In press, Proc. of International Workshop on Visual Form, Capri, Italy, May 1991.

COMPUTING EGOMOTION AND SHAPE FROM IMAGE MOTION USING COLLINEAR POINTS

Niels da Vitoria Lobo and John K. Tsotsos[1]

Department of Computer Science, University of Toronto,
Toronto, ON, Canada M5S 1A4

INTRODUCTION

Many years ago researchers (Helmholtz 1925, Gibson 1957) hypothesized that the 3-D motion and the shape of the environment are perceivable from the projected motion arising out of the relative motion between a monocular observer and the scene. In this paper, we summarize our recent computational solution to computing egomotion and show how shape information can be computed robustly.

The paradigm we work within assumes that an approximation to instantaneous image velocity (also termed *image flow* or *optic flow*) can be measured, and some progress has been made towards achieving such measurements (see Schunck 1986, Watson & Ahumada 1985, Heeger 1988, Fleet 1990, Woodham 1990). The subsequent step, that of computing the 3-D motion parameters and shape information from the instantaneous image velocity, has received ample attention from researchers. However, in addition to the fact that every 3-D algorithm proposed so far is not robust to noise in the input instantaneous velocity, the algorithms typically suffer from other important problems. Algorithms that permit general rigid motion and unrestricted shapes have to search in spaces that have at least three dimensions (Bruss & Horn 1983, Mitiche 1986), and the non-linear numerical methods used are very sensitive to the initial guess. Others assume some restricted form of motion (eg., no rotation in a certain dimension; see Barron 1988), or restrict the allowable shapes (eg., planarity), in order to get closed-form solutions for the unknowns (Waxman & Wohn 1987), or assume that some pa-

[1]John K. Tsotsos is the Canadian Pacific Fellow of the Canadian Institute for Advanced Research. This work was supported by the Natural Sciences and Engineering Research Council and through the Information Technology Research Center, a Province of Ontario Center of Excellence. The range data used in this paper came from the Range Image Database of NRC Canada.

rameters are known and solve for the others (Ballard & Kimball 1983, Matthies *et al.* 1989) — all of them too restrictive for general use.

The basis of our approach to computing egomotion is a technique for combining collinear image points which allows rotation to be cancelled in an exact manner. Along any straight line in the image, the rotational contribution to the image velocity component orthogonal to the line varies linearly with length, so that taking approximations to the second derivative of this component of velocity cancels out rotation. Thus despite the unrestricted motion and unrestricted shape involved in the problem, the motion parameters can be unlocked by a search for the correct direction of translation, which is a mere 2-dimensional search and incurs a far lower computational cost than other algorithms that search in higher dimensions. The algorithm, termed the *FOE Algorithm*, uses an operator that simultaneously cancels out rotation exactly and samples the translation contribution to find the direction of translation. Earlier, da Vitoria Lobo & Tsotsos (1990) showed that for three non-collinear image points, the pair-wise relative depths of the three scene points are dependent only on the unknown 3-D direction of translation and the known image velocities and image positions (i.e., that, in principle, knowledge of rotation is irrelevant to the calculation of shape from motion). For other work that cancels rotation see Prazdny (1983), Nelson & Aloimonos (1988), Heeger & Jepson (1990) and Weinshall (1990). Our approach also straightforwardly detects points that do not move in a manner consistent with the assumption of a rigid scene. We first review our *FOE algorithm* and its extension to detecting independent motion. Then we discuss the issue of acquiring shape information from the flow, knowing the location of the FOE. We propose conducting this in several stages: detecting discontinuities in depth and orientation, computing qualitative shape, and lastly combining the discontinuities and the qualitative information to produce quantitative shape.

