Proceedings of the National Conference on Artificial Intelligence,

AAAI - 90, Boston, Mass, 1990

Constraints for the Early Detection of Discontinuity from Motion

Michael J. Black and **P. Anandan**
Department of Computer Science
Yale University
New Haven, CT 06520-2158

Abstract

Surface discontinuities are detected in a sequence of images by exploiting physical constraints at early stages in the processing of visual motion. To achieve accurate early discontinuity detection we exploit five physical constraints on the presence of discontinuities: *i)* the shape of the sum of squared differences (SSD) error surface in the presence of surface discontinuities; *ii)* the change in the shape of the SSD surface due to relative surface motion; *iii)* distribution of optic flow in a neighborhood of a discontinuity; *iv)* spatial consistency of discontinuities; *v)* temporal consistency of discontinuities. The constraints are described, and experimental results on sequences of real and synthetic images are presented. The work has applications in the recovery of environmental structure from motion and in the generation of dense optic flow fields.

Introduction

The relative motion of surfaces can provide information about the presence of surface discontinuities. We detect these discontinuities over time by exploiting physical constraints at early levels in the processing of visual motion. As noted by Marr (Marr 1982), the human visual system efficiently detects object boundaries using only relative surface motion as a cue. For example, figure 1 shows one image in a random dot sequence, in which a square patch is translating with respect to a stationary background. Human observers easily detecting the boundary of the square when presented with these images in sequence even when noise is added. Thus lines of discontinuity can provide evidence about the presence of surface discontinuities and the structure of the environment.

In this paper five physical constraints on the presence of discontinuities will be explored. We will describe how these constraints can be exploited to detect lines of discontinuity from a sequence of densely sampled images.

The first constraint is derived from the observation that multiple surfaces in relative motion will have different best displacements between a pair of images. We use the standard *sum of squared differences* (SSD) correlation measure for computing displacements. In the presence of surface discontinuities, the shape of the SSD surface provides information about the number, and relative motion of, the surfaces present (Anandan 1987). In the simplest case, the surface is *multi-modal* with local minima corresponding to the motion of the surfaces at the discontinuity.

While a multi-modal error surface indicates the presence of a discontinuity, the aperture problem means that the *absence* of a multi-modal surface does not guarantee that no discontinuity is present. A second constraint uses information about how the intensity structure in an area of the image changes with motion. This change is measured by comparing the SSD surface obtained under motion with a translation invariant *auto-correlation* surface which would result if no motion were present. If a discontinuity is present, the shape of the surface will change with motion, while uniform motion will result in a surface with similar shape.

A third constraint (*neighborhood flow modality*) exploits the fact that if multiple surfaces are moving relative to each other, then the optic flow will be different for each surface. This, in turn, will be reflected in a histogram of flow vectors within a neighborhood; the histogram will be multi-modal (Spoerri & Ullman 1987). Naively applied, this approach has serious limitations but by using confidence measures (Anandan 1989) associated with the flow field we can effectively exploit this constraint.

The fourth constraint exploits the *continuity of discontinuities* (Marr 1982). Discontinuities correspond to surface boundaries in the world, and hence it is reasonable to assume that such boundaries have spatial extent. A *spatial consistency* constraint is developed using *controlled continuity splines* (Kass, Witkin, & Terzopoulos 1987).

Finally, assuming a fairly stable environment, discontinuities tend to persist over time and move continuously across the image plane. A *temporal consistency* constraint is developed using active contour models in

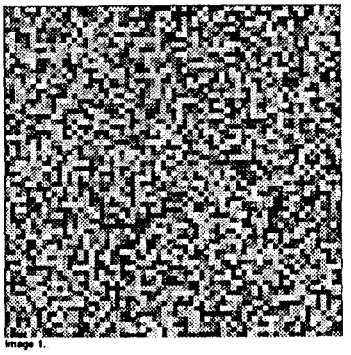


Figure 1: **Random Dot Image.** One image from a motion pair in which a foreground patch is undergoing a 2 pixel displacement with respect to a stationary background. Zero mean gaussian noise with a standard deviation equal to 10 percent of the standard deviation of the image has been added to the second image.

which an energy term constrains the motion of the contour to be consistent with image flow.

