [13] used various forms of the Viterbi dynamic programming algorithm to match edges, and Ohta and Kanade [14] extended Baker's interscanline search, again using dynamic programming, to find an optional matching surface (however, their 3-D search is very expensive). Medioni and Nevatia [15] match linear edge segments that cut through the epipolar lines and thus automatically ensure interscanline continuity. Hoff and Ahuja [16] attempt to combine the feature matching, contour detection, and surface interpolation into one process. Their results are very impressive, but they fail when their matching features (zero crossings) are too sparse; in addition, they cannot accurately locate the surface discontinuities.

### C. Interpolation

In both area- and feature-based stereo correspondence methods, it is often necessary to interpolate values for those regions for which no disparity can be found. Some methods such as those by Hoff and Ahuja [16] and Boulton and Chen [17] have included the interpolation into the normal processing. In addition to interpolating, these methods are also used to smooth surfaces and to isolate depth discontinuities.

Grimson [12] interpolated surfaces from sparse depth data using variational methods. Terzopoulos [18] attempted to locate discontinuities by locating significant inflection points on the resultant surface. Blake and Zisserman [19] introduced their "graduated nonconvexity" algorithms, which allow the direct search for depth discontinuities and orientation discontinuities, respectively.

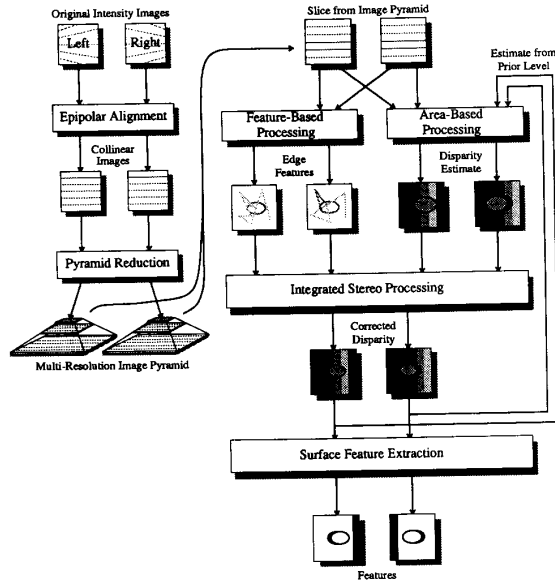


Fig. 1. Stereo vision system flow diagram.

Saint-Marc *et al.* [20] present a method to smooth the surface while preserving discontinuities, which facilitates the detection of discontinuities, and Sinha and Schunck [21] have recently introduced another process for preserving discontinuities while performing the surface reconstruction using a weighted bicubic spline adapted across discontinuities.

Any one of these interpolation methods may be used to extract the discontinuities in the manner that we suggest by adding a strong preference for the existence of the discontinuity contour (either depth or orientation) to occur at edgels or along edges.

#### D. Stochastic Methods

Geman and Geman [22] introduced the idea of using Gibbs distributions to model the spatial correlations between neighboring regions within images along with the use of an additional cost constraint (the "line-process") that represents the presence or absence of a surface discontinuity. Marroquin [23] and Gamble *et al.* [24], [25] have used this model to perform similar fusion between early vision processes. The results are comparable with those obtained from traditional methods, but stochastic approaches are computationally expensive in a typical implementation and contain nonrobust parameters that need specific tuning to each scene. Our approach has the advantage of being deterministic and is robust with respect to its internal parameters.

### III. DESCRIPTION OF THE METHOD

#### A. Overview

Fig. 1 shows the overall control flow through the stereo vision system. First, the original images are adjusted so that their corresponding epipolar lines lie along corresponding rasters and that the range of observable disparities is approximately



Fig. 2. Renault intensity image pair, with row 188 highlighted.

centered around zero. Next, the resulting images are reduced by a convolution with a Gaussian and subsampled by a factor of two along each axis to form a pyramid of three image pairs. Each pair of slices from this stereo pyramid is separately processed starting with the coarsest (most reduced) pair.

The initial feature- and area-based processes proceed independently to produce a set of edge features and a dense disparity estimate. These are then combined to produce a dense disparity estimate with less blurring.

At each level of the image pyramid, this estimate is used to improve the matching in the next finer level, and the last set of results are used to extract the surface features: specifically labeling all of the points as being visible or not, providing a confidence for the generated values, and marking the depth discontinuity contours. Further processing may then be performed to smooth the discrete surface, interpolate through unknown areas, and extract the orientation discontinuities.

#### B. Initial Processing

The initial processing is performed partly by hand and generates the registered image pyramid expected by the actual stereo processing. The first step is to crop and align the images. This is achieved in two steps: First, the image pair is cropped to give a zero disparity near some fixation point. This serves as the center of the fusion interval during the later processing. The second step is to provide an **approximate** epipolar alignment. This is done by providing matches for four points, where one is in the neighborhood of each corner of the images. These are then used to generate a translation and skewing of all of the points in the images.

Once the collinear images are generated, a series of coarser resolution images are generated by a series of reductions in size by a factor of two using a Gaussian convolution filter. We have used three-level pyramids for all of the examples in this paper.

Fig. 2 shows the Renault part stereo pair.<sup>1</sup> The image is  $251 \times 256 \times 8$  bits, and the total disparity ranges from -30 (near) to 15 (far) pixels. This example is being used as an illustration of each step of the process.

#### C. Area-Based Process

Area-based processing matches a measure of the local

<sup>1</sup>The images are displayed as side-by-side pairs for viewing with standard stereo lenses. Many of the results will also be shown as a stereo pair representing the values with respect to the left and right images.

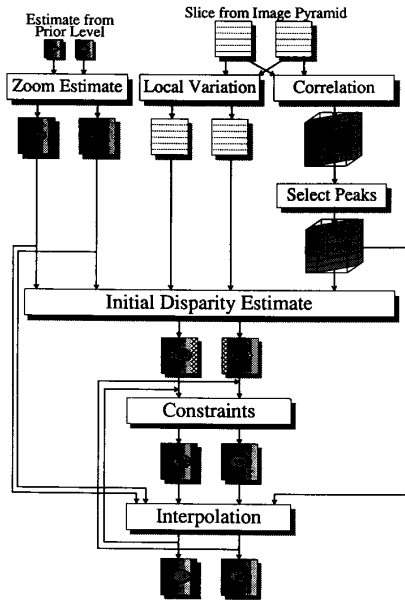


Fig. 3. Area-based processing flow diagram.

texture in the image pair to produce a dense disparity map. The input to the area-based process is a stereo pair of intensity images and an optional pair of disparity estimates from a prior pass. The overall process is shown in Fig. 3.

1) *Local Variation*: We first compute an estimate, which we call the local variation of the image texture. This is similar to an inverted autocorrelation and returns a small value when there is little or no matchable texture. This is used to mark areas for which no match should be generated during the initial disparity estimation. In the special case in which there are no features within the correlation window, as indicated by a very low value in a local variation measure, a value of zero is assigned. This "local variation" is defined by the formula

$$V_{x,y} = \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} 1 - \frac{\bar{x}_j}{\sqrt{\left(\frac{n-1}{n}\right)\sigma_j^2 + \bar{x}_j^2}}$$

where the mean  $\bar{x}_j$  and variance  $\sigma_j^2$  are calculated along each row of an  $n \times m$  window.  $V_{x,y}$  becomes a minimum of zero when all of the values in each row of the window are the same, reaches a maximum when variance of the row is large, and is indexed by the mean intensity so that any texture present in darker regions is emphasized. The reason why the value is summed over the raster rather than within the window is that a horizontal striped pattern is not matchable and must be treated as a textureless region. Any similar measure of texture that is applied along the individual epipolar lines within the correlation window gives similar results.

