Accurate Early Detection of Discontinuities

James J. Little
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada V6T 1Z2

Abstract

Displacement must be determined in both stereo and optical flow computation. Correlation or Sum of Differences can produce robust displacement estimates. To improve its performance near occlusions and disocclusions, we construct new estimators from a priori analysis. We introduce smoothly decaying summation functions into Sum of Differences operators to improve their reliability. Moreover, we show how occlusions can be simply detected by comparing the outputs of support windows of differing size and orientation. We report the behavior of the system on real and synthetic images. Displacement must be determined in both stereo and optical flow computation. Correlation or Sum of Differences can produce robust displacement estimates. To improve its performance near occlusions and disocclusions, we construct new estimators from a priori analysis. We introduce smoothly decaying summation functions into Sum of Differences operators to improve their reliability. Moreover, we show how occlusions can be simply detected by comparing the outputs of support windows of differing size and orientation. We report the behavior of the system on real and synthetic images.

1 Introduction

This investigation describes a component to be used in a vision system that integrates outputs of early vision modules for tasks such as recognition and navigation. The integration stage computes maps of scene properties augmented by an explicit representation of discontinuities in the scene, identifying their physical origin. We develop techniques for locating displacement discontinuities using information internal to the stereo and motion modules, rather than by post-processing the output. Later processing to detect discontinuities [11] can then operate with substantially more information about their location.

The first stage determines an estimate of displacement. This is accomplished by a displacement (stereo or motion) algorithm using correlation or sum of squared differences (SSD). The output of that stage is displacements, confidence measures and a local estimate of the location of boundaries. This provides orientation of local detectors which aid in analysis of the next frame in an image sequence.

The system finds the most suitable collection of oriented filters to refine motion/stereo boundary detection.

2 Displacement and discontinuities

Window size and shape is a critical problem in computing stereo or motion displacements by correlation, or, more properly, sum of squared differences (SSD). [7, 1, 2, 10, 6, 8, 4]. If the window is too large, the displacement may vary within the window. On the other hand, a small window may not cover enough brightness variation to permit correct detection. Levine *et al.*[7] described a scheme for varying window size when the brightness variation is not sufficient. Kanade and Okutomi [6] show a method that, based on a statistical model of brightness and disparity variation, adaptively alters the window size.

We are motivated by considerations of biological explanation and computational efficiency to consider algorithms that begin with an assembly of several windows of different sizes and orientations.

Typical systems for displacement identification by SSD suffer from two defects, aside from window size determination: first, the window size is chosen small for efficiency, leading to ambiguity, and, second, the summation function is a simple rectangular windowing function that has poor behavior, introducing aliasing effects.

We report on a method for configuring such systems, an engineered system, driven by considerations of simplicity. The interaction among windows is specified by rules derived from weak a priori assumptions. We report the behavior of the system on real and synthetic images.

The engineered system includes a simple smooth windowing function for summation that improves results dramatically. Experiments with several windowing functions shows that any of several smoothly decaying centrally weighted windowing function improves both detection and localization. We test detection and localization by counting errors on synthetic images.

3 Engineered Detectors

The feature images input to an SSD system consist of simple difference of Gaussian processing. We base this decision not only on previous experiments[5] but

*This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Networks of Centres of Excellence Institute for Robotics and Intelligent Systems, Project A-1.

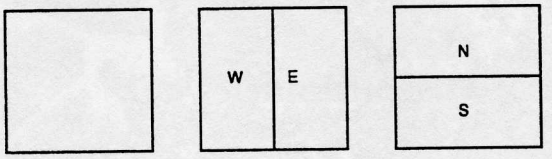


Figure 1: Half neighborhoods: a support region and sub-regions

also on the nature of cross-correlation[12]; we choose the sum of absolute values of differences (SAD). This has several benefits, first, that the magnitude of intermediate results is comparable to the magnitude of the inputs and, second, the effect of outliers is reduced, increasing the reliability of the detector near occlusions.