Motivation and Previous Work

This work has two main motivations. First, one of the primary goals of computer vision is to recover the structure of the environment; surface discontinuities provide a great deal of structural information. Traditional edge detection techniques have well known limitations for boundary detection. They may fail to detect boundaries between textured surfaces, and detect many edges which do not correspond to structural properties of the environment but are artifacts of surface marking. Motion based discontinuity detection, may be able to be combined with edge detection schemes to produce more accurate and complete descriptions of the environment.

Secondly, we would like to incorporate information about discontinuities into the computation of dense optic flow fields over time (Schunk 1989) (Spoerri & Ullman 1987). One of the key observations underlying work in optic flow computation is the notion that surface discontinuities are not dense. Or phrased another way, that flow changes gradually across the field of view. This allows smoothness constraints to be introduced into the flow field computation which correct for noise in image correlation. This assumption of a smoothly changing flow field is violated at object boundaries, and hence the smoothness constraint is not appropriate across these boundaries. By explicitly computing surface discontinuities at early stages of motion processing, discontinuity information can be incorporated into the smoothness constraint to produce more accurate flow fields.

Most previous attempts at detecting discontinuities from motion have focused on an analysis of this flow field using region growing (Potter 1980) or edge detection techniques (Thompson, Mutch, & Berzins 1982). A variation on this approach (Mutch & Thompson 1988) computes *accretion/deletion* regions using correlation techniques. Another approach, which we will also exploit, uses information about the distribution of flow vectors in a neighborhood about a point to decide if a discontinuity is present (Spoerri & Ullman 1987).

These previous approaches suffer from two problems. First, they all ignore valuable information that is present in the correlation surface from which the flow is derived. Secondly, many occur too late in the flow computation; to work they must be applied to a smoothed flow field. When these techniques are applied to the raw, unsmoothed, flow field, the results are poor. There is a Catch-22: you need the information about discontinuities to derive an accurate smoothed flow field, and you need the smoothed flow field to detect the discontinuities.

Our approach is novel in that we develop constraints on the location of lines of discontinuity using information present in the SSD surface as well as physical properties of discontinuities to achieve robust early detection. While the constraints have intuitive appeal, and the experimental results are promising, we currently have no probabilistic justification for the confidence measures associated with these constraints and no probabilistic interpretation is implied. This is an area of ongoing research.

In the following sections each of the five constraints is developed in detail and illustrated with experimental results on synthetic data. We then present experimental results with a real motion pair. Before concluding, we discuss our current research directions.

Shape of the Error Surface

Correlation-based matching is a common technique used in the computation of optic flow (Anandan 1989). The approach is appealing for a variety of reasons; it is simple, it captures the intuitive notion of similarity between two image regions, and is inherently parallel. The sum of squared differences (SSD) is a common correlation measure which is computationally simple and performs well in empirical tests when applied to band-pass filtered images (Burt, Yen, & Xu 1982).

Given a point in an image and a set of points G in a neighborhood of size $n \times m$ around the point, we define the data error term for a displacement (u, v) of that point as:

$$E(u,v) = \sum_{i,j \in G} (I_1(i,j) - I_2(i+u, j+v))^2,$$

where I_1 and I_2 denote the intensity functions of two successive images. The SSD *surface*, S, is defined over the space of possible displacements (u, v) with the height of the surface corresponding to the data error, $E(u, v)$, of that displacement (figure 2 shows an example SSD surface at a corner point).

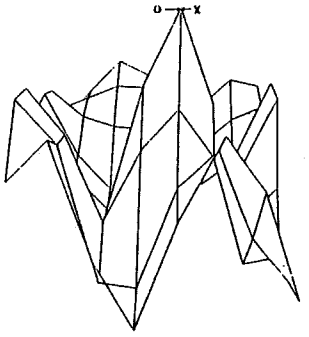


Figure 2: Example SSD Surface at a Corner (inverted for display).

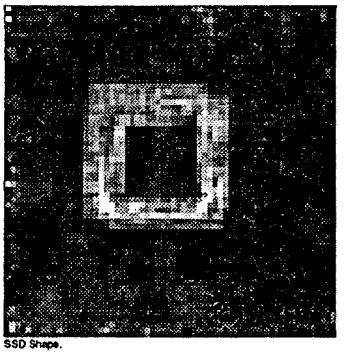


Figure 3: Confidence, C_S, based on shape of the SSD surface.