2) *Correlation*: Next, a normalized cross correlation (see pp. 15–16 of [26]) of the two intensity images is calculated. This operation results in a volume begin generated, whose dimensions are the width of the left image by the width of the right image by the number of rows. Since we are looking for matches within a fusion interval about the fixation point, we

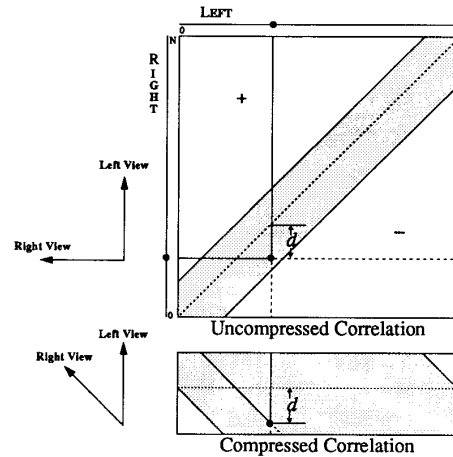


Fig. 4. Compressed cross-correlation array.

do not need to calculate values for each point in this volume (only in a band about zero disparity). Fig. 4 depicts a slice through this volume showing the cross correlation generated for one row of data. The slices through the left and right images form the "axes" of a square image with the correlation data distributed in a narrow band along a diagonal. Here, each grid point represents the result of correlating the values in a window centered at a pixel from the left slice with a similar group of values from a window centered at a pixel from the right slice according to the correlation formula

$$C_{x,y,d} = \frac{(\sum_A \text{Left}_{x,y} \text{Right}_{x+d,y})^2}{\sum_A \text{Left}_{x,y}^2 \sum_A \text{Right}_{x+d,y}^2}$$

where  $A$  is the correlation window centered at the pixels  $(x, y : x + d, y)$ . In all of the examples shown here, we use an  $11 \times 11$  window.

However, this representation is difficult to work with for two reasons: First, as a data structure, a lot of space is wasted since the disparity is usually an order of magnitude smaller than the  $x$  and  $y$  dimensions of the images. The second reason is conceptual. Since the angle between the camera origins and a point is actually about  $10\text{--}15^\circ$ , it is easier to visualize the matching when the data structure also has a narrow angle between the viewpoints. The cross correlation represents the viewpoints at right angles. The compressed cross-correlation format that we use is a compromise that provides a closer-to-normal  $45^\circ$  between the viewpoints and wastes very little storage space. This is accomplished by sliding slices from the filled portion of the array such that the zero-disparity (dashed) line is horizontal.

Fig. 5(a) shows a slice through this compressed cross-correlation volume at row 188 (see Fig. 2) from the bottom, which cuts through the two lobes of the Renault part. The better matches are shown as lighter regions.

Fig. 5(b) shows the peaks (likely matches) selected from this slice, where "peaks" are defined to be the set of points in the cross-correlation volume that exhibit the following:

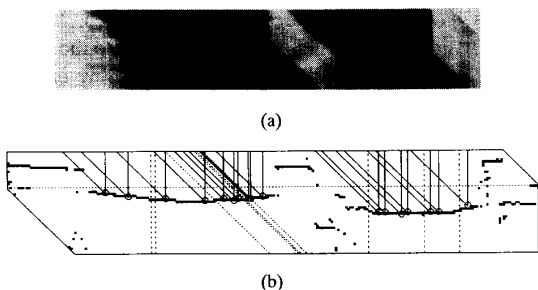


Fig. 5. Cross-correlation slice (row 188 of the Renault pair in Fig. 2). The corresponding edge matches from the same row in Fig. 9 are shown in (b): (a) Left view of the cross-correlation; (b) correlation peaks and edge matches.

1. Each peak has a value greater than or equal to its 4-connected neighbors in the uncompressed correlation volume.
2. Each peak has a value greater than or equal to one half of the value of the strongest peak along each viewing direction.

The first inequality is the real definition of a peak. In this case, a peak means that the correlation centered at a given pair of pixels must be greater or equal to that of the values centered at a pixel to either side (along the epipolar row) of the given one in either of the two images. The second inequality limits the search space by removing any peaks that are relatively weak, which we have arbitrarily chosen to define as one half the value of the strongest peak. This restriction is applied along each view; therefore, every peak is a member of two sets of possible matches: those corresponding to a single pixel in the left image (left view) and those to a single pixel in the right image (right view).

3) *Estimate from a Prior Level:* There may or may not be an estimate from a prior level of the image pyramid. If there is one, then it must be scaled up by doubling the image size and the disparity values. This is used to force a preference for a weaker match near the estimate over a strong match that is farther away. "Near" is defined as less than or equal to three pixels, which allows for a pixel error at the prior level and at the current level. In the region near a discontinuity (i.e., within one half the correlation mask width), points near either surface are accepted, and those points that have no prior estimate are unbounded.

4) *Initial Disparity Estimate:* Fig. 5(b) shows the extracted peaks (best correlation-based matches) overlaid with the matched (solid) and unmatched (dashed) edgels (see Fig. 9).<sup>2</sup> By looking at Fig. 5, it should be clear that for most points, the correct disparity can be found by simply extracting the peaks in the array. Once this is done, we need to focus our attention on the problem areas only. The first occurs in areas without measurable texture. For these, it is important *not* to generate matches from the cross-correlation information. Where multiple matches (peaks) occur, we prefer the highest peak (best correlation) on which both views can agree. It is possible (although unlikely) for the correct peak not to

<sup>2</sup>The edges were matched using Medioni's segment matcher [15].

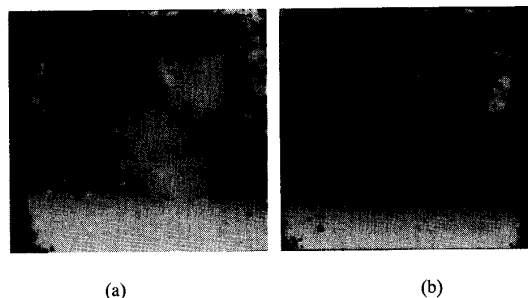


Fig. 6. Initial disparity estimate derived by selecting the best peaks of the cross correlation of the images in Fig. 2: (a) Disparity (left); (b) disparity (right).

be selected; therefore, we also must impose a smoothness constraint on the surfaces, that is, the peaks should form a piecewise continuous "ridge-line." In addition, we assume that order reversals due to the spatial shift between viewpoints cannot occur within the relatively narrow fusion interval. An example of such a reversal is the two small ridges in the middle of Fig. 5(b). The remaining incorrect areas represent points that are either occluded in one of the views or "visible" points on which the two views cannot agree. These areas may be removed by using the results from multiple resolutions to help select the longer ridges in preference of locally higher ones. Finally, there is the "blurring" beyond the actual edge, which can be seen in Fig. 5(b), where the peak ridges extend beyond the matched edgels (the solid vertical and 45° lines) that represent the object boundaries.

Now, the initial disparity estimate is made. Here, the strongest peak (disparity) is selected at each location from both the left and right viewpoint, subject to having enough local texture as indicated by the local variation, and optionally guided by a prior estimate.<sup>3</sup> This gives us an estimate of the disparity such as that shown in Fig. 6.

