Direct Evidence for Occlusion in Stereo and Motion

James J. Little
University of British Columbia
Vancouver, BC, Canada V6T 1W5

Walter E. Gillett
MIT
Cambridge, MA, USA 02139

Abstract: Discontinuities of surface properties are the most important locations in a scene; they are crucial for segmentation because they often coincide with object boundaries. Standard approaches to discontinuity detection decouple detection of disparity discontinuities from disparity computation. We have developed techniques for locating disparity discontinuities using information internal to the stereo algorithm of [2], rather than by post-processing the stereo data. The algorithm determines displacements by maximizing the sum, at overlapping small regions, of local comparisons. The detection methods are motivated by analysis of the geometry of matching and occlusion and the fact that detection is not just a pointwise decision. Our methods can be used in combination to produce robust performance. This research is part of a project to build a "Vision Machine" [7] at MIT that integrates outputs from early vision modules. Our techniques have been extensively tested on real images. [1]

1 Introduction

This investigation describes a component of the MIT Vision Machine [7], which integrates outputs from early vision modules for tasks such as recognition and navigation. The integration stage computes maps of scene properties augmented by an explicit representation of scene discontinuities, identifying their physical origin. Our major achievement in this paper is the development of techniques for locating disparity discontinuities using information internal to the stereo and motion modules, rather than by post-processing the output. Later processing to detect discontinuities [6] can then operate with substantially more information about their location. We have devised techniques for discontinuity location, based on an analysis of patchwise matching scores internal to the algorithm, and based on the effects of occlusion.

Stereo and motion both compute similar quantities – image displacements of image elements. We use a dense set of overlapping matching operators to compute displacements between the two images. Both stereo and motion apply uniqueness and continuity constraints. Scene geometries differ, however, and so do interpretations of ordering constraints.

1.1 The parallel stereo algorithm

The Drumheller-Poggio algorithm [2] served as an experimental testbed for the research described here. Stereo matching is an ill-posed problem [1] that cannot be solved without taking advantage of natural constraints. The continuity constraint asserts that the world consists primarily of piecewise smooth surfaces. If the scene contains no transparent objects, then there can be only one match along the left or right lines of sight (uniqueness). The ordering constraint [11] states that any two points must be imaged in the same relative order in the left and right eyes.

The specific assumption used is that the disparity of the surface is locally constant in a small region surrounding a pixel. It is restrictive, but may often be a satisfactory local approximation (it can be extended to more general surface assumptions in a straightforward way but at high computational cost). Let $E_L(x,y)$ and $E_R(x,y)$ represent the left and right image of a stereo pair or some transformation of the images. We look for a discrete disparity $d(x,y)$ at each location (x,y) in the image that minimizes

$$||E_L(x,y) - E_R(x+d(x,y),y)||_{N_{(x,y)}} \tag{1}$$

where the norm is a summation over a local neighborhood $N(x,y)$ centered at each location (x,y); $d(x,y)$ is assumed constant in the neighborhood. The algorithm actually implemented is somewhat more complicated, since it involves geometric constraints (ordering and uniqueness) that affect the way the maximum operation is performed (see [2]). The algorithm is composed of the following steps:

[1]This report describes research done within the Artificial Intelligence Lab at MIT as well as at UBC, supported, at MIT, by DARPA under Army contract DACA76-85-C-0010 and in part under ONR contract N00014-85-K-0124. Primary support for Gillett came from NIGMS Training Grant T32-GM07484, under by the MIT Dept. of Brain and Cognitive Sciences. This research was also supported by NSF Contract No. MIP-8814612, and by a grant from the Natural Sciences and Engineering Research Council of Canada.

1. Compute features for matching (edge detection or band-pass filtering).
2. Compute matches scores between features.
3. Determine the degree of continuity around each potential match.
4. Identify disparities based on the constraints of continuity, uniqueness, and ordering.

Potential matches between features are computed as follows. The images are registered so that the epipolar lines are horizontal. We compute match score planes, one for each horizontal disparity. Let $p(x,y,d)$ denote the value of the (x,y) entry of the match score plane at disparity d. For edge-based tokens, the results of comparison are binary. We set $p(x,y,d) = 1$ if there is a token at (x,y) in the left image and a compatible token at $(x-d,y)$ in the right image; otherwise set $p(x,y,d) = 0$. For brightness-based matching, the matching score continuously varies (E_L and E_R vary over some finite range and the norm of their difference assumes a range of values, not just 0 and 1 - see Equation 1).

