3-D MOTION RECOVERY FROM TIME-VARYING OPTICAL FLOWS

Kwangyoen Wohn and Jian Wu

Division of Applied Sciences
Harvard University
Cambridge, MA 02138

Abstract

Previous research on analyzing time-varying image sequences has concentrated on finding the necessary (and sufficient) conditions for a unique 3-D solution. While such an approach provides useful theoretical insight, the resulting algorithms turn out to be too sensitive to be of pratical use. We claim that any robust algorithm must improve the 3-D solution adaptively over time. As the first step toward such a paradigm, in this paper we present an algorithm for 3-D motion computation, given time-varying optical flow fields. The surface of the object in the scene is assumed to be locally planar. It is also assumed that 3-D velocity vectors are piecewise constant over three consecutive frames (or 2 snapshots of flow field). Our formulation relates 3-D motion and object geometry with the optical flow vector as well as its spatial and temporal derivatives. The deformation parameters of the first kind, or equivalently, the first-order flow approximation (in space and time) is sufficient to recover rigid body motion and local surface structure from the local instantaneous flow field. We also demonstrate, through a sensitivity analysis carried out for synthetic and natural motions in space, that 3-D inference can be made reliably.

1. INTRODUCTION

When an object moves relative to a viewer, the projected image of the object also moves in the image plane. By analyzing this evolving image sequence, one hopes to extract the instantaneous 3-D motion and surface structure of the object. The path from time-varying imagery to its corresponding 3-D description may be divided into two relatively independent steps: (1) computation of 2-D image motion from the image sequence, and (2) computation of 3-D motion and structure of objects from 2-D motion. This paper deals with the latter issue.

This work was supported by the Office of Naval Research under Contract N00014-84-K-0504 and Joint Service Electronics Program under Contract N00014-84-K-0465.

The relations between 2-D motion and 3-D environment are formulated in terms of non-linear equations. This non-linearity prevents us from solving them in a trivial way, which makes the problem mathematically interesting. The schemes used to interpret 2-D motion information can be classified into two categories depending on the kind of 2-D motion representation utilized. One may use the motion of distinct, well-isolated feature points [6]. The other approach uses the continuous flow field within a small region [4]. While either method has its own merits and drawbacks, the second approach leads to stable solutions provided that the partial derivatives of the flow field are available [8]. Waxman and Wohn [9] developed the methods of extracting the partial derivatives of the flow field directly from evolving contours over time. They have demonstrated that the combined algorithms of 2-D flow computation and 3-D structure and motion computation are quite stable with respect to input noise and changes in surface structure.

Although the approach of Waxman et. al. behaves much better than its predecessors, it is still questionable that all the partial derivatives of flow up to second order could be recovered reliably enough to produce a meaningful 3-D solution. While no rigorous analysis on the behavior of this algorithm has been so far conducted, it has turned out that, in general, the second-order derivatives determine the accuracy of 3-D solution, and they are not reliable as the field of view decreases under 20°.

Our approach is based on the derivatives of flow up to the first order. In our recent experiments conducted on various natural images, we found that the first-order derivatives can be recovered with greater accuracy than those of second-order. However, the first-order derivatives alone do not provide the sufficient condition for solving 3-D motion. We obtain additional constraints by introducing the temporal derivatives of optical flow. The new representational scheme of 2-D motion then consists of four first-order spatial derivatives and two first-order temporal derivatives of local flow field. We call these partial derivatives "deformation parameters of the first kind". The idea of utilizing multiple frames has been proposed by several researchers for a restricted class of rigid-body motion [3], for semi-rigid motion

under orthographic projection [7], and for determining the focus of expansion (or contraction) [1].

In Section 2, we begin our discussion of the 3-D motion recovery process, adopting the new flow representation scheme. The formulation given here relates this new local representation of an optical flow to object motion and structure, in terms of eight non-linear algebraic equations. This formulation requires that the object surface be approximated as locally planar and that 3-D velocities do not change over a short period of time. Our method reveals certain families of degenerate cases for which the temporal derivatives of flow do not provide sufficient independent information. In certain cases, the temporal change of the first-order derivatives may be used. In some other cases, multiple solutions may result due to the non-linearity. The complete solution tree is also presented. We conduct a stability analysis of the algorithm in Section 3. Some results of experiments conducted on synthetic data as well as real time-varying images are presented. Concluding remarks and our future direction follow in Section 4.

2. SPATIO-TEMPORAL IMAGE DEFORMATION

Adopting a 3-D coordinate system (X, Y, Z) as in [4] (See Figure 1), the relative motion is represented in terms of viewer motion: the translational velocity $\boldsymbol{V} = (V_X, V_Y, V_Z)$, and the rotational velocity $\boldsymbol{\Omega} = (\Omega_X, \Omega_Y, \Omega_Z)$. The origin of the image coordinate system (x, y) is located at $(X, Y, Z) = (0, 0, 1)$.