5) *Constraints:* The last part of the area-based processing consists of iteratively applying a set of constraints to the matched points to identify and remove unlikely matches and then performing a surface interpolation guided by the cross-correlation peaks to build on the remaining, high-confidence matches. This process is repeated three times and allows the first, "Two-Views," constraint (see below) to be relaxed so that a many-to-one match can be accepted. This allows matching of surfaces that have a gradient larger than 1:1 in disparity space.

a) *Two-Views Constraint:* This first constraint makes use of the redundancy inherent in the two views of the same scene: one matching pixels of the left image to pixels of the right and the other matching pixels of the right image to those of the left. If these initial estimates do not agree, that is, if a pixel of the left image is "matched" to one in the right image that, in turn, has not been matched to that same pixel in the left image, then we can safely assume that either one or both estimates

<sup>3</sup>Points for which there is no current match and those for which the match is "removed" are set to an unused, null, disparity value. In the examples of disparity surfaces, the lighter regions indicate the pixels that are closer to the viewer, and the darker regions indicate those that are farther away. However, totally black pixels are those that have no match and are considered to have an unknown disparity.

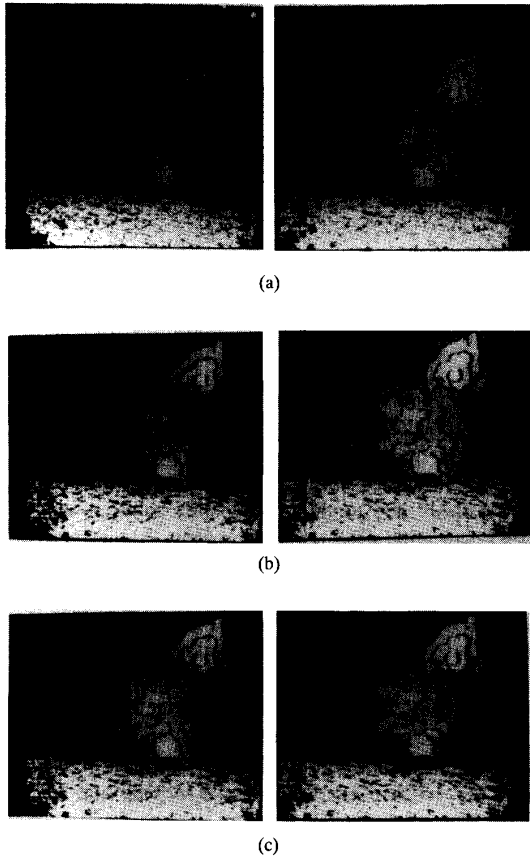


Fig. 7. Application of constraints to the raw estimate in Fig. 6: (a) After checking for agreement of both views; (b) after removing order reversals; (c) after removing isolated pixels.

are incorrect. At first, this agreement between matched points must be exact, but during the later interpolation, this constraint is relaxed to allow a disagreement of from two to four pixels so that the disparity surface is allowed to slope and is not forced to be flat and perpendicular to the camera axis. Fig. 7(a) shows the disparity points for which the estimates from each view are in agreement.

*b) Order Constraint:* The second constraint forces the removal of small patches whose order from left-to-right is reversed from one view to the other. This is reasonable since we assume that there is a limited fusion interval about the selected vergence, and it is very unlikely for small wire-like objects to be strung in front of the scene. The effect of this constraint is shown in Fig. 7(b). Each scanline is processed left-to-right, and each piecewise continuous patch is numbered. Then, the corresponding patches are examined via their matches from the other view. When the ordering of a pair is reversed, then the patch with the least local support is removed (that is, its matches are set to a null value).

*c) No Isolated Pixels Constraint:* Isolated pixels are those pixels that have a disparity that differs more than 2.5 pixels from the average value in the surrounding  $5 \times 5$  neighborhood. Furthermore, pixels for which fewer than six (of its 24)

neighbors have nonnull values are also removed. Now, the estimate, which is shown in Fig. 7(c), is reduced to those points that have a high probability of being correct. Any erroneous matches that remain at this point have sufficient local support to be retained; therefore, it is worthwhile to remove a few good points in this process, and they are likely to be replaced by the interpolation in the next step. The remaining points are those in which we have the highest confidence.

These constraints are applied in a heuristic manner that was found to work well for the range of images that we have studied. The limitation to three passes was determined from observation as being sufficient.

*6) Interpolation:* The last step of area-based processing is the interpolation through the unknown areas. This is done by using the remaining (high-confidence) matches as "seeds," expanding the disparity estimate along the disparity surface formed by adjacent peaks. A cross section of these surface patches is visible as the lines of peaks in Fig. 5(b). The interpolation proceeds in three steps, and between each of these steps, the constraints are reapplied. Correlation peaks that are adjacent to the existing "seed" matches are added to the surfaces according to the following rules:

- If more than one adjacent match is possible, then the strongest is used.
- If two or three are equal, then they are selected according to the ordering: same disparity, closer, farther.

*a) Pass One:* During the first interpolation pass, only exact matches of peaks between the two views are allowed, that is, only peaks that exist in both views with the same disparity. The interpolation is applied for six cycles (one half the width of the correlation mask), and the resulting surface is shown in Fig. 8(a).

*b) Pass Two:* During the second pass, a small difference in disparity is allowed so that surfaces that are changing rapidly may be represented by a one-to-many match between the two views. This means that the selected matching points from the two views may differ by some small amount. This threshold represents the amount of gradient that is allowed for a smooth surface (rather than restricting the matches to form only flat surface patches perpendicular to the camera axis). This surface cannot be represented by either single view with only one disparity value for each cell; however, the two views together can adequately model it. We always use a difference of  $\pm 2$  for the coarsest level of the pyramid and allow another pixel difference for each finer level. Fig. 8(b) shows the disparity surface at the finest level for the Renault part after this step is performed for six cycles with an allowed disparity difference of  $\pm 4$ .

*c) Pass Three:* Finally, small holes of up to six pixels in extent are filled with a median value. This is because the definition of peaks is quite restrictive, and therefore, holes exist in the disparity surfaces formed by the peaks. A median filter works quite well for this, although a better approach would be to reevaluate the local peaks in the cross-correlation volume. Fig. 8(c) shows the final disparity surface. This step is performed for three cycles.

The remaining unknown points are considered to be either

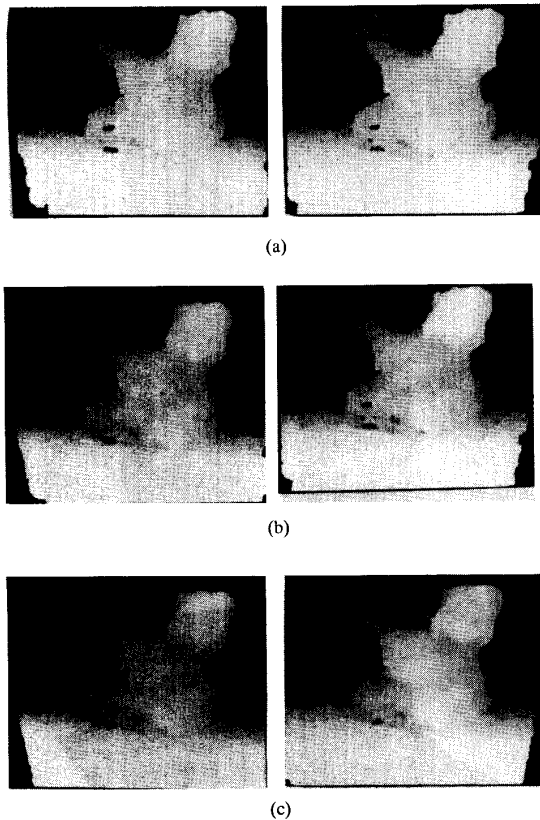


Fig. 8. Application of Interpolation Passes to the "seed" estimate in Fig. 7(c): (a) After the first pass (strict agreement); (b) after the second pass (weak agreement); (c) after the final pass (median filter).

occluded or to have insufficient texture to be matched, depending on the associated value of the local-variation image. Fig. 8(c) represents an impressive result for an area-based process but does not have the accuracy that we desire due to the tendency to "blur" beyond the object contours (the more strongly textured surface "leaks" into the less textured region).