The value computed by Equation 1 measures the degree of continuity around each potential match at (x,y,d). For edge-based matching, the summation counts the "votes" for the disparity d in the d^{th} match plane. If the continuity constraint is satisfied near (x,y,d) then $N(x,y)$ will contain many votes and the score $s(x,y,d)$ will be high (see Equation 1). We mostly will discuss the edge-based methods in stereo and therefore will maximize the normalized correlation and will speak of peaks in the measured values. Finally, we select the correct matches by applying the uniqueness and ordering constraints. Under the uniqueness constraint, a match suppresses all other matches along the left and right lines of sight with weaker scores. To enforce the ordering constraint, if two matches are not imaged in the same relative order in left and right views, we discard the match with the smaller support score. Each match suppresses matches with lower scores in its forbidden zone [11][8] (see Section 2.2).

The matching scores of the stereo algorithm are valuable information. They provide a confidence level for each match that can discriminate between competing matches, as in forbidden zone suppression (using the ordering constraint). Matching scores are computed everywhere with no additional computation (because of homogeneous computation in SIMD machines), both for edge-base and brightness-based matching, producing dense information. The scores also help to suppress bad matches within occluded areas of the scene (Section 2.2).

2 Disparity discontinuities

We describe two discontinuity detection techniques, arising from analysis of the behavior of matching methods near occluding boundaries. One method is based on an analysis of matching scores for different disparities and the other arises from the effects of geometric constraints near occlusions.

2.1 Close winners

The close winners technique analyses matching scores. For each point $p = (x,y)$ in the left image and $q = (x+d,y)$ in the right image, the matcher computes a score $s(x,y,d)$ indicating the likelihood that p matches q, i.e., that p and q are images of the same physical point in the scene. The score at a point, $s(x,y,d)$, is the sum of pointwise match scores in a region $N(x,y)$ (see Equation 1). The matcher examines only disparities in the fixed interval $[id, fd]$, where id and fd are the initial and final disparities. Define the score vector $v(p) = \{s(id), s(id+1), ..., s(fd)\}$, the sequence of matching scores for point p.

We begin with a simple example, a random-dot stereogram (RDS) which fuses to yield the impression of a 192×192 square floating in front of the background (256×256). Figure 1 shows a schematic representation of the scene; the dark strip on the left-hand side is an occluded part of the background seen in the left view but not the right. Point B is located on the boundary of the square. The local support neighborhood of point B, N_B, is divided between the square and the background. Approximately half of the edges in N_B will vote for the wrong disparity, namely the background disparity. The graph of $v(B)$ is bimodal, with one peak at the foreground disparity and another peak at the background disparity. Spoerri and Ullman[9] use a similar observation to derive a different scheme for motion segmentation. Matching scores vary over possible disparities (displacements) and will be maximal at the two displacements of the foreground and background. In contrast, $v(A)$ and $v(C)$ are unimodal, since their support regions cover constant disparity regions. Figure 2 shows score vectors computed for the RDS - high scores represent best matches. It is critical that the diameter of the support region be larger than the largest disparity gap in the image

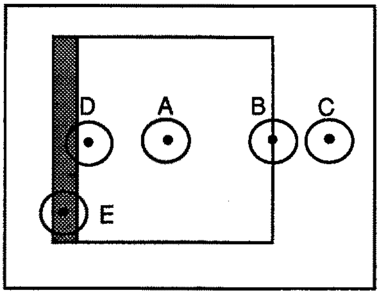
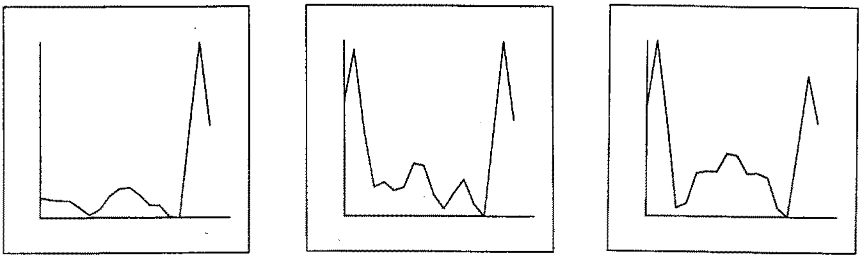


Figure 1: Line drawing of scene: floating square.



(a) (b) (c)

Figure 2: Score vectors for RDS. (a) A: (128,128). (b) B: (192,128). (c) D: (85,128).