In previous work[8], we determined several schemes for detecting occlusions directly from the output of the stereo/motion detection stage. One of the schemes involves varying the size and shape of detectors.

The output of the stereo and motion modules depends on the size of the support neighborhood. When a depth or motion discontinuity bisects the support region, the results are less reliable, but can be analysed to identify occlusions. By using a set of smaller support regions that divide a support region (see Figure 1), we can get the response of the matching module over the range of supports. At a discontinuity, the set of detectors should give different outputs, and otherwise should agree.

Increasing the window size, when the contrast range in the window is small, increases total sum of differences. Nevertheless, a large summation can eliminate ambiguity when even there is only small image variation.

To increase the accuracy of flow estimation, we treat the images symmetrically, matching from left to right and right to left for stereo. Points at which the flows do not agree, to some tolerance, are marked as indeterminate. Similar methods are shown in [9, 5, 4]. Symmetric matching is a simplification of the geometric imaging constraints [15]. The original Marr-Poggio stereo algorithm [9] and a parallel implementation [3] with many similar geometric features use two lines-of-sight and optionally the full forbidden zone (FFZ) to determine support for matching. A selected match must the better than all others in its FFZ, or just the two lines of sight. To simplify matching we write results into two arrays simultaneously and then compare final results. This symmetry is also appealing from a biological standpoint. Eliminating mismatches (points not symmetrically matched) reduces the number of errors significantly.

The support windows vary in size, shape and orientation. They must compete with each other to determine the correct displacement at a pixel. Window summation for all summation filters are normalized to unity. The windows compete, and the window with smallest error wins. In the composite result, detection of discontinuities improves significantly.

The details of the SSD and SAD implementations vary, although most have used a box filter for summation of the difference terms. The box filter introduces unnecessary ringing into the system, so from signal considerations alone one should use a windowing (summation) function that decays smoothly to zero at the boundary of the box [12]. Other considerations, namely regularization, suggest that the local weighting function should decay smoothly [14].

Several such filters were implemented and tested both on real and synthetic images. The following section describes the results of the tests.

4 Experiments

We use a random dot stereogram for synthetic image experiments. A series of square regions are arranged at displacements of zero, five and ten pixels, as in a wedding cake. The box filter extends to $n = 7$ pixels and the Gaussian filter uses an equivalent $\sigma = 3$.

Several points can be made about the results. First, in every case, regardless of the filter shape, the composite result has a much lower error rate, approximately one third that of the symmetric filter. We use several support functions, including a half Gaussian (a Gaussian whose output is zero for negative x) and a half box. They have comparable results (less than 0.2 percent errors on synthetic images). The following table lists the numbers of erroneous points after pruning for mismatched points, for a variety of standard windowing or summation functions [12]. In each case, the 2d filter is generated for the appropriate windowing function and used for summation. The asymmetric or oriented summation filters are created by zeroing half of the filter coefficients in the appropriate half plane. The composite result (joint) uses the displacement for the window with least SAD.

Window	Joint	Isotropic	Matched	% wrong
Gaussian	92	286	48925	0.188
Box	94	454	48946	0.192
Rayleigh	375	1803	—	—
Blackman	341	819	—	—
Hamming	294	606	—	—
Cosine	288	554	—	—

Table 1: Errors on cake example for various summation filters.