#### D. Disparity-Map Refinement

The blur just mentioned can be refined using edge information. We can use edgels (or edges) from any source; here, Fig. 9 shows the edges resulting from the intensity images in Fig. 2 using Nevatia and Babu's LINEAR feature detector [27].

Instead of using edgels matched by an independent feature-based process, we use the monocular edgels in the 2-D image and associate with them the disparity obtained from the area-based process. Since depth and orientation discontinuities usually give rise to these intensity edges, some of the detected edgels should correspond to discontinuities. Therefore, we smooth the disparity map, keeping the disparity at the edgels fixed. This is implemented using the adaptive smoothing formalism developed by Saint-Marc *et al.* [20]. All points whose disparity shifts by more than a constant amount (we use 1.0 pixels) are discarded, removing the "blurred" fringe around the actual contour edges. Note that we are using the edgels as

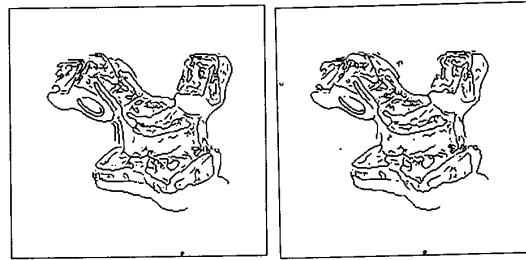


Fig. 9. Edges generated by LINEAR from the images of Fig. 2.

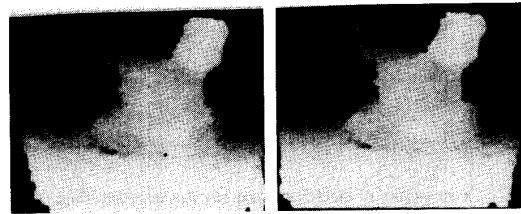


Fig. 10. Disparity map after incorporation of edge information.

monocular cues so that even those edges with an orientation close to the epipolar line directions play an active role, whereas they are discarded in all other feature-based approaches.

Fig. 10 illustrates the result of this process on the results of the refined disparity map in Fig. 8(c). From this, we can now put the pixels into four groups. Fig. 11 shows a cross section depicting the blurred regions that were removed, the visible (but unmatched) surfaces now uncovered, and the occluded regions. Those unknown points that are left over may be considered points that should be visible but were unmatched for various reasons. The occluded regions are further subdivided into those areas occluded by a surface and those occluded by border of the other image. In Fig. 12, the intensity values mean the following:

1. Known disparity values are white.
2. Points that should be visible, but were not matched, are light gray.
3. Points visible only from this view because of image clipping in the other view (or initial masking) are dark gray.
4. Penumbral points, which are visible only from this view, are black.

Given this data, we now attempt to reproduce the original object as shown in Fig. 2. First, we interpolate along those regions labeled in Fig. 12 as being visible. This interpolated disparity image is shown in Fig. 13(a). The shaded representations of these disparity images are shown in Fig. 13(b). These were produced by assigning the smoothed surface a simple reflectance function and positioning an imaginary light in space. Most of the error is due to digitization noise and some flattening at the limb edges.

Fig. 13(c) is a 3-D plot of the left-image-disparity data from Fig. 13(a) sampled at every third pixel. The disparity ranges from -15 (far) to +30 (near), and the zero plane of the plot is moved backward 17 from the zero plane of the image. A

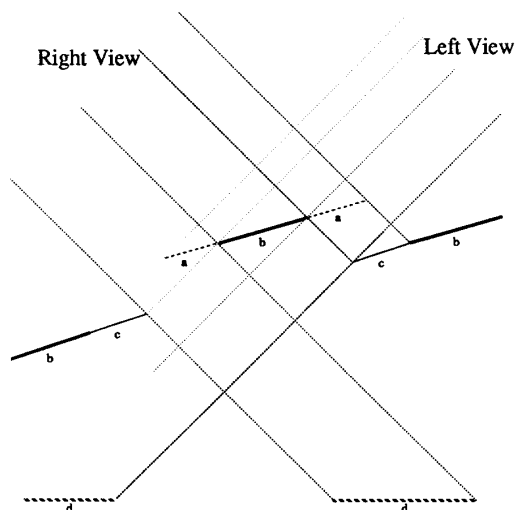


Fig. 11. Cross section of labeled surface: (a) Removed blurring; (b) visible surface; (c) visible but unmatched surface; (d) occluded regions.



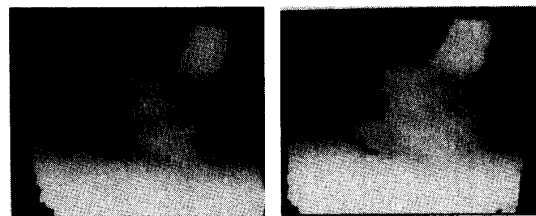
Fig. 12. Labeled points from Fig. 10.

rendered view is also shown in Fig. 13(d) in which the original intensity values (Fig. 2, left) were projected onto the surface rotated  $45^\circ$  about the  $y$  axis and scaled by 4 in the original  $z$  axis.

In addition, we can provide a rough estimate of the confidence at those points that have been assigned a disparity. The points with the highest confidence are the "seed" points from Fig. 7(c) that have been retained after mixing with the edgel information. Next, the other points that correspond to peaks of the cross correlation are assigned a medium confidence, whereas the remainder of the points are given a low confidence.

### E. Surface Segmentation

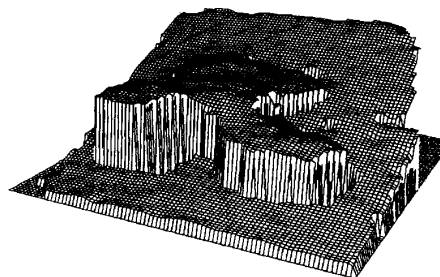
Our goal is not simply to clean up the disparity surface or to produce a rendered image of the scene but also to produce a description of the scene. We would like to be able to segment the scene into surface patches that represent the faces of the objects composing the scene and eventually to group these individual patches into objects in the scene. In order to do this, we must locate the borders or edges of these patches. These borders are the locations where either the surface breaks at depth discontinuities or where it creases at orientation discontinuities.



(a)



(b)



(c)



(d)

Fig. 13. Final disparity maps and their shaded representation: (a) Interpolated; (b) shaded. Renault part—reconstructions: (c) 3-D plot of the final result; (d) 3-D rendering of the final result.

From the prior processing, we know the location of the penumbral, visible, and off-the-edge areas, and we know the resolution of disparity surface: We can locate the depth discontinuities by looking for any jumps that are greater than the resolution step size and occur between points that are visible or between borders of the penumbral and the visible areas. These discontinuities are shown in Fig. 14(a).