– else the two peaks will not be detected using close winners. Also, the maximum value for the match score at B will at most be *half* that of the score at points such as A and C; this leads to a method for discontinuity identification using local spatial extrema of the match score (see [4]).

We call point B a close winner because the "winning" disparity has a close competitor; such points are likely to be located at disparity discontinuities. For all points p in the left image, use the following procedure to determine whether p is a close winner:

1. Identify peaks in $v(p) = \{s_{id}, s_{id+1}, \ldots s_{fd}\}$.
2. If $v(p)$ has two or more peaks, pick the two largest, α and β, $\alpha \geq \beta$. Let the margin $m = (\alpha - \beta)/\alpha$. If $m \leq M$ (0.2 for the results here), then p is a close winner.

Figure 3 shows close winners for several stereo scenes. Note that close winners can be correctly located for point B, but for point E, they identify locations in the center of the occluded area. These can be corrected by using a symmetric matching scheme, combined with mapping close winners into a common coordinate system.

2.2 Suppression Using Ordering Constraint

When one surface lies in front of another, the foreground surface occludes a portion of the background surface. The location of the occluded region depends on the viewpoint. Since the boundary on the near side of an occluded region is the discontinuity contour, identifying an occluded region leads us directly to the associated disparity discontinuity. This technique can be used to locate any disparity discontinuity except extended horizontal boundaries, which are not associated with occlusion.

Let us consider right-occluded areas, i.e., areas visible from the left but not the right view (see Figure 4b). Such an area does not have a match in the right image. Every potential match is surrounded by an hourglass-shaped region extending through the d and x dimensions, the forbidden zone (see [11]), as pictured in Figure 4a.

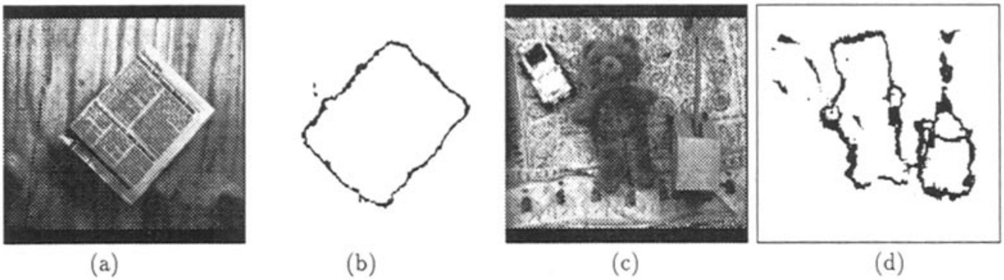


Figure 3: Close winners. (a) Newspaper on wood: left view. (b) Close winners. (c) Left view of truck, teddy bear, and crane. (d) Close winners for teddy.

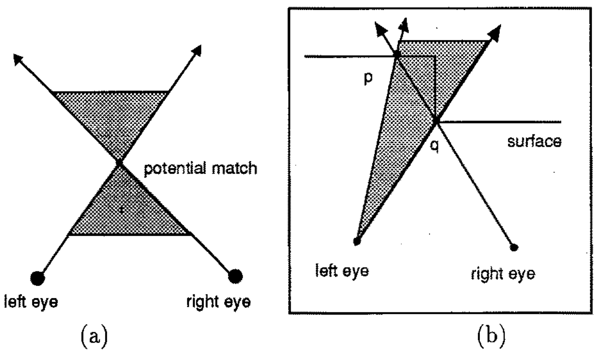


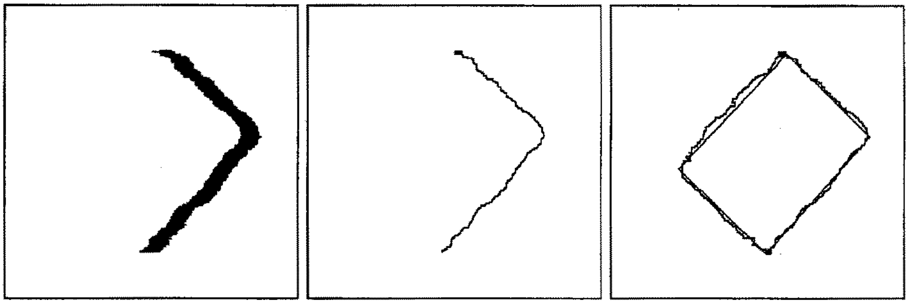
Figure 4: (a) The forbidden zone (shaded) for a particular potential match. (b) The shaded region is contained within the union of the forbidden zones for points p and q, showing that no match will be permitted there.