These effects are noticeable on the synthetic images chiefly at the occluding edges of the objects. Each figure shows the output of the symmetric filter, the results of the competition among the summation regions and finally the orientation of the summation filter chosen at each point. The choice is either the central (C) isotropic filter or one of the subneighborhoods as depicted in Fig. 1. Figure 2 shows the output where the filter has a Gaussian profile, decaying smoothly, while Figure 4 shows the output where the filter is simply a box. Figure 4 shows the results of the competition between the oriented filters: there are

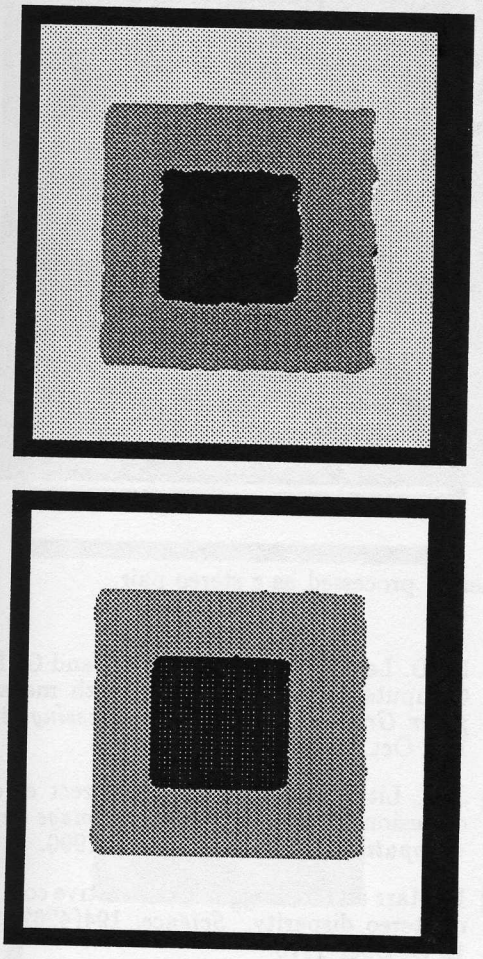


Figure 2: (a) Gaussian summation — circularly symmetric operator. (b) Asymmetric operator.

five brightness categories according to the filter with least error, including C=center or isotropic, N=north, E=east, S=south, and W=west.

Figure 4 also shows the points in the right image at which symmetric matching fails. Each point in the right image follows its displacement vector into the left image, and then follows the displacement vector found there back into the right image. When the trip is not a round trip, symmetry fails. Formally, these are points p_R where the displacement to the right image from the left $V_R(p_R)$ is such that

$$V_R(p_R) \neq -V_L(p_R + V_R(p_R))$$

These are either points where the image content is ambiguous or which are visible in the right image but not in the left.

4.1 Test on Real Images

We also used images (256 × 233), taken by a translating robot, of an outdoor scene containing trees (Figure 5). The results of the symmetric, asymmetric and

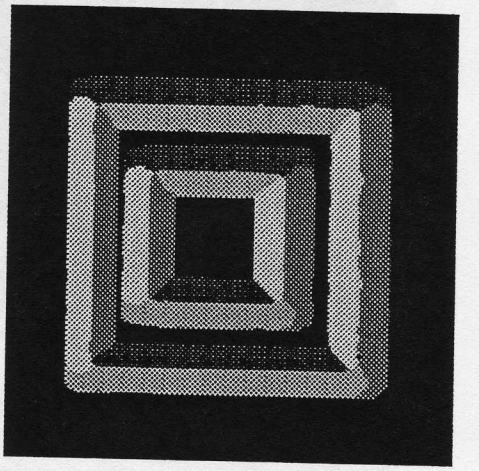


Figure 3: Gaussian summation — choice output (C=0,N=50,E=100,S=150,W=200).

the choice output are shown in Figure 6. Where symmetric matching fails is shown in black, overlaid on the disparity data. There is no interpolation in this example so the results appear quantized in the figure.

We include subpixel interpolation from the SAD data. When the two minimal SAD values represent adjacent displacements, d_1 and d_2, the optimum value is assumed to lie between the two and its location is estimated by weighted interpolation $d = (w_1 * d_2 + w_2 * d_1)/(d_1 + d_2)$, where w_i is the SAD value at the i^{th} point. Both examples use Gaussian summation; Fig. 6 does not use interpolation, while Fig. 7 does. The scene structure becomes more visible with interpolation; note the tree in the midground on the left that appears only in the interpolated data. The choice Figure 6 shows that points where either east or west filters dominate lie along the boundaries of the trees. North and south filters dominate in the ground plane where disparity variation is primarily vertical. igure 8, output of the box filter summator, shows that the detection properties of the box filter are more stable, but relatively poor at localization.