COMPUTING THE EGOMOTION PARAMETERS

We sketch our egomotion computation here. Details appear in da Vitoria Lobo & Tsotsos (1991). Computing a discrete approximation to the second derivative of the normal component of image velocity for three collinear image points cancels out the rotational contribution from observer motion. Call the result of this computation a triplet *Sum*. This *Sum* is zero iff the three scene points are collinear, or there is no translation component of egomotion, or the line passing through the three image points also passes through the Focus of Expansion (FOE), which is the intersection of the direction of observer translation and the imaging surface. Figures 1a,b,c elaborate on the FOE. So we can design an operator as shown in Fig 1d, that when applied to the image flow associated with a rigid scene will give a zero response only at the FOE, unless there is no translational component of egomotion, or the whole scene is a single plane. In the latter two cases, the response is zero everywhere (so these two cases can be eliminated easily).

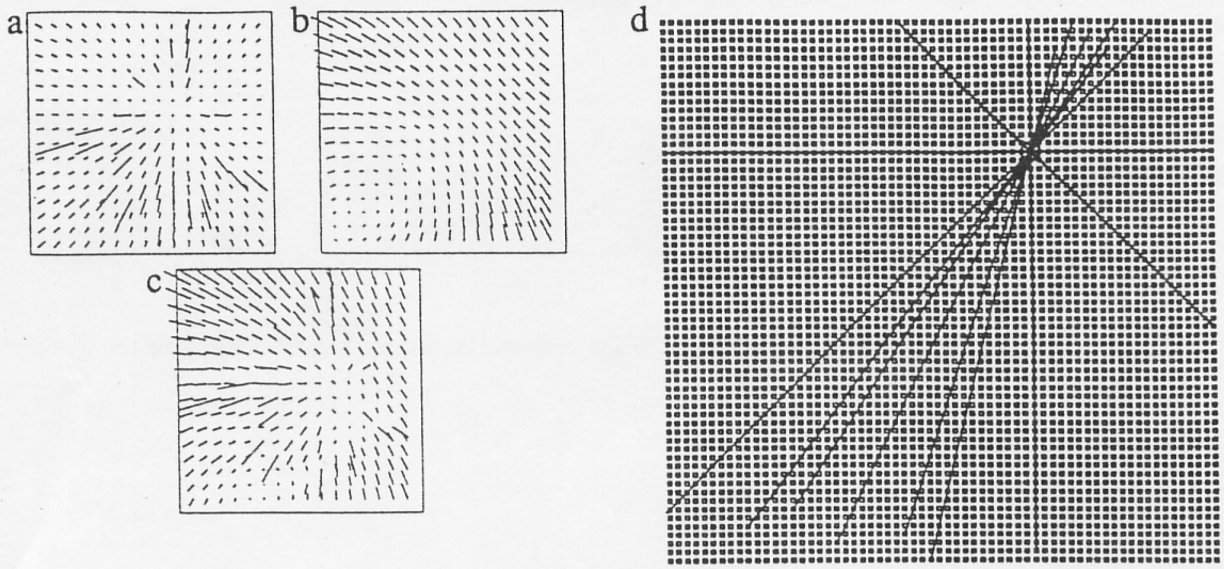

Figure 1:

Figs a, b, and c show flow fields generated when the observer a) only translates, b) only rotates, and c) the more typical case where the rotation and the translation have been combined. Note that in the first and third cases the FOE is just above and to the right of the image center and in the second there is no FOE because there is no observer translation. Fig d shows how an operator is made up of intersecting lines of points in the image. On a regular 256x256 square grid of points, at a single point we can have about 60-70 intersecting lines each of which passes through at least several image points. Here, only 9 lines are shown. The image is a dense grid of points, shown as hollow squares, each with a flow estimate associated with it. Some of the points used by each line have been blackened to identify them. Along each line triplet *Sums* are calculated for overlapping triplets. The absolute values of these are added to give a *LineSum*. The Line Sums are added to give the response of the operator at position (x,y) which is where the lines intersect.

Fig 2 shows the FOE being computed for a synthetic flow field. Even with noise (as high as 8% on average) in the flow, the FOE is found easily. With noise, the pit of the minumum broadens out but is still pronounced (somewhat like Fig 3b). Fig 2d shows how knowing the direction of 3-D translation and given ideal flow data, we can recover relative depth by an algorithm by da Vitoria Lobo & Tsotsos (1990). However, this relative depth algorithm is very sensitive to noise, and needs further enhancements to be of practical use.