The shape of the SSD surface is typically quite complex and contains information about the motion of the surfaces that gave rise to it (Anandan 1984). In particular, in certain well defined cases, if there are multiple surfaces undergoing relative motion in a neighborhood, then they will each have different best displacements. This gives rise to a *multi-modal* error surface with minima corresponding to the displacements of each surface.

In an ideal situation, minima are easily detected by examining the first and second partial derivatives of the surface. Of course, when dealing with real imagery, detecting minima may not be so easy. In the presence of noise, true minima may be obscured and spurious minima may be introduced. Additionally, if the relative motion of the surfaces is small, then due to discretization, the peaks may merge together and be indistinct. In practice then, we must settle for a heuristic measure of *peakness*. One heuristic, ψ, takes into account the steepness of the peak, by measuring the distance of a point from its neighbors:

$$\psi(u,v) = \sum_{i=-1}^{1}\sum_{j=-1}^{1} S(u,v) - S(u+i,v+j).$$

The more negative ψ, the more likely a steep peak exists. Other measures of peak shape and steepness exist. For example, the scalar confidence measures of (Anandan 1984), which are based on normalized directional second derivatives of the surface, provide a measure of peakness based on curvature.

We desire an estimated confidence, C_S, that a particular image location corresponds to a discontinuity given the shape of the SSD surface at that location. Such a confidence measure should take into account the number of peaks present in the surface and some notion of how good these peaks are. We also take into account that, for some distance on either side of a discontinuity, the SSD surface may contain evidence of multiple motion. Our confidence measure should be highest at the actual boundary.

Experimental results with many confidence measures and heuristics indicate that simple measures, like the ratio of the depths of the two best peaks, perform nearly as well as more complex measures. If the first and second best peaks, as defined by ψ, have displacements (u_0, v_0) and (u_1, v_1) respectively, and $\rho_0 = S(u_0, v_0)$ and $\rho_1 = S(u_1, v_1)$ are the depths of the peaks, then we define C_S as:

$$C_S = \rho_0/\rho_1.$$

This function will have a global maximum approaching 1.0 at the actual boundary and will fall off as distance from the boundary increases. Figure 3 shows the values of C_S obtained from the SSD surface generated between the images described in figure 1. Bright values correspond to locations where there is a high confidence of a discontinuity.

An empirical study of the behavior of the SSD surface indicates that in areas of sufficient texture the surface contains enough information for discontinuity detection. However, if one or both of the surfaces present are homogeneous, the aperture problem prevents us from deriving meaningful information from the surface.

Weakening the Continuity Assumption

The SSD surface provides only approximate information about the displacement of multiple surfaces. It embodies the assumption that the intensity structure of a surface patch remains constant over time. This assumption generally holds for surfaces which are continuous but is violated at surface discontinuities.

When using the quadratic SSD measure, the presence of a poorly correlated surface introduces noise which influences the overall correlation. As the data error increases without bound so does the SSD measure. Instead, we desire a function which weights highly differences which fall within the expected range of error and remains uncommitted about data outside this range.

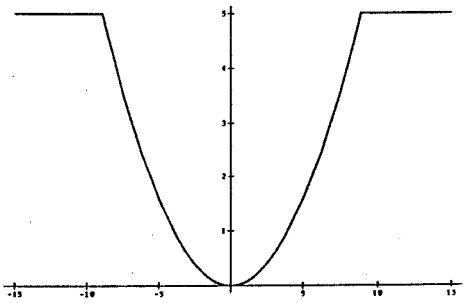


Figure 4: The ϕ function embodying the weak continuity constraint.

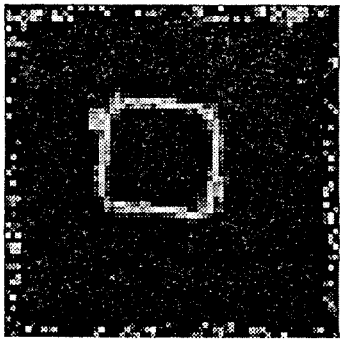


Figure 5: Confidence, C_S, using the weak continuity assumption.