These experiments only show the results for the four summations (NESW); other experiments in progress show promising, improved results for 8 summation operators including diagonally oriented filters. So far, we have used asymmetric summation filters composed of a step edge convolved with a symmetric filter. Extensions to filters where the parameter along the step boundary varies from the across boundary parameter appear promising and preliminary results are encouraging. Ralph[13] has investigated learning these detectors from examples, with significant improvements comparable to those reported here.

5 Conclusion

Correlation stereo and motion, or more properly, SAD stereo and motion, is a robust technique for de-

Figure 5: Images from translating robot sequence, processed as a stereo pair.

tecting displacements. However, SAD stereo can be improved near occluding boundaries. Symmetric support regions are subdivided into asymmetric support regions competing to predict the output. Experiments with a simple set of four overlapping subregions show significant improvements, both in synthetic images and real images. In addition, we show how the use of a smoothly decaying windowing function (the support function) improves both localization and detection.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[2] H. Bulthoff, J. J. Little, and T. Poggio. A parallel algorithm for real-time computation of optical flow. *Nature*, 337:549–553, February 1989.

[3] M. Drumheller and T. Poggio. On parallel stereo. In *Proceedings of IEEE Conference on Robotics and Automation*, pages 1439–1448, Washington, DC, 1986. Proceedings of the IEEE.

[4] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. Technical Report 1369, INRIA, Jan. 1991.

[5] J. M. Hakkarainen, J. J. Little, H. Lee, and J. W. Jr. Interaction of algorithm and implementation for analog VLSI stereo vision. In *Proceedings of 1991 SPIE Conference on Visual Information Processing: From Neurons to Chips,*, Apr. 1991.

[6] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. CMU-CS-90-120, Carnegie Mellon University, Apr. 1990.

[7] M. D. Levine, D. A. O'Handley, and G. M. Yagi. Computer determination of depth maps. *Computer Graphics and Image Processing*, 2(4):131-150, Oct. 1973.

[8] J. J. Little and W. Gillett. Direct evidence of occlusion in stereo and motion. *Image and Vision Computing*, 8(4):328–340, Nov. 1990.

[9] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194(4262):283–287, 15 October 1976.

[10] L. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209–236, 1989.

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

[12] W. K. Pratt. *Digital Image Processing (2nd edition)*. John Wiley & Sons, New York, NY, 1991.

[13] S. Ralph. A neural network implementation for integrating discontinuity and displacement information. Master's thesis, The University of British Columbia, Vancouver, BC, 1991. M.Sc., supervisor J.J. Little.

[14] A. L. Yuille and N. M. Grzywacz. A computational theory for the perception of coherent visual motion. *Nature*, 333:71–74, May 1988.

[15] A. L. Yuille and T. Poggio. A generalized ordering constraint for stereo correspondence. AI-Memo-777, MIT AI Laboratory, Cambridge, MA, 1984.