Locating the creases in the disparity surface is more difficult since the estimated curvature is extremely sensitive to quantization noise [28], which requires us to smooth the

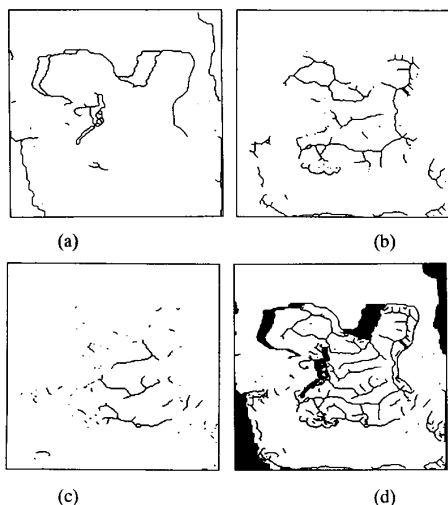


Fig. 14. Discontinuities from Fig. 13(b): (a) Depth discontinuities; (b) convex folds; (c) concave folds; (d) combined discontinuities and unknown areas.

local surface as much as possible. Smoothing, however, also removes or attenuates the desired features. We need a process that will smooth away noise due to the quantization and will preserve the creases due to the underlying surface orientation discontinuities. Our approach is to first repeatedly smooth the surface with a binomial (or other approximate-Gaussian kernel) but attenuate sharply any change over one half of the step size from the original discrete surface. This gives us a smooth surface with some ripple while preserving the creases where there is sufficient resolution. To finish the process, we apply the adaptive smoothing approach proposed by Saint-Marc *et al.* [29] to the derivative of the partially smoothed disparity data. This process preserves the stronger creases while removing the weaker ones. Finally, we calculate the local curvature at each point and threshold and thin the maximum positive and negative curvatures. Fig. 14(b) and (c) shows the estimated negative and positive creases, whereas Fig. 14(d) combines all of the the discontinuities with the unknown regions to show where the depth is known and where surface features are known or suspected.

The poor localization of the creases and, to a lesser extent, of the depth discontinuities are due, in large part, to the lack of available resolution. Fig. 15(a) and (c) shows cross sections through a small section of the Renault image both before and after the smoothing described above. Fig. 15 (b) and (d) shows the same cross sections through a similar section at greater resolution. (This section was cropped from an image at twice the resolution, doubled in size, and then processed separately.) Fig. 16 shows the extracted discontinuities from each image. The low- and high-resolution images have been adjusted for comparison. The disparity change across the face in the low-resolution image is about 3 pixels, whereas for the high-resolution, the change is around 7 pixels. Although the sharp horizontal crease across the bottom of the images is clear at both resolutions, the more rounded vertical crease to the right of center is completely distorted at the lower resolution.

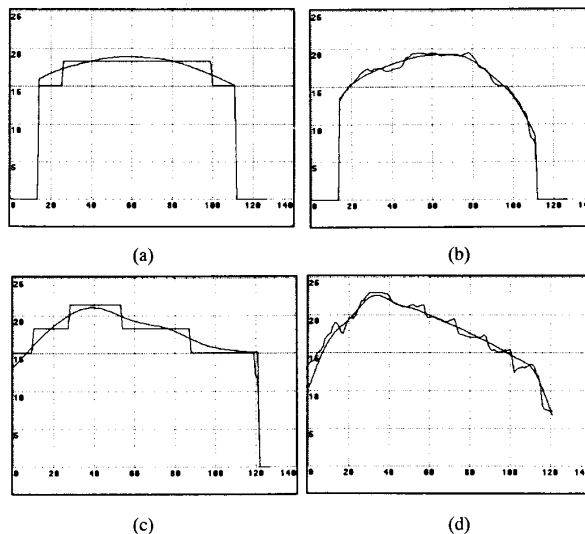


Fig. 15. (a) and (b) are horizontal slices through part of the Renault image (shown in Fig. 15(b)), whereas (c) and (d) are vertical slices: (a) Row 66, low resolution; (b) row 66, high resolution; (c) column 60, low resolution; (d) column 60, high resolution.

#### IV. EXPERIMENTAL RESULTS

Some other observations and specific problems are discussed in context below, with several examples of the stereo images processed by our algorithm. The overall complexity of the current serial implementation of this process is  $\mathcal{O}(H \cdot W \cdot D)$ , where  $H$  is the height,  $W$  the width, and  $D$  the depth of the fusion interval, all in pixels. However, the processing is inherently local and can be performed in constant time, except for the search for the strongest peak of the disparity that can be done in  $\mathcal{O}(\log(D)^2)$ . In the following examples, the times are for a serial implementation written in Lisp and running on a Symbolics 3650 under Genera 7.2. The running time may be estimated as  $H \cdot W \cdot D \cdot 12$  ms/voxel.

Typical times for a serial implementation are therefore about 1 hr for a  $128 \times 128$  image with 25 pixels total disparity or about 5 hr for a  $256 \times 256$  image with 25 pixels total disparity.

##### A. Random Dot

The first example to be shown is an extreme case: a random-dot stereogram “wedding cake” with a disparity change of 8 pixels between planes. The test image, which is shown in Fig. 17, is  $128 \times 128 \times 1$  b and has an approximately equal number of white and black pixels. This stereo pair shows the area-based processing under the best conditions—where each correlation window has a large amount of information and there is no noise. Table I shows the processing time for each level of processing. The total disparity of the finest level is 21, which is  $\pm 8$  pixels about zero plus an extra two pixels at the extremes. The errors come from two sources: First, the corners are rounded because, with no edge to guide the processing, the cross correlation smoothes over corners. Second, where there is an occluded region, chance clustering yields mismatches that

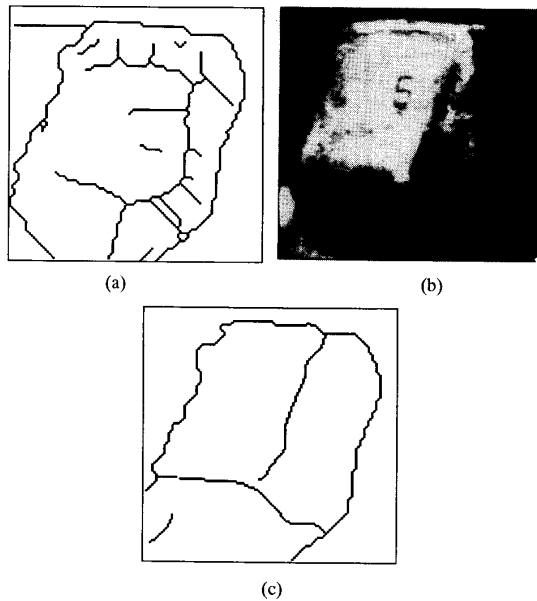


Fig. 16. Combined discontinuities from low- and high-resolution images: (a) Low resolution; (b) cropped intensity image; (c) high resolution.

TABLE I  
WEDDING CAKE—PROCESSING TIMES

	<i>H</i>	<i>W</i>	<i>D</i>	Time
Coarse	32	32	9	3m 29s
Medium	64	64	13	15m 02s
Fine	128	128	21	1h 30m 20s
Total				1h 48m 51s

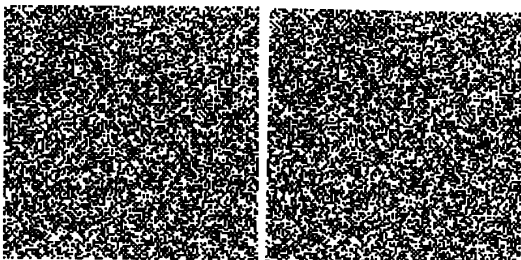


Fig. 17. Wedding cake—Original intensity images.

give rise to the rough vertical boundaries visible in Fig. 18. The most important thing to be noted here is the conservative result that yields 0.76% error and marking 2.97% unknown rather than simply guessing when insufficient information is present.