This weakening of the SSD assumption corresponds to Blake and Zisserman's *weak continuity constraint* (Blake & Zisserman 1987). The following function, ϕ, has the desired properties:

$$\phi(x) = \begin{cases} \lambda^2 x^2 & \text{if } |x| < \sqrt{\alpha}/\lambda, \\ \alpha & \text{otherwise} \end{cases}$$

where λ and α are constants chosen with respect to the expected noise. The resulting data error term is a quadratic function of the difference in intensity values as long as the magnitude of the difference is below a threshold $\sqrt{\alpha}/\lambda$, and stabilizes to a fixed value α beyond the threshold (see figure 4).

This function weights well correlated points highly and diminishes the importance of poorly correlated points. If there are multiple surfaces in relative motion, there will be multiple displacements where a high number of points correlated well, and hence the correlation surface will contain multiple peaks.

The data error is now:

$$E(u,v) = \sum_{i,j \in G} \phi(I_1(i,j) - I_2(i+u, j+v)).$$

The error surface is generated as before and peaks are detected. Using the same confidence measure, C_S, as before we see that the area of possible discontinuity is more precisely located (figure 5).

Change in Surface Shape

As indicated in the previous section, the SSD surface may not have multiple peaks even when there are multiple surfaces in relative motion. In certain cases repetitive structures can cause multiple peaks in the SSD surface when only a single motion is present. Hence, we need a different approach to detect the absence of discontinuities. The key observation is that if an area is undergoing a uniform motion then the *cross-correlation* surface, S, between successive frames will have the same shape same as the *auto-correlation* surface, A, generated by correlating the the first image with itself (Anandan 1984).

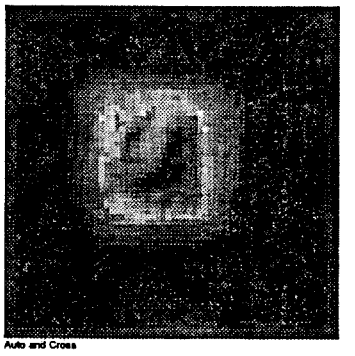


Figure 6: Confidence, $C_{S,A}$, based on the change between the auto and cross correlation surfaces.

Intuitively, if the cross-correlation surface is similar to the auto-correlation surface, given an appropriate translation, then the likelihood of a discontinuity is low. We define a confidence measure, $C_{S,A}$, based on this intuition. We translate the auto-correlation surface so that it is centered at the point of best match, (u,v), and compute the difference between the auto and cross-correlation surfaces:

$$C_{S,A} = \sum_{u=-m}^{m} \sum_{v=-n}^{n} (S(u,v) - A(u,v))^2.$$

This measure will be large at discontinuities and small in areas of consistent motion. Multiple peaks in the cross-correlation surface which are the result of repetitive structure will also appear in the auto-correlation surface and hence $C_{S,A}$ will be low. This measure is illustrated in figure 6 where $C_{S,A}$ is displayed for the sample random dot pair.

Neighborhood Flow

By taking the displacement of minimum error in the SSD surface, we arrive at a raw, unsmoothed, flow field, F. Each point in the field contains the best displacement of that point. Looking in a neighborhood around

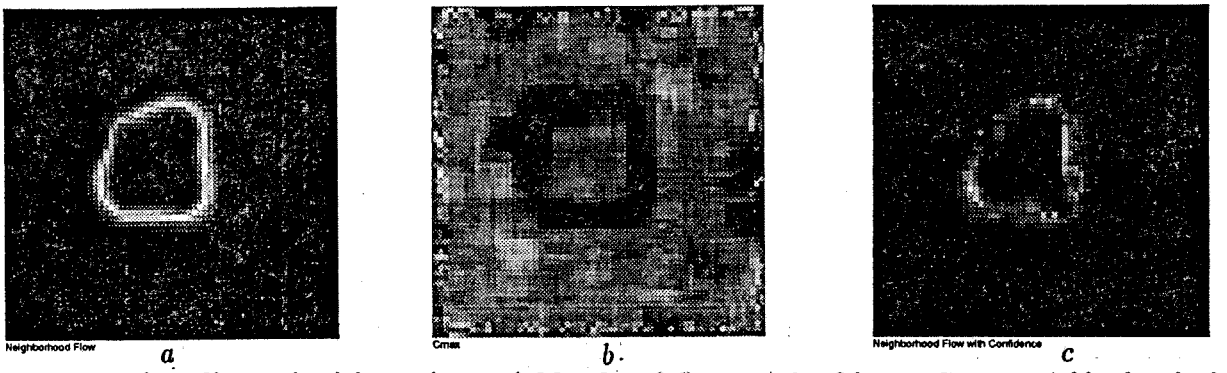


Figure 7: **Detecting discontinuities using neighborhood flow.** *a)* Confidence, C_F, on neighborhood of raw flow vectors. *b)* Confidence, C_{max}, in raw flow vectors. *c)* Confidence, $C_{F,C_{max}}$, combined neighborhood flow and flow confidence.