DETECTING INDEPENDENT MOTION

Once the FOE has been recovered for a scene (even with independently moving regions in the scene; see Fig 3a,b), we can detect the independently moving regions. This is done by centering the FOE operator at the located FOE and scanning through its triplets in all directions looking for non-zero triplet *Sums*. These correspond to triplets in which at least one of the three points is moving with motion not consistent with the rigid scene's motion. These regions may need additional attention to ascertain their identity, and at the very least they should be segmented out to avoid corrupting the shape recovery process. Our algorithm's performance at detecting independent motion is shown in Fig 3c,d.

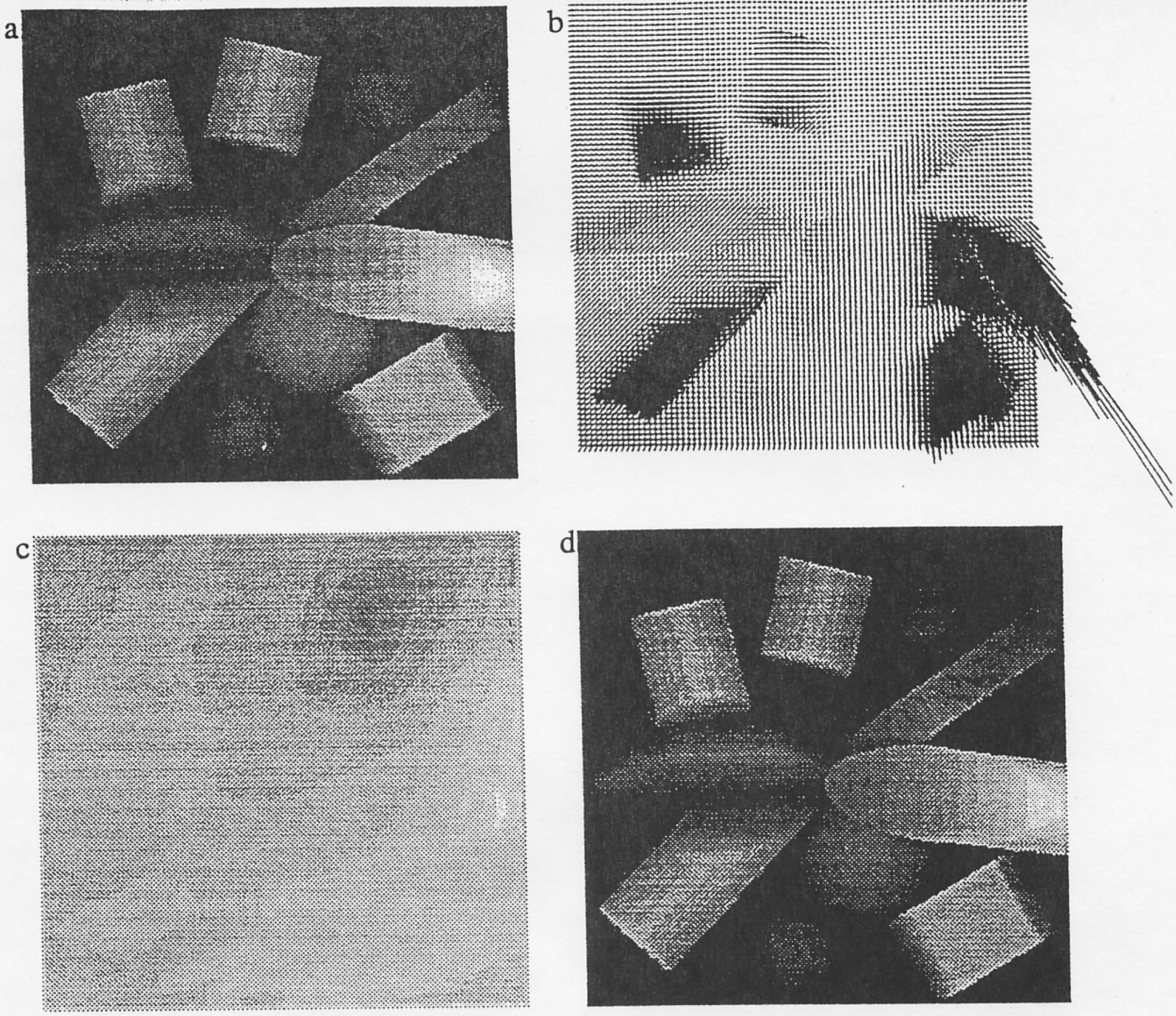

Figure 2:

a) Range data that was used for generating the flow field used in the experiments. Brightness represents nearness. An observer moves with instantaneous 3-D rotation and translation with respect to these depth points and the flow equations that relate image velocity to scene parameters are used to produce an instantaneous image velocity vector at each image point. b) Flow field generated using some 3-D motion and the range data in Fig 2a. c) FOE operator response map using the noise-free instantaneous flow field as input. The darkest point is the global minimum, and it is in exactly the correct position as the true FOE. In this and all subsequent response maps, brightness is proportional to the *Log* of the true response.

d) The reconstructed depth from the computed relative depth that is produced using a result involving non-collinear points by da Vitoria Lobo & Tsotsos (1990). Unfortunately, this reconstruction algorithm is very sensitive to noise.

ROBUST RECOVERY OF SHAPE INFORMATION

Now we turn our attention to recovering quantitative shape in a robust manner. We found that the relative depth computation used for the ideal data to produce Figure 2d, does not perform well when as little as 3% noise was introduced into the flow field values. We propose that the following intermediate steps can help to achieve robust shape recovery.