Table II shows the correct and incorrect matching statistics at the finest level of the matching process. The first column gives the viewpoint: either "Left" or "Right." The second column shows the correct matches: The **binocular** values are for those points that were correctly matched, whereas the **monocular** ones are those that were correctly left unmatched. The third column gives the incorrect matches, which are also broken down into incorrectly matched points that are visible in both views (binocular) and those occluded points that were

TABLE II  
WEDDING CAKE—MATCHING STATISTICS

View	Good Match/Label		Bad Match/Label	
	Binocular	Monocular	Binocular	Monocular
Left	84.64%	11.49%	0.16%	0.62%
Right	84.84%	11.57%	0.19%	0.54%
Total	96.27%		0.76%	

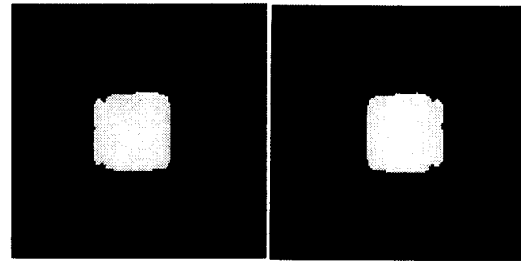


Fig. 18. Wedding cake—Disparity surface.

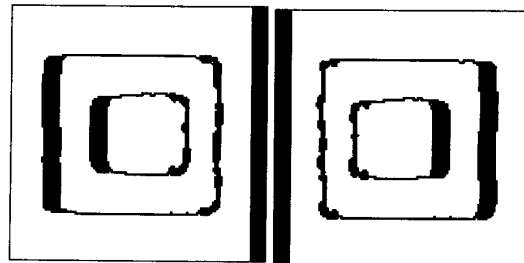


Fig. 19. Wedding cake—Depth discontinuities and occluded regions.

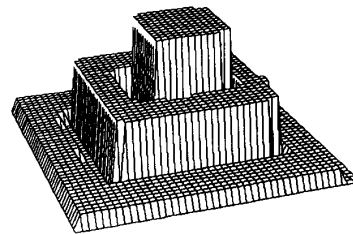


Fig. 20. Wedding cake—3-D plot of the integrated results.

mistakenly matched with some point (monocular). The total does not add to 100%. This is because the program did not give a match for some points that were visible in both views, which represents an acceptable result that is neither a correct nor an erroneous value. Therefore, the final results shown here have 96.27% correctly matched points with 0.76% in error and 2.97% unknown.

Fig. 19 shows the depth discontinuities and the occluded regions of this stereo pair, and Fig. 20 shows a 3-D plot of the image reconstructed from the matched pixels.

### B. Books

Fig. 21 shows a  $488 \times 488 \times 8$  stereo pair of a stack of books obtained from the University of Illinois, courtesy of Dr. W. Hoff, who is now with Martin Marietta. The disparity of the stereo pair ranges from 50 pixels (far) to -18 pixels (near). Ta-

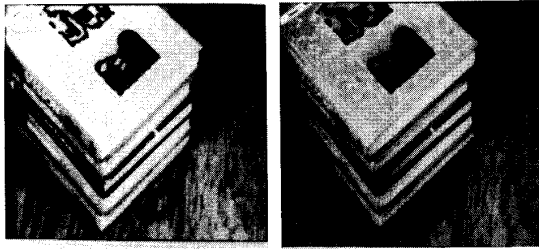


Fig. 21. Books—Original intensity images.

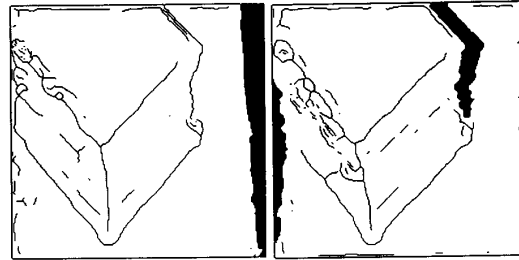


Fig. 24. Books—Surface features.

TABLE III  
BOOKS—PROCESSING TIMES

	<i>H</i>	<i>W</i>	<i>D</i>	Time
Coarse	122	122	22	54m 03s
Medium	244	244	39	4h 57m 16s
Fine	488	488	73	28h 49m 32s
Total				34h 40m 52s



Fig. 22. Books—Disparity surface with only the finest level.

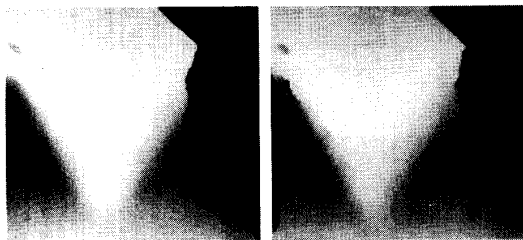


Fig. 23. Books—Disparity surface.

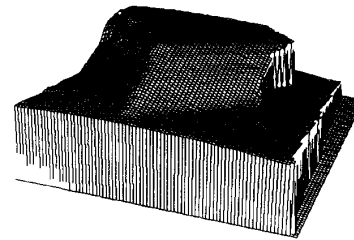


Fig. 25. Books—3-D plot of the integrated results.

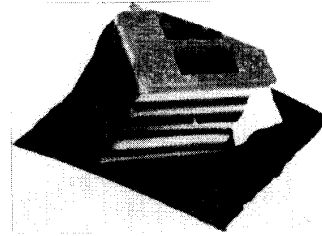


Fig. 26. Books—3-D rendered view of the integrated results.

ble III shows the processing time for each level of processing.

This example most clearly demonstrates the improvement provided by the multilevel pyramid. Fig. 22 is the result at the finest level without the use of the image pyramid. The multiple incorrect matches are due to local patches that form better matches than the correct patches because of geometric and photometric distortions. Fig. 23 shows the final disparity views, and Fig. 24 shows the surface features located. Finally, Figs. 25 and 26 show the 3-D views of the disparity and the rendered scene, respectively.

C. Blocks

Fig. 27 is a  $295 \times 295 \times 8$  image taken with a large angular separation and areas of both high and very low texture. In addition, the fusion interval was intentionally chosen such that the back wall falls outside of it; therefore, the correct value cannot be obtained there. In addition, the surface markings

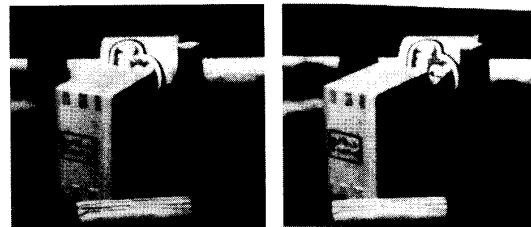


Fig. 27. Blocks—Original intensity images.

TABLE IV  
BLOCKS—PROCESSING TIMES

	<i>H</i>	<i>W</i>	<i>D</i>	Time
Coarse	73	73	23	18m 43s
Medium	147	147	42	2h 32m 33s
Fine	295	295	80	14h 18m 33s
Total				17h 09m 49s

on the "Play-Doh" container have a higher contrast than the texture of the wood figures. Finally, the lower wooden cylinder has a grain pattern that is approximately parallel to the epipolar lines of the image. Thus, most of the adverse conditions for our approach are present in the image. Table IV shows the processing time for each level of processing.

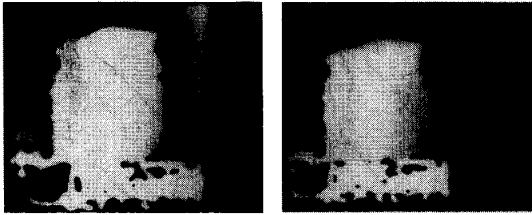


Fig. 28. Blocks—Disparity surface.