a given point in F, if there are multiple surfaces moving relative to each other, then there will be clusters of points with different flow vectors. A histogram of displacement vectors in a neighborhood will contain multiple peaks if a discontinuity is present (Spoerri & Ullman 1987). Peaks are detected in the histogram and a confidence measure, C_F, can be created by comparing the relative heights of the two highest peaks. At a boundary this measure has a global maximum approaching 1.0.

Since the traditional smoothness process blurs the distinction between neighboring flow vectors, the neighborhood flow constraint must be applied prior to smoothing. However, the unsmoothed flow field is usually noisy and error-prone, hence, the resulting histogram will itself be unreliable. This is illustrated by the example shown in Figure 7*a*, which is computed from the random dot test pair. The measure is maximum near the actual boundary but, due to noise in the unsmoothed flow field, it produces only a rough approximation to the boundary.

The solution our dilemma is contained in the use of confidence measures such as those described in (Anandan 1989). These provide a measure of confidence in a flow estimate based on the curvature of the SSD surface. Figure 7*b* shows the confidence in the optic flow estimates for the sample image pair. Areas where confidence in the flow estimate is low appear dark in the figure. During the computation of the histogram, we simply weight the contribution of each vector according to its associated confidence and find peaks in the histogram as before. Flow vectors near the discontinuity that are unreliable will contribute less than in the unweighted scheme. This approach cannot find a discontinuity if the information is not present in the flow field. Its usefulness is in reducing the confidence of spurious discontinuities which are the result of flow errors. The resulting confidence, $C_{F,C_{max}}$, for the test images is shown in figure 7*c*.; confidence in the erroneously located discontinuity at the occluding corner has been reduced.

Note that our simple scheme for detecting multiple peaks may fail due to the discretization of the histogram. For instance, two adjacent peaks in the histogram may simply be a broad single peak. More sophisticated clustering techniques may be needed to deal with such problems, and will be considered in the future.

Spatial Consistency

Until now we have only discussed the assignment of a confidence to a point in the optic array which has some likelihood of corresponding to a discontinuity in the environment. Since discontinuities result from objects and their boundaries, and hence have spatial extent, our goal is not to detect points, but to detect lines of discontinuity which provide the best interpretation of the evidence supplied by the other constraints. The approach taken here is to construct a *confidence field* based on the previous constraints and use controlled continuity splines, or *snakes*, (Kass, Witkin, & Terzopoulos 1987) to detect local minima in the field.

We view the task of detecting lines of discontinuity as an energy minimization task where the internal spline forces E_{int} impose a smoothness constraint and the pointwise discontinuity confidence imposes external forces E_{disc} on the shape of the curve:

$$E_{sc} = \int_0^1 E_{int}(s) + E_{disc}(s)ds.$$

Local minima of the energy function correspond to possible lines of discontinuity and temporal, or higher level, processes may then be able to choose the global minima corresponding to actual discontinuities.

Our spatial consistency assumption gives us a model of discontinuities as continuous curves in the environment. The shape of these curves can be described by an internal spline energy function:

$$E_{int} = (\alpha(s)|\mathbf{v}_s(s)|^2 + \beta(s)|\mathbf{v}_{ss}(s)|^2)/2$$

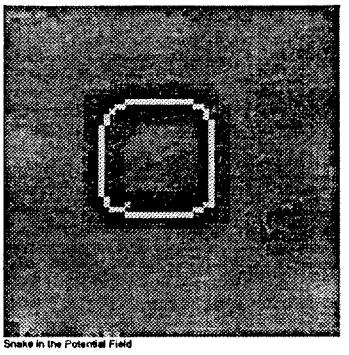


Figure 8: **Discontinuity confidence and spatial consistency.** Line of discontinuity detected using spatial consistency constraint superimposed on the confidence field, generated using C_S, $C_{S,A}$, $C_{F,C_{max}}$

where $\mathbf{v}(s) = (x(s), y(s))$ represents the position of the snake parametrically, $\mathbf{v}_s$ and $\mathbf{v}_{ss}$ are the first and second derivatives of the spline, and $\alpha(s)$ and $\beta(s)$ control to what extent the snake acts like a membrane and a thin plate respectively.