Consider a simple step discontinuity (Figure 4b) where the portion of the surface between points p and q is right-occluded. The shaded region contains all points that are imaged between p and q in the left view. Note that the shaded region is contained entirely within the union of the forbidden zones for p and q: the area above the line joining p and q is in the forbidden zone for q, and the area below the line is in the forbidden zone for p. Therefore all possible matches in the left view between the images of p and q will be suppressed. Match suppression is the key to locating occluded areas.

2.2.1 The Mechanics of Match Suppression

While computing matches and applying the ordering constraint, we can keep track of suppressed matches. Since matching scores are computed at all points, stereo produces dense suppression of competing matches at occlusions.

A point (x, y) in the left image is suppressed if, for all disparities d, the potential match at (x, y, d) has been suppressed. Suppressed points collectively determine regions of suppression that correspond to right-occluded areas. Disparity discontinuities are points on the right-hand side of suppressed regions, the near side in the case of right-occlusion. Others [10] have noted the connection between matching and identification of occlusions, but do not tie it in to the full ordering constraint. Figure 5 shows the suppressed regions for the newspaper scene. Some suppressed points are part of significant occluded regions and others result from incorrect matches or disparity quantization effects. As a simple measure to select significant regions, we threshold the width of contiguous strips of suppressed points. Figure 5 shows the suppressed regions and filtered suppressed regions for the left-occluded regions of the newspaper-on-wood scene. Left-occluded areas (visible in right image



(a) (b) (c)

Figure 5: Left-occluded regions for newspaper. (a) Filtered suppressed points (left-occlusion). (b) Merged discontinuities (left/right). (c) Merged discontinuities over newspaper silhouette.

not from left image) are found by a symmetric analysis of the right image. The associated disparity discontinuities then lie on the left-hand side of the occlusion, and are mapped from the right into the left image, using the disparity value. The results of the analysis are shown in Figure 5.

Finally, there is an additional benefit of identifying occluded areas. Knowledge of occlusion can improve naive interpolation. Interpolation blurs discontinuities, filling in occluded areas with depth data from both sides. A better approach assumes that an occluded area has the same disparity as the background, e.g., filling in right-occluded regions with disparity values from left to right [3].

3 Conclusion

We have addressed the detection of discontinuities in stereo and motion, within the context of efficient, parallel implementation. The techniques we have examined all use information internal to the correspondence process to identify discontinuities. Any later processing to determine the figure/ground relation and to improve surface description begins with an almost complete description of the location of discontinuities. Further, these techniques all can easily be implemented on a SIMD parallel computer and simple circuits. A detailed discussion may be found in [5].

References

[1] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proc. of the IEEE*, 76(8):869–889, Aug. 1988.

[2] M. Drumheller and T. Poggio. On parallel stereo. In *Proc. of IEEE Conf. on Rob. and Auto.*, pages 1439–1448, Washington, DC, 1986. IEEE.

[3] W. Gillett. Issues in parallel stereo matching. Master's thesis, Massachusetts Institute of Technology, 1988.

[4] J. J. Little, H. H. Bülthoff, and T. Poggio. Parallel optical flow using local voting. In *Proc. Int. Conf. on Comp. Vision*, pages 454–459, Tarpon Springs, Florida, Dec. 1988. IEEE, Washington, DC.

[5] J. J. Little and W. E. Gillett. Direct evidence of occlusion in stereo and motion. Tr-90-5, UBC Dept. of Computer Science, Vancouver, BC, 1990.

[6] T. Poggio, E. B. Gamble, and J. J. Little. Parallel integration of vision modules. *Science*, 242(4877):436–440, October 21 1988.

[7] T. Poggio and the staff. The MIT Vision Machine. In *Proc. Image Under. Work.*, Cambridge, MA, April 1988. Morgan Kaufmann, San Mateo, CA.

[8] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. Disparity gradients and stereo correspondences. *Perception*, 1987.

[9] A. Spoerri and S. Ullman. The early detection of motion boundaries. In *Proc. Int. Conf. on Comp. Vision*, pages 209–218, London, England, June 1987. IEEE, Washington, DC.

[10] J. Weng, N. Ahuja, and T. S. Huang. Two-view matching. In *Proc. Int. Conf. on Comp. Vision*, pages 64–73, Tarpon Springs, Florida, Dec. 1988. IEEE, Washington, DC.

[11] A. L. Yuille and T. Poggio. A generalized ordering constraint for stereo correspondence. A.I. Memo No. 777, Art. Intell. Lab, MIT, 1984.