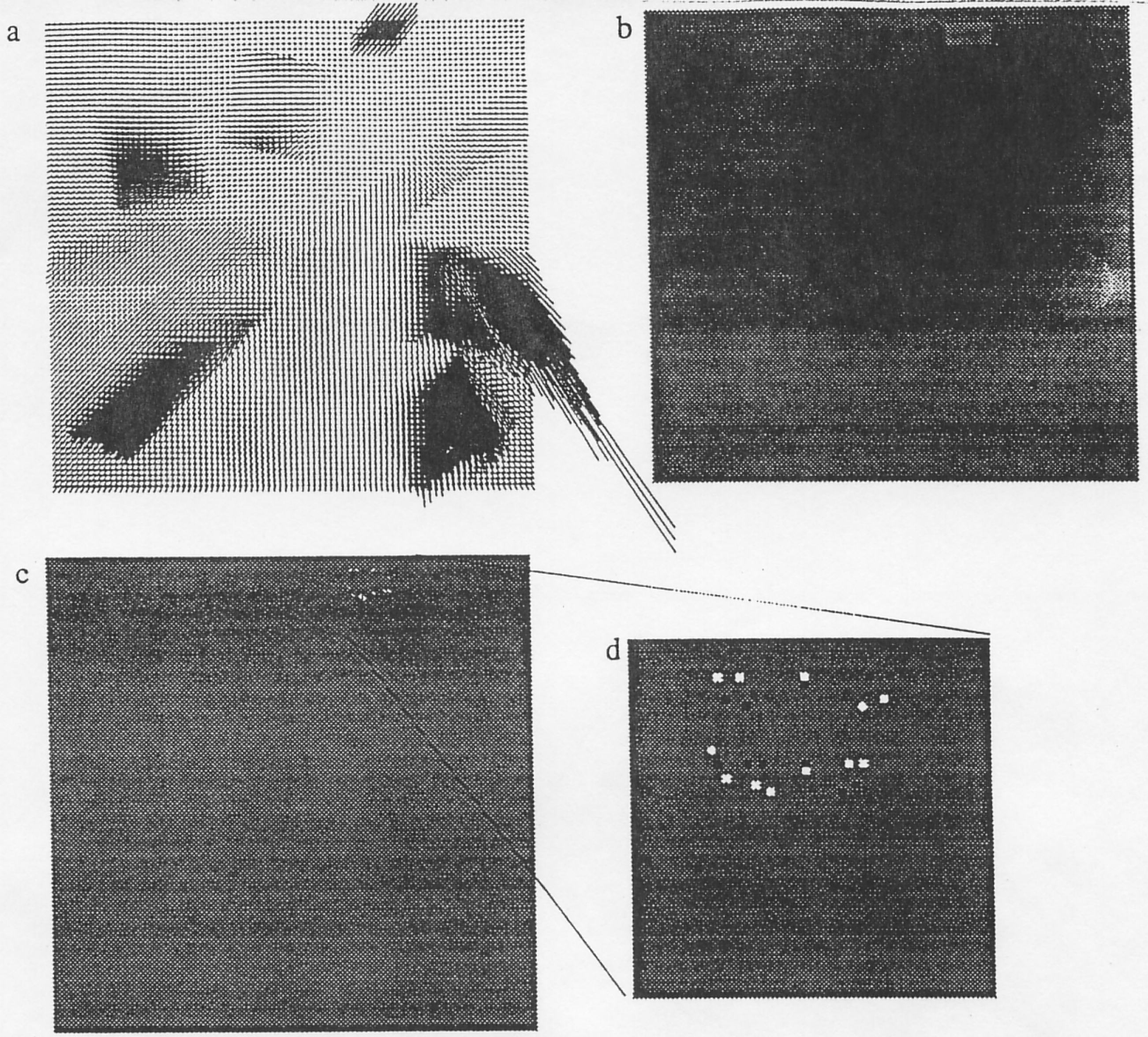

Figure 3:

a) Flow field that combines the original rigid flow field and an independently moving patch in upper right part of image. The frontoparallel rectangular patch is translating upward and to the right with an image flow magnitude that is quite significant. b) Response map for the flow associated with the nonrigid scene in which a patch moves independently. The global minimum is in the same position, indicating that the FOE computation is robust to non-rigidity in the scene. c) This figure shows how the points in triplets that overlap the patch in the flow field of Fig 3a have been automatically marked, as inconsistent triplets. d) shows an enlargement of the inset of inconsistent points. Note that because the patch itself is moving in a planar fashion, the triplet sum, when all three points are inside the patch, is zero. Hence these internal points have not been marked.

1. Use knowledge of the FOE to qualitatively identify depth discontinuities and discontinuities in the first derivative of depth (i.e., discontinuities in surface orientation.)

2. Use knowlegdge of the FOE to qualitatively characterize scene curvature information (convexity, concavity, planarity; see da Vitoria Lobo & Tsotsos 1990, and Weinshall 1990)