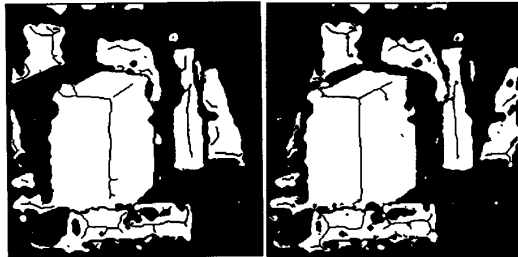


Fig. 29. Blocks—Surface features.

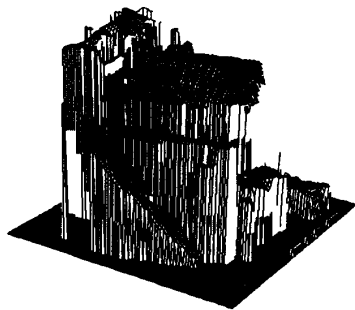


Fig. 30. Blocks—3-D plot of the integrated results.

Fig. 28 shows the recovered disparity surfaces, whereas Fig. 29 shows the unknown regions along with a few depth discontinuities and the orientation discontinuities. Although much of the scene was discarded as unmatchable due to occlusion or lack of texture, the remainder is present and has accurate disparity values. Even though most of the brick wall is absent, some small parts that were just within the fusion interval were correctly matched and can be seen in the 3-D plots in Fig. 30. Finally, Fig. 31 shows a rendered view of the matched parts of this scene. This example provides an illustration of graceful degradation of our algorithm in the presence of a difficult image pair.

#### D. Pentagon

Fig. 32 shows an aerial view of the Pentagon in Washington, DC (about  $38.5^\circ$  north latitude and  $77.0^\circ$  west longitude), which was obtained from Prof. T. Kanade of the Carnegie-Mellon University Computer Science Department. The image is  $512 \times 512 \times 8$ , and the disparity ranges from  $-9$  (near) to  $8$  (far) pixels. Table V shows the processing time for each level of processing.

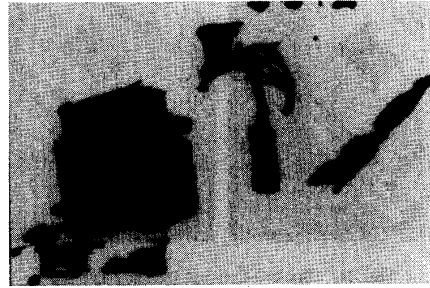


Fig. 31. Blocks—3-D rendered view of the integrated results.



Fig. 32. Pentagon—Original intensity images.

TABLE V  
PENTAGON—PROCESSING TIMES

	H	W	D	Time
Coarse	128	128	10	47m 12s
Medium	256	256	14	3h 39m 05s
Fine	512	512	22	18h 17m 59s
Total				22h 44m 16s

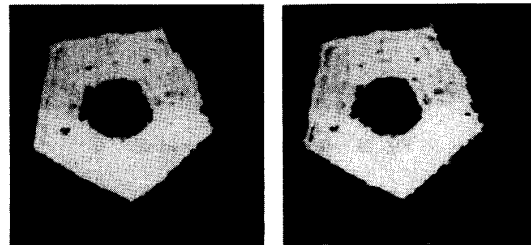


Fig. 33. Pentagon—Disparity surface.

The results are quite impressive but also point out two weaknesses with the method. The first is that when the disparity difference between two surfaces is only about 1 pixel, they cannot be accurately separated, as can be seen around the underpass of the bridge in the upper right corner. In addition, when two edges are very close together, the "blurring" can extend past more than one edge. We have no solution for this except to work at a higher level of resolution or to attempt to obtain subpixel accuracy.

Fig. 33 shows the disparity surface of Fig. 32, Fig. 34 shows a 3-D plot of the disparity, and Fig. 35 shows a rendered view of the same figure.

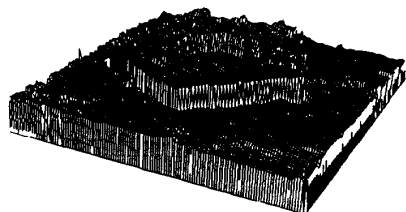


Fig. 34. Pentagon—3-D plot of the integrated results.



Fig. 35. Pentagon—3-D rendered view of the integrated results.

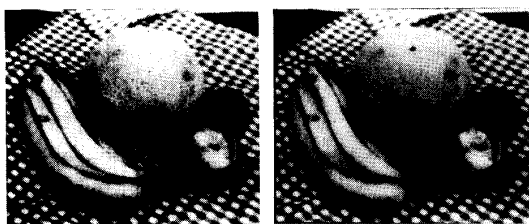


Fig. 36. Fruit scene—Original intensity images.

#### E. Fruit on a Table

Fig. 36 is another image obtained from the University of Illinois courtesy of Dr. W. Hoff, who is now with Martin Marietta. This image is  $256 \times 256 \times 8$  bits. Table VI shows the processing time for each level of processing. Here again, the background is more textured than the foreground. In addition, the tablecloth pattern ripples with a frequency of about 16 pixels, which generates aliasing within the fusion interval. The multilevel matching succeeds, and the image matches are generally correct, except along the lower part of the cantaloupe. In addition, the border of the cantaloupe is rough since the boundary between it and the tablecloth is really a texture boundary and not specified by an intensity edge. This, and the lack of texture in the cantaloupe, cause some points to be marked as unknown along that edge. In addition, the upper left edge was incorrectly matched.

Fig. 37 shows a 3-D plot of this scene, whereas Fig. 38 shows a rendered view.

#### V. FUTURE RESEARCH

Thus far, we have only used the simplest integration of area and feature-based methods. We plan to extend the interaction using more elaborate monocular and stereo cues to guide the area-based stereo matching, especially edge matching to improve the accurate location determination of detected surface features. In addition, we plan to add a multilevel process to propagate and apply information from features at different spatial frequencies to help in the reduction of false matches and blurring.

TABLE VI  
FRUIT—PROCESSING TIMES

	<i>H</i>	<i>W</i>	<i>D</i>	Time
Coarse	64	64	12	11m 27s
Medium	128	128	17	53m 28s
Fine	256	256	29	4h 47m 41s
Total				5h 52m 36s

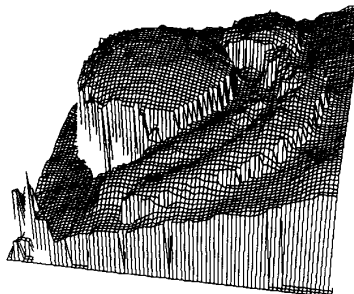


Fig. 37. Fruit scene—3-D plot of the integrated results.

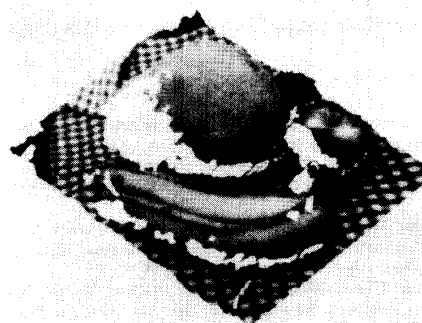


Fig. 38. Fruit scene—3-D rendered view of the integrated results.

#### VI. CONCLUSION

We have developed a complete and working implementation of a stereo algorithm and have presented some ideas whose use in standard stereo correspondence provides some impressive results on several images. In particular, we provide one use for those edgels that are parallel to the epipolar lines and, along with all edgels, serve as boundaries to limit the search for discontinuities in depth or orientation. We have also applied a number of constraints from feature-based approaches to stereo matching into the process of selecting a peak from the correlation volume. In particular, we have demonstrated an important paradigm in stereo matching: the agreement by two views for each disparity value to resolve the ambiguity and verify the matches.