We can combine the pointwise information about discontinuity to form a confidence field Ψ where wells in the field correspond to areas where there is high confidence that a discontinuity is present:

$$\Psi = 1/(w_1 C_S + w_2 C_{S,A} + w_3 C_{F,C_{max}}),$$

where the w_i are scalar weights. Other formulations of the field are possible. The external energy force on the discontinuity spline is then just where $E_{disc} = w_{disc}\Psi$.

Figure 8 shows the confidence field for the random dot sequence with noise. A closed snake was initialized manually with an initial starting position roughly near the discontinuity. The figure shows one local minimum found by the snake as bright against the darker confidence field. The deep well about the discontinuity means initial placement of the snake can be fairly inaccurate. In our current work we are exploring ways of automating this instantiation process.

Temporal Consistency

Lines of discontinuity correspond to boundaries of surfaces in the environment. Under the reasonable assumption that surfaces tend to persist in time, we can expect that the discontinuities will also persist. This *temporal consistency* of discontinuities provides a powerful constraint which can be used to disambiguate between possible lines of discontinuity.

Temporal consistency implies that lines of discontinuity will move steadily across the optic array. This can be formulated as a constraint on the location and the motion of the *snakes*. In particular, the snake-velocity $\dot{\mathbf{v}}(s)$ should be consistent with the flow field of the frontal surface which gives rise to the discontinuity. Similarly, we may require that the snake acceleration $\ddot{\mathbf{v}}(s)$ be small.

The experiments reported in this paper have been based on two frames, and hence do not exploit temporal consistency. It appears, however, that the inclusion of this constraint for multiple-frame analysis will provide significant improvements.

Experimental Results

The constraints and associated confidence measures provide accurate discontinuity detection in random dot images, even in the presence of noise. These images, however, contain more texture than is common in images of natural scenes. A sequence of 64 × 64 pixel images of a cluttered office scene was used to test the constraints on real data. The densely sampled sequence contains two relatively homogeneous bars in the foreground moving across a stationary background containing areas of varying amounts of texture. The closest bar is undergoing approximately a 2 pixel displacement while the more distant bar is displaced by approximately 1 pixel. Noise, multiple discontinuities, and nearly homogeneous surfaces make this a challenging sequence for discontinuity detection.

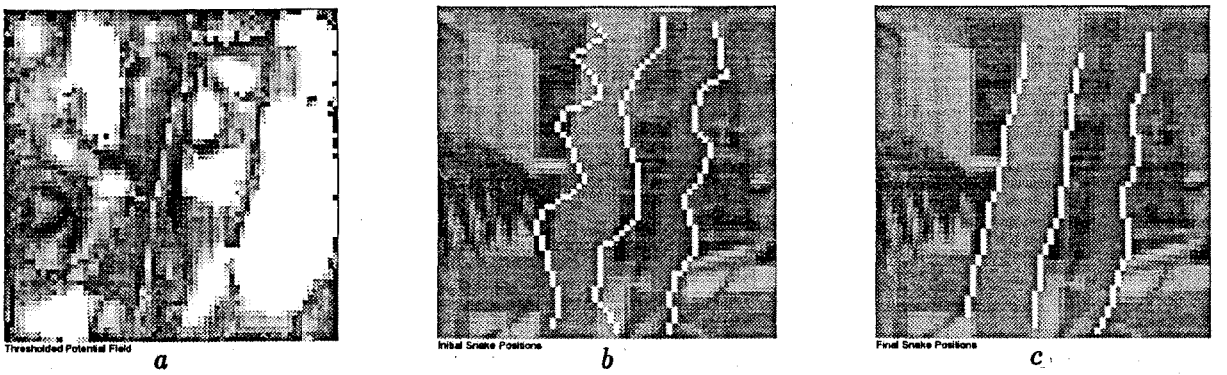
The images were first band-pass filtered. The SSD computation for the auto and cross correlation surfaces used a 7×7 search area and a 7×7 neighborhood with a uniform distribution. A 7×7 neighborhood was used for computing neighborhood flow. Figure 9*a* shows a thresholded image of the potential field generated using C_S, $C_{S,A}$, and $C_{F,C_{max}}$. Dark areas correspond to locations where there is high confidence that a discontinuity exists.