As a point P in space (located by position vector $\boldsymbol{R}$) moves with a relative velocity $\boldsymbol{U} = -(\boldsymbol{V} + \boldsymbol{\Omega} \times \boldsymbol{R})$, The corresponding point p moves with a velocity:

$$v_x = \left\{ x \frac{V_Z}{Z} - \frac{V_X}{Z} \right\} + \left\{ xy\,\Omega_X - (1 + x^2)\,\Omega_Y + y\,\Omega_Z \right\}, \tag{2.1a}$$

$$v_y = \left\{ y \frac{V_Z}{Z} - \frac{V_Y}{Z} \right\} + \left\{ (1 + y^2)\,\Omega_X - xy\,\Omega_Y - x\,\Omega_Z \right\}. \tag{2.1b}$$

These equations define an instantaneous image flow field, assigning a unique 2-D image velocity $\boldsymbol{v}$ to each direction (x, y) in the observer's field of view.

2.1. Spatial Coherence of Optical Flow

Equations (2.1) constitute two independent relations among seven unknowns. Various techniques for recovering the 3-D parameters differ by the way the additional constraints are provided. We shall consider only a single, smooth surface patch of some object in the field of view, such that the surface is differentiable with respect to the image coordinates. Let us further assume that such surface patch be locally approximated by a planar surface in space as:

$$Z = pX + qY + Z_0 \tag{2.2}$$

Substituting the above into the flow relation (2.1), we get expressions in the form of a second-order flow field with respect to the image coordinates (x,y). On the other hand, non-planar surfaces generate flows which are not simple polynomials in the image coordinates.

Following [8], we form the partial derivatives of flow in equations (2.1) with respect to the image coordinates and evaluate them at the image origin; we get the following four independent relations:

$$v_{x,x} \equiv \left(\frac{\partial v_x}{\partial x} \right)_0 = V'_Z + V'_X\, p, \tag{2.3a}$$

$$v_{x,y} \equiv \left(\frac{\partial v_x}{\partial y} \right)_0 = V'_X\, q + \Omega_Z, \tag{2.3b}$$

$$v_{y,x} \equiv \left(\frac{\partial v_y}{\partial x} \right)_0 = V'_Y\, p - \Omega_Z, \tag{2.3c}$$

$$v_{y,y} \equiv \left(\frac{\partial v_y}{\partial y} \right)_0 = V'_Z + V'_Y\, q, \tag{2.3d}$$

where we have introduced the normalized motion parameters

$$V'_X \equiv \frac{V_X}{Z_0}, \quad V'_Y \equiv \frac{V_Y}{Z_0}, \quad V'_Z \equiv \frac{V_Z}{Z_0}. \tag{2.4a,b,c}$$

The quantities on the left-hand side represent the relative motion (or geometrical deformation) in an infinitesimal neighborhood. The above process adds 4 relations while replacing the unknown Z with p and q. Now we have 6 relations (equations (2.1) evaluated at the origin and (2.3)) and 8 unknowns.

2.2. Temporal Coherence of Optical Flow

Unlike [4] and [8] in which the additional constraints are obtained by introducing higher-order derivatives of flow, we observe that the flow field itself changes smoothly over time unless the following situation(s) occur:

1) abrupt change of 3-D motion in time, and/or
2) abrupt change of object distance in time.

For the latter case, it can be easily shown that the temporal change of object distance is given by:

$$\frac{1}{Z_0} \left(\frac{\partial Z}{\partial t} \right)_0 = -V'_Z + p\,(V'_X + \Omega_Y) + q\,(V'_Y - \Omega_X) \tag{2.5}$$

and the previous assumption on surface planarity eliminates this possibility. This observation leads us to investigate the way flow field changes over time. Now, let us assume that 3-D motion parameters do not change at all during an infinitesimal time period. Differentiating (2.1) with respect to time, and utilizing

(2.5):

$$v_{x,t} \equiv \left(\frac{\partial v_x}{\partial t}\right)_0 = V'_X \delta, \quad (2.6a)$$

$$v_{y,t} \equiv \left(\frac{\partial v_y}{\partial t}\right)_0 = V'_Y \delta, \quad (2.6b)$$

where $\delta \equiv -V'_Z - p\ v_x - q\ v_y$. δ describes the rate of depth change in time, normalized by Z, the absolute distance. The above quantities represent the flow changes in time, at the center of the image coordinates. Notice that we have obtained two independent relations without introducing any additional unknown.