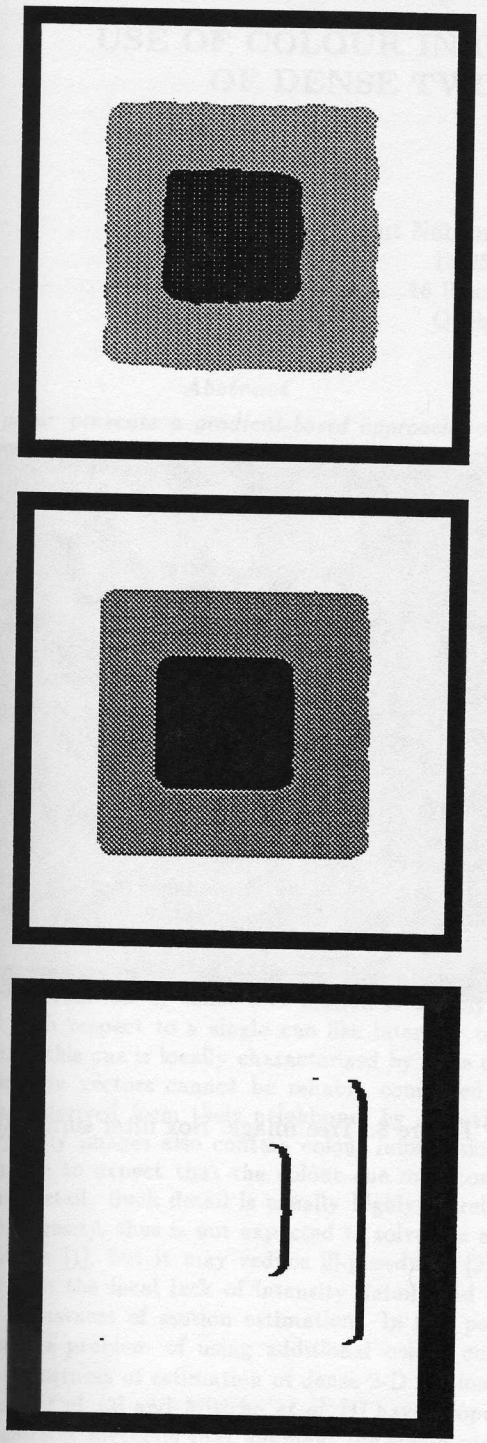


Figure 4: (a) Box summation — square summation. (b) Box with subsets. (c) Points where symmetric matching fails.

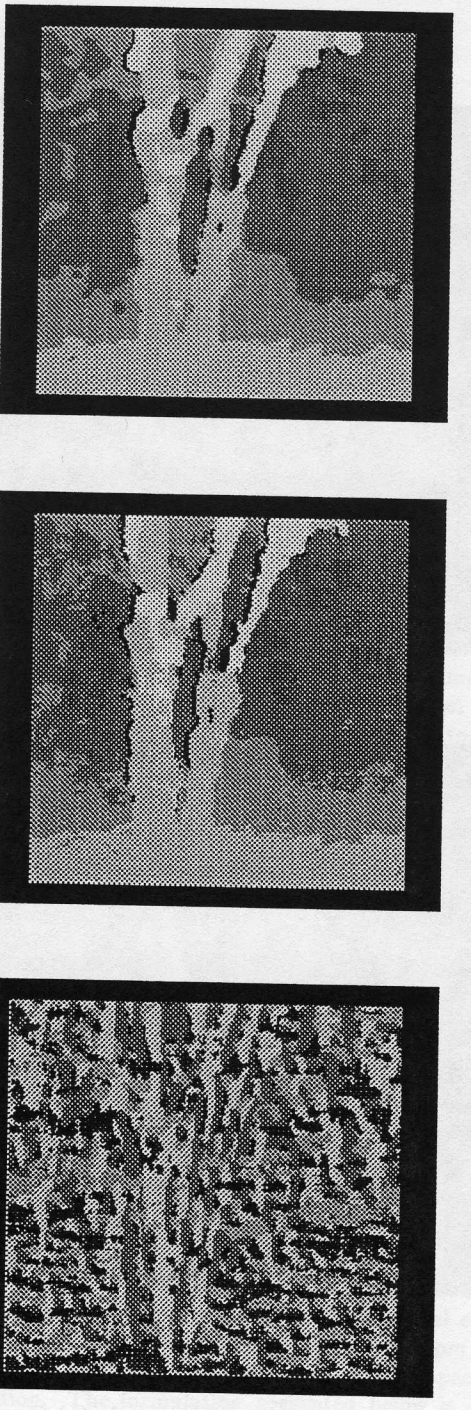
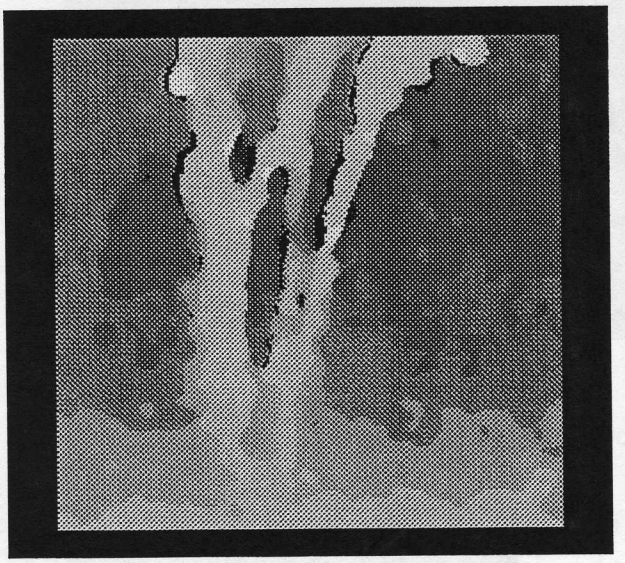