Snakes were initialized manually (figure 9*b*) in the general area of the discontinuity. This initialization process could be automated by using curves generated from intensity-based edge detection and perceptual grouping. Figure 9*c* shows the snakes resting at local minima in the field.

Conclusion

This paper has presented physical constraints which can be exploited to perform early detection of motion discontinuities over time. We have also presented a way of combining the various constraints in the form of an optimization problem, along the lines of the active contour models developed by (Kass, Witkin, & Terzopoulos 1987). Our approach is suitable for early stages in the processing of visual motion, and produces useful results even using our current formulation of the constraints, which is admittedly somewhat simple.

We are currently working on a Bayesian interpretation for our constraints. A conditional probability for a discontinuity can be obtained from each constraint and these can then be combined. The Bayesian model of the uncertainty developed in (Szeliski 1988) for flow-field computation provides hope that such a rigorous treatment is possible.



a *b* *c*

Figure 9: **Experiments.** *a)* Threshold of potential field using C_S, $C_{S,A}$ and $C_{F,C_{max}}$. *b)* initial snake positions. *c)* final positions.

There is also work to be done extending the constraints themselves; in particular, temporal constraints need to be incorporated. The possibility that discontinuities may appear, merge, split, grow, and shrink presents a number of interesting challenges in the use of snakes. The shape of the SSD surface and the use of weak continuity constraints deserve additional study, as do the possibilities for additional constraints. For example, it may be possible to combine dynamic discontinuity analysis with static image analysis.

There are also possibilities for feeding these discontinuities back into the correlation process. By explicitly accounting for discontinuities when computing the correlation it may be possible to achieve better estimates of flow. This idea relates to work in Markov Random Fields in which line processes are introduced to account for discontinuities (Geman & Geman 1984).

Finally, the value of this work will be demonstrated when it is applied to the problems of motion understanding. In particular, the incorporation of discontinuities into the smoothness constraint in flow field computation needs to be examined. The test will be whether early discontinuity detection can indeed be used to produce more accurate dense flow fields.

References

[1] P. Anandan. Computing dense displacement fields with confidence measures in scenes containing occlusion. In *SPIE Int. Conf. Robots and Computer Vision, 521*, pages 184–194, 1984.

[2] P. Anandan. *Measuring Visual Motion from Image Sequences.* PhD thesis, University of Massachusetts, Amherst, 1987. COINS TR 87-21.

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

[4] A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press, Cambridge, Massachusetts, 1987.

[5] P. J. Burt, C. Yen, and X. Xu. Local correlation measures for motion analysis a comparative study. Technical Report IPL-TR-024, Image Processing Lab., Rensselaer Polytechnic Institute, 1982.

[6] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6), November 1984.

[7] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proc. 1st ICCV*, pages 259–268, June 1987. London, UK.

[8] D. Marr. *Vision.* W. H. Freeman and Company, New York, 1982.

[9] M. Mutch, K. and B. Thompson, W. Analysis of accretion and deletion at boundaries in dynamic scenes. In W. Richards, editor, *Natural Computation*, pages 44–54. MIT Press, Cambridge, Mass., 1988.

[10] L. Potter, J. Scene segmentation using motion information. *IEEE Trans. on Systems, Man, and Cybernetics*, 5:390–394, 1980.

[11] B. G. Schunk. Image flow segmentation and estimation by constraint line clustering. *IEEE PAMI*, 11(10):1010–1027, Oct. 1989.

[12] A. Spoerri and S. Ullman. The early detection of motion boundaries. In *Proc. 1st ICCV*, pages 209–218. London, UK, June 1987.

[13] R. S. Szeliski. *Bayesian Modeling of Uncertainty in Low-Level Vision.* PhD thesis, Carnegie Mellon University, 1988.

[14] W. B. Thompson, K. M. Mutch, and V. Berzins. Edge detection in optical flow fields. In *Proc. of the Second National Conference on Artificial Intelligence*, August 1982.