There is a class of motion for which equations (2.6) become redundant and do not provide any new constraint. In such a case, we observe that the temporal change of spatial deformation may also provide independent relations:

$$v_{x,xt} \equiv \left(\frac{\partial v_{x,x}}{\partial t}\right)_0 = -V'_Z \delta + V'_X p \gamma + V'_X q \Omega_Z + V'_X \Omega_Y, \quad (2.7a)$$

$$v_{x,yt} \equiv \left(\frac{\partial v_{x,y}}{\partial t}\right)_0 = V'_X q \gamma - V'_X p \Omega_Z - V'_X \Omega_X, \quad (2.7b)$$

$$v_{y,xt} \equiv \left(\frac{\partial v_{y,x}}{\partial t}\right)_0 = V'_Y p \gamma - V'_Y q \Omega_Z - V'_Y \Omega_Y, \quad (2.7c)$$

$$v_{y,yt} \equiv \left(\frac{\partial v_{y,y}}{\partial t}\right)_0 = -V'_Z \delta + V'_Y q \gamma + V'_Y p \Omega_Z + V'_Y \Omega_Y, \quad (2.7d)$$

where $\gamma = V'_Z - pV'_X - qV'_Y$.

2.3. Recovery of 3-D Motion

Once the deformation parameters are obtained, we can proceed to recover the 3-D motion and structure of a surface patch. The eight equations to be solved ((2.1), (2.3) and (2.6)) involve eight unknowns: V'_X, V'_Y, V'_Z, Ω_X, Ω_Y, Ω_Z, p and q. We first observe that V'_X and V'_Y are coupled with p and q as seen in equations (2.3). This suggests to us a two-step method: (1) to solve for the products as a whole, and (2) to separate them into individual parameters. The detailed derivation may be found in [11]. When $V'_Z = 0$, 3-D motion is restricted to the motion parallel to the image plane. Substituting $V'_Z = 0$ into the original equations (2.3), one can see that the coupled terms $V'_X p$, $V'_X q$, $V'_Y p$, $V'_Y q$ cannot be separated. We now utilize the temporal change of the spatial deformation parameters: $v_{x,xt}$, $v_{x,yt}$, $v_{y,xt}$ and $v_{y,yt}$, as depicted in equations (2.7). They provide sufficient constraints for solving for 3-D parameters.

In summary, one can determine 3-D motion uniquely except the following cases:

(i) Stationary flow: The flow field does not change over time. Our method fails to recover 3-D motion.

(ii) $V'_Z = 0$: Motion is parallel to the image plane. Dual solution may exist.

The complete solution tree is presented in Figure 3.

3. SENSITIVITY ANALYSIS AND EXPERIMENT

Unlike other existing algorithms such as [5,8], our method solves for 3-D motion without involving any searching or iterative improvement. Given the perfect deformation parameters (or flow field), the algorithm recovers the exact 3-D solution. In practice, however, there are various factors which reduce the accuracy of deformation parameters. Beside the sources of error common to any early vision processing such as the error due to digitization, camera distortion, noise, truncation error due to discrete arithmetic, etc., there are several other factors which affect the accuracy of calculated values of deformation parameters. In principle, the deformation parameters may be obtained first by recovering optical flow and secondly by taking the partial derivatives of optical flow. But since the differentiation process will amplify noise, we are unlikely to recover these partial derivatives reliably.

3.1. Recovery of Optical Flows

In our experiment conducted on the natural time-varying image, contours are used as a primary source of information. The following factors concerning various stages of the flow computation must be considered and their effects must be analyzed.

Imperfect contour extraction and false matching: Since we utilize the contours to sample the image deformation experienced by a neighborhood in the image, for all points on a contour in the first image there must be corresponding points on the matching contour in the second image, and vice versa. Hence contours such as those which corresponds to extremal boundary or shadow boundary must be excluded. However, there is no reliable method for classifying edges into meaningful categories, based on the local intensity measurement. In fact, the analytic boundaries of flow field suggest a way of classifying edges [2].

The performance of edge operators severely affects the forthcoming computation. In our current implementation, contours are obtained from the zerocrossings of $\nabla^2 G^* I$. Although zerocrossings possess many attractive features by itself, they may not correspond to the actual meaningful edges.

Imperfect normal flow estimation: Having determined the pair of matching contours, one can measure only the *normal flow* around the contour, whereas the motion along the contour is invisible. We make use of geometrical displacements normal to contours as shown in Figure 2. Since the points along the contour generally have some component of motion tangential to the contour as well, the measured normal flow in this way is not exactly equal to the true normal flow. In most cases, the effect of this tangential component on the resulting full flow is not negligible and the 3-D solution becomes of no use at all. In this regard, we developed an algorithm which iteratively improves the normal flow estimates [12].