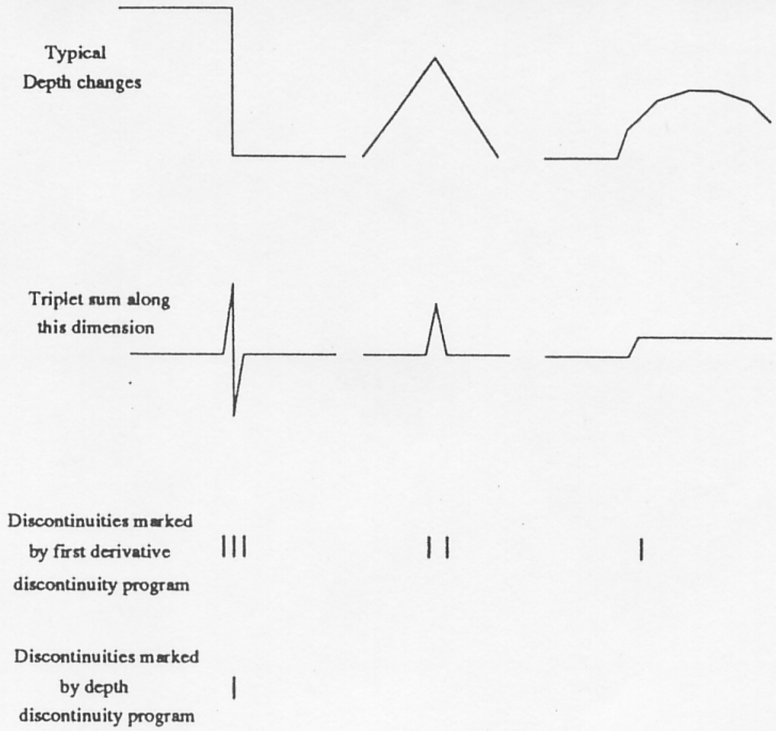

Figure 4:
This shows the kinds of depth and surface orientation discontinuities we are interested in, shown along one dimension. These however are embedded within the flow field. When the triplet operator is applied to the flow field along that dimension, the response has a characteristic response, shown in the second row. The lower two rows show the discontinuities that are to be found.

3. Use the computed information about boundaries and information about the qualitative curvature of regions to constrain a surface-fitting process.

The steps are not completely original. For instance, Waxman (1984) suggests similar ideas in computing "boundaries of analyticity". However, the boundary detection task in the first step has defied solution in noisy conditions. In this paper, we will not discuss the third step, as that is an area of work in progress (but see Terzopoulos 1987, for example.) The second step has been dealt with in other work (da Vitoria Lobo & Tsotsos 1990, and Weinshall 1990). The first step itself can be further subdivided into two stages, a *local discontinuity* detection stage, followed by a grouping process such as that computed by Zucker *et al.* (1988). We discuss the local discontinuity detection stage below.

To obtain depth discontinuities we are going to traverse a given line of collinear flow estimates, computing our triplet *Sums* along that line. This corresponds to an approximation to the second derivative of the flow component normal to that line, and depth discontinuities are signaled by zero-crossings (transitions from positive to negative sums, or vice-versa.) See Fig 4. Similarly, discontinuities in the first derivative of depth can be located by finding any transitions in sign of the *Sum* computations along the same line. Such transitions are more general than zero-crossings as they can include transitions from, for example, zero to

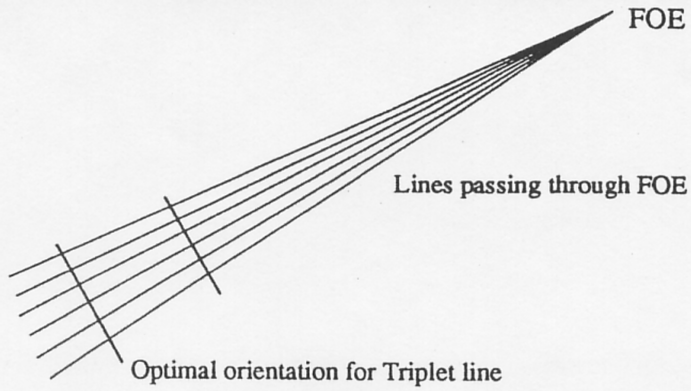

Figure 5:
The Translation part of the flow projects maximally along lines towards the FOE. Our triplet sum cancels out rotation while simultaneously sampling the translation part that projects normal to the line passing through the triplet. Hence we want to use collinear measurements that are oriented perpendicular to the lines through the FOE, so that we sample the larger velocity values. This increases robustness to noise, because triplet sums that sample larger values of velocity are less sensitive to the choice of threshold for deciding whether the sum is near-zero or positive or negative.

---

positive. It is clear from the edge detection literature (eg., Marr & Hildreth 1981, Witkin 1983) that the discontinuities located will be dependent on the scale of the second derivative operator (i.e., how far apart in the image the three points are). We will not pursue this here. What we address are the questions: given a certain scale, where does one find the most robust measurements, and how does one decide on thresholds for positive/negative/zero labels.

It has already been pointed out that the larger the velocity estimates are the more robust the estimates of velocity derivatives will be (Barron 1988). So, in a given region of the image, can we exploit our information about the FOE to help us ascertain in advance which triplet computations are likely to be more robust than others? Consider Figure 5. Note that the translation component of the flow vectors are of greatest magnitude in the direction along the lines passing through the FOE. i.e., for the translation component alone (ignoring rotation), flow vectors project maximally along lines through the FOE. Thus to sample these projections of the flow vectors, we need to pick our collinear points so that they lie along lines almost perpendicular to the lines passing through the FOE. Other orientations of triplet lines, will sample considerably smaller projections of the flow field. For instance, in the extreme case, if the triplet line is oriented almost towards the FOE, the velocity components it samples are almost zero. With noisy data, picking a threshold to determine whether a triplet *Sum* is zero or not is difficult. Thus, in principle, we should use triplets along small segments tangent to concentric circles around the FOE. Then along the same segment, we merely need to compare the signs of adjacent triplet *Sums* to detect the discontinuities. In practice, we will use a small range of orientations around the ideal tangent orientation.

As one gets closer to the FOE (independent of the orientation of the triplet line), the over-

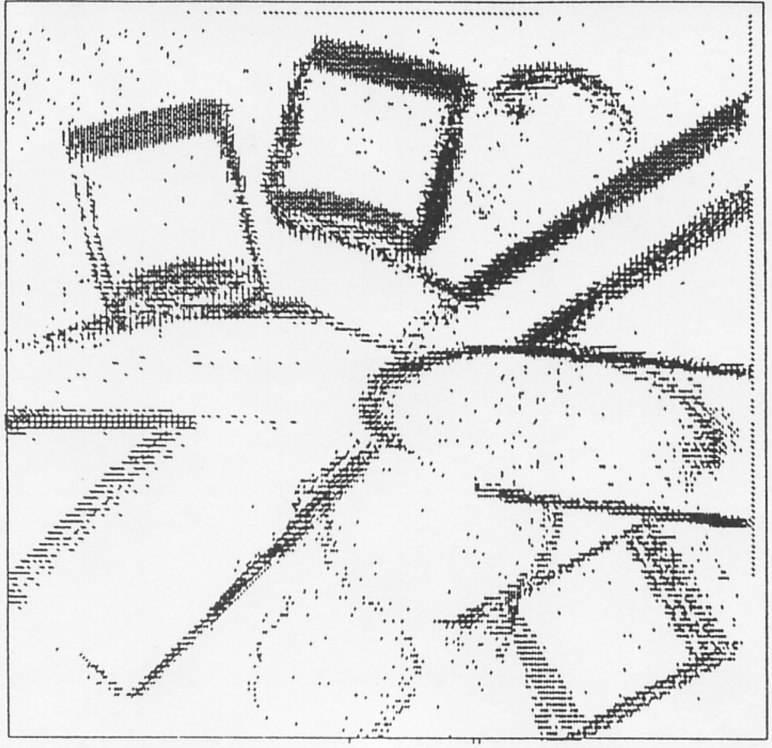

Figure 6:
Discontinuities found in the first derivative of depth, by applying the discontinuity detection algorithm to the flow field with 4 percent noise.

---

all magnitude of the flow vectors gets smaller, hence thresholds to determine whether or not there is a discontinuity in the depth or its first derivative, must be smaller as one gets closer to the FOE. Note that this variation in magnitude of the translation component across the image is proportional to the magnitude of the W component of 3-D velocity. (The translational component is given by $(-U + xW/Z, -V + xW/Z)$; see Longuet-Higgins & Prazdny 1980.) Another factor influencing magnitude of the translational part of the flow is depth; there is an inverse dependence here. Both these influences combine to give the magnitude of the translational contribution to flow at a point. Thus the magnitude of the flow in a local neighbourhood should influence the selection of a threshold.

We combined the two ideas — adaptive thresholding and using triplets tangent to concentric circles around the FOE — to demonstrate that they improve robustness to noise. The adaptive threshold for a region was chosen according to the average value of the lowest three flow magnitudes in a 3x3 patch of values. Figs 6 and 7 display the local discontinuities found in the first and zeroth derivative of depth respectively. The flow field used had upto 4% noise added to all its values (this in addition to the noise already present in the original range data). These results are promising, and we hope to be able to group discontinuities into continuous curves shortly. Many of the gaps in the discontinuity output are due to insufficient sampling along the tangent orientations. This was an artifact of the fact the our program only used 24 lines through an FOE operator, rather than the possible 60–70.

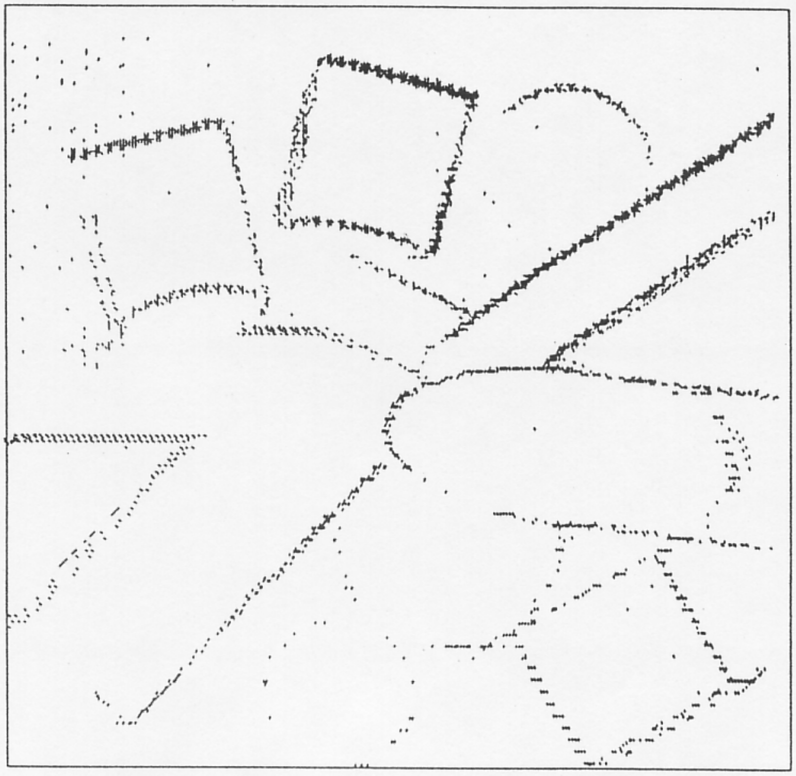

Figure 7:
Discontinuities found in depth, by applying the discontinuity detection algorithm to the flow field with 4 percent noise.

CONCLUSION

We have presented a novel approach to computing the translation component of egomotion, for an observer moving with unrestricted motion. The approach uses only collinear subsets of points. This enables the exact cancellation of rotation from the flow field. For these collinear points, only the component of the flow that is normal to the line of collinearity is needed. The tests of robustness of the egomotion computation are promising. We can also use this approach to detect independent motion in the scene.

The shape recovery process is still sensitive to noise. We propose that this be broken into a discontinuity detection stage, a qualitative stage, and a quantitative reconstruction stage. Some ideas for robustly detecting the discontinuities have been presented.

REFERENCES

Ballard D.H. and Kimball O.A. (1983), "Rigid body motion from depth and optical flow," *CVGIP*, 22, 95-115.

Barron J. (1988), "Determination of egomotion and environmental layout from noisy time-varying image velocity information in monocular image sequences," *Ph.D. Thesis*, Dept. of Computer Science, Univ. of Toronto. Available as RBCV-TR-88-24.

Bruss A.R. and Horn B.K.P. (1983), "Passive navigation," *CVGIP*, **21**, 3-20.