These have allowed us to produce a system that infers more than just the disparity at each point as we can also mark boundary contours that correspond to depth discontinuities and thereby label points as to whether or not they are occluded. In addition, we have made progress in locating orientation discontinuities that should allow image segmentation based on surface properties detected from passive stereo.

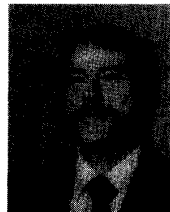
This methodology is robust. When there are no edge features, it acts as a good area-based process, and when there is no texture, it acts as a feature-based system. Furthermore, many existing systems may be easily modified to search for discontinuities in the manner that we suggest by adding a strong preference for the existence of the discontinuity contour (either depth or orientation) to occur at edgels or along edges. Thus far, we have only used the simplest integration of area and feature-based methods to demonstrate the feasibility of a more complete system. Although the serial implementation is slow, we have, with one exception, selected processes that are local in scope so that the routines may be implemented on a parallel-architecture machine. That one exception is the search for the peaks of the cross correlation, which is bound by the depth of the fusion interval.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. P. Saint-Marc of Matra MS2I, France, for the use of his software for adaptive image smoothing and for the shading and rendering of the final results. They would also like to thank the anonymous reviewers for their suggestions to improve this paper.

#### REFERENCES

- [1] S. Barnard and M. Fischler, "Computational stereo," *ACM Comput. Surveys*, vol. 14, no. 4, pp. 553-572, Dec. 1982.
- [2] U. R. Dhond and J. K. Aggarwal, "Structure from stereo—A review," *IEEE Trans. Syst. Man Cybern.*, vol. 19, no. 6, pp. 1489-1510, Nov./Dec. 1989.
- [3] 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.
- [4] L. Quam, "Hierarchical warp stereo," in *Proc. DARPA Image Understanding Workshop* (New Orleans), Oct. 3-4 1984, pp. 149-155.
- [5] M. J. Hannah, "Computer matching of areas in stereo images," Ph.D. thesis, Comput. Sci. Dept., Stanford Univ., Stanford, CA, July 1974; Tech. Rep. STAN-CS-74-438.
- [6] ———, "Bootstrap stereo," in *Proc. DARPA Image Understanding Workshop* (College Park, MD), Apr. 1980, pp. 201-208.
- [7] ———, "SRI's baseline stereo system," in *Proc. DARPA Image Understanding Workshop* (Miami Beach, FL), Dec. 1985, pp. 149-155.
- [8] R. D. Arnold, "Automated stereo perception," Ph.D. thesis, Stanford Univ., Stanford, CA, Mar. 1983; Tech. Rep. AIM-351 and STAN-CS-83-961.
- [9] ———, "Local context in matching edges for stereo vision," in *Proc. DARPA Image Understanding Workshop* (Boston, MA), May 1978, pp. 65-72.
- [10] D. Marr and T. Poggio, "A theory of human stereo vision," Tech. Rep. AI Memo 451, Artificial Intell. Lab., Mass. Inst. Technol., Cambridge, Nov. 1977.
- [11] J. E. W. Mayhew and J. P. Frisby, "Psychophysical and computational studies toward a theory of human stereopsis," *Artificial Intell.*, vol. 17, pp. 349-385, Aug. 1981.
- [12] W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*. Cambridge, MA: MIT Press, 1981.
- [13] H. H. Baker, "Depth from edge and intensity based stereo," Tech. Rep. AIM-347 and STAN-CS-82-930, Stanford Univ., Stanford, CA, Sept. 1982.
- [14] Y. Ohta and T. Kanade, "Stereo by two-level dynamic programming," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 7, no. 2, pp. 139-154, Apr. 1985.
- [15] G. Medioni and R. Nevatia, "Segment-based stereo matching," *Comput. Vision Graphics Image Processing*, vol. 31, pp. 2-18, 1985.
- [16] W. Hoff and N. Ahuja, "Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 2, pp. 121-136, Feb. 1989.
- [17] T. E. Boulton and L.-H. Chen, "Synergistic smooth surface stereo," in *Proc. IEEE Int. Conf. Comput. Vision* (Tampa, FL), Dec. 1988, pp. 118-122.
- [18] D. Terzopoulos, "The computation of visible-surface representations," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 10, no. 4, pp. 417-438, July 1988.
- [19] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
- [20] P. Saint-Marc, J. S. Chen, and G. Medioni, "Adaptive smoothing: A general tool for early vision," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 13, no. 6, pp. 514-529, June 1991.
- [21] S. S. Sinha and B. G. Schunck, "Discontinuity preserving surface reconstruction," in *Proc. Conf. Comput. Vision Patt. Recogn.* (San Diego, CA), June 1989, pp. 229-234.
- [22] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984.
- [23] J. L. Marroquin, "Deterministic Bayesian estimation of markovian random fields with applications to computational vision," in *Proc. IEEE Int. Conf. Comput. Vision* (London, England), June 1987.
- [24] T. Poggio, E. B. Gamble, and J. J. Little, "Parallel integration of vision modules," *AAAS Sci.*, vol. 242, pp. 436-440, Oct. 21, 1988.
- [25] E. B. Gamble, "A discontinuity detector based on the pervasive noise in surface property data," in *p-cvpr* (Lahaina, Hawaii), June 3-6 1991, pp. 757-758.
- [26] R. Nevatia, *Machine Perception*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [27] R. Nevatia and K. Babu, "Linear feature extraction and detection," *Comput. Graphics Image Processing*, vol. 13, no. 3, pp. 257-269, July 1980.
- [28] P. J. Flynn and A. K. Jain, "On reliable curvature estimation," in *Proc. Conf. Comput. Vision Patt. Recogn.* (San Diego), June 1989, pp. 110-116.
- [29] P. Saint-Marc, J. S. Chen, and G. Medioni, "Adaptive smoothing: A general tool for early vision," in *Proc. Conf. Comput. Vision Patt. Recogn.* (San Diego), June 1989, pp. 618-624.



**Steven D. Cochran** (S'85-M'91) was born in Burbank, CA. He received the B.A. degree (cum laude) in mathematics and the M.S. degree in computer science from the California State University at Northridge in 1977 and 1980, respectively, and the Ph.D. degree in computer engineering from the University of Southern California, Los Angeles, in 1990.

From 1980 to 1988, he was a project engineer at TRW, where he worked in software engineering, and from 1988 to 1990, he worked as a research assistant at the Institute for Robotics and Intelligent Systems at the University of Southern California. He is presently a research associate in the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, where his research interests are machine vision, aerial image understanding, and man-made interface design.

Dr. Cochran is a member of the American Association for the Advancement of Science and the Association for Computing Machinery, as well as Sigma Pi Sigma and Blue Key.



**Gérard Medioni** (S'83-M'83-SM'92) received the Diplôme d'Ingenieur Civil from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1977, and the M.S. and Ph.D. degrees in computer science from the University of Southern California (USC), Los Angeles, in 1980 and 1983, respectively.

Since then, he has been an Assistant Professor of Computer Science with a joint appointment in Electrical Engineering at USC. His research interests include computer vision, artificial intelligence, and robotics.

Dr. Medioni served as Program Co-Chair of the 1991 IEEE Conference on Computer Vision and Pattern Recognition in Hawaii. He is a member of the American Association for the Advancement of Science and the Association for Computing Machinery.