Inaccurate flow model: Given the normal flows, the deformation parameters (or equivalently optical flow) can be recovered by the Velocity Functional Method proposed by [9]. It considers the second-order flow approximation as the starting point of flow recovery. It then computes the best-fitting second-order flow from the local measure of normal flow. Although the second-order terms will not be used at the later stage of 3-D motion computation, they are included here in order to "absorb" noise, and thereby to obtain more accurate spatial deformation parameters. Currently temporal deformation parameters are obtained by subtracting the spatial deformation parameters over two consecutive image frames. Alternatively, as a better approach, the velocity can be approximated as the truncated Taylor series in the spatio-temporal coordinates (x, y, t).

In general, such approximation is not exact. It involves the truncation error which is characterized by many quantities. However, as we have mentioned in Section 2, the second-order approximation is exact for planar surfaces. For curved surfaces, it has been shown that the exact error formula is determined mainly by the surface curvature and the size of the neighborhood [10]. By keeping the neighborhood size small enough, one can still rely on the second-order model.

Non-uniform 3-D motion: In the mathematical sense, non-uniformity of motion is not a problem since all we need is the change of flow in infinitesmal time period. In practice, we have to subtract (or differentiate) two flow fields, which means three snapshots of images are needed at least. The motion is assumed to be constant during this time interval. The effect of 3-D motion change, i.e. acceleration, on the first-order image deformation can be shown to be proportional to the amount of acceleration.

3.2. Experiment

In the first experiment, we test the sensitivity of the proposed algorithm by using synthetically generated data. Typical sensitivity is well illustrated by the following case. A planar surface in space is described by $Z = Z_0 + pX + qY$ with $Z_0 = 10$ units, and p and q corresponding to slopes of 30° and 45°, respectively. The observer moves with translational velocity $V = (5, 4, 3)$ units/frame and rotational velocity $\Omega = (20, -10, 30)$ degrees/frame. On the image plane we specify two algebraic curves along which the normal flow can then be computed exactly (See Figure 4a). We then perturb the magnitude and direction of the normal flow vectors randomly (from a uniform distribution), bounded by a specified percentage of the exact normal flow (Figure 4b). The viewing angle is fixed at 20°. The Velocity Functional Method is then applied to the input normal flows (Figure 4c). We consider three measures of sensitivity which characterize the relative error in surface orientation e_S, in translation e_T, and in rotation e_R. One can see from Table 1 that the sensitivity of 3-D structure and motion predictions is fairly linear to the normal flow. We have found that perturbations to the normal flow of about 20% can be tolerated down to fields of view of about 10°.

In the next example, the normal flow vectors are measured from the synthetic images. The same contours defined above undergo the following motion: translational velocity $V' = (0.06, 0.03, -0.04)$ and rotational velocity $\Omega = 0$. The slope of the surface is measured roughly as $p = 30°$ and $q = 45°$. Three frames of digital images (of size 256 by 256) were generated from the graphical simulation program (Figures 5a). Normal flows along the contours are measured from the pair of consecutive frames (Figure 5b). The iterative process in [12] was used to recover the full flows (Figure 5c). The 3-D parameters computed from this full flow are: $V' =$ (0.059, 0.028, −0.038) units/frame, $\Omega =$ (−0.0021, 0.0006, −0.0007) degrees/frame, $p = 31.87°$ and $q = 46.26°$.

As the last example, Figure 6a shows one of three consecutive images obtained from the natural environment. A CCD camera with known viewing angle and focal length was attached to a robot arm so that the camera can follow a predetermined trajectory. The images are 385 by 254 in size, with 6 bits of resolution. The motion between the successive frames is given as $V' = (0.6, -0.25, 0.4)$ and $\Omega = 0$. The orientation of the table was given as $p = -10°$ and $q = 55°$. A pyramid-based flow recovery scheme being developed is applied to the input images. Figure 6b shows the flow field obtained from the first two frames, assuming that the entire image consists of a single planar surface. 3-D parameters computed from this flow field are: $V' =$ (0.45, 0.18, 0.57) units/frame, $\Omega =$ (0.028, −0.037, 0.021) degrees/frame, $p = -25.11°$ and $q = 56.29°$.

4. CONCLUDING REMARKS

This work is a part of the hand-eye coordination project at Robotics Laboratory, Harvard University. In order to provide 3-D geometrical information of scene from visual data, binocular/time-varying images will be used as main visual source.

We presented an algorithm which recovers 3-D structure and motion from the first-order deformation parameters. In most cases, such deformation parameters can be recovered quite reliably. We carried out a sensitivity analysis for synthetic and natural motions in space, to demonstrate that 3-D information can also be recovered reliably. Although many independent factors affect the accuracy of 3-D solution, we have found, through various experiments, that the temporal deformation given as $v_{x,t}$ and $v_{y,t}$ plays the major role on the sensitivity. More work should be done on the flow recovery scheme. We are currently investigating a muti-resolution approach which integrates several different methods of flow computation. The 3-D solution can also be improved further by exploiting the idea of 3-D motion coherence on the level of 3-D motion computation. We are developing a predictive filtering scheme based on a dynamic model.