Figure 6: Tree image. (a) Gaussian summation — circularly symmetric operator. (b) Asymmetric operator. (c) Choice output.



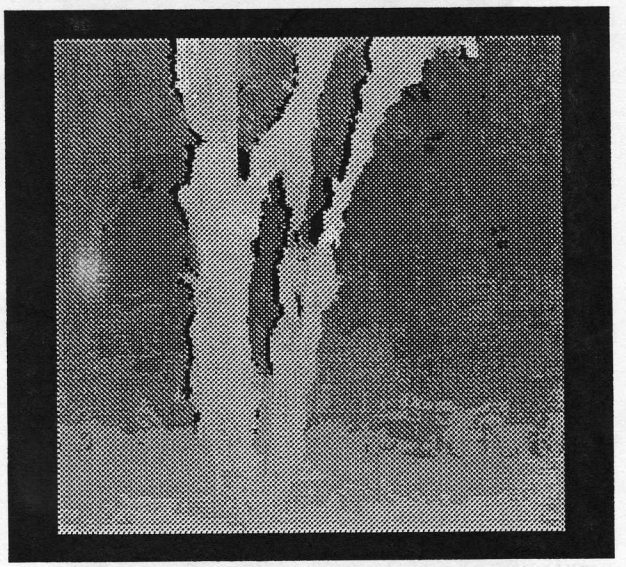


Figure 7: Tree image, interpolated. (a) Gaussian summation — circularly symmetric operator. (b) Asymmetric operator.

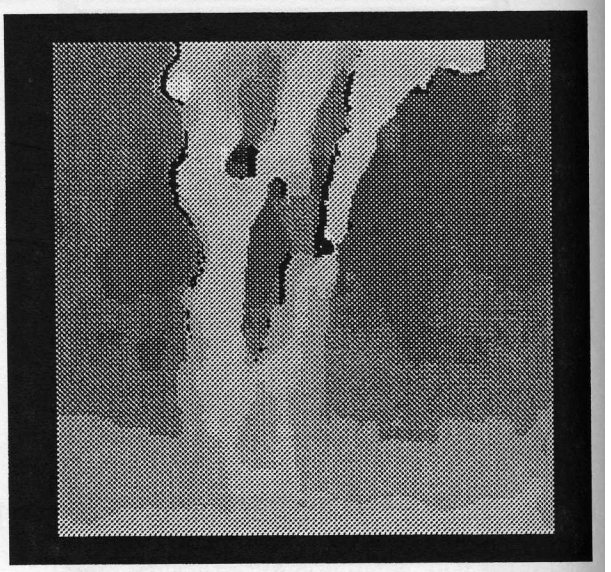


Figure 8: Tree image. Box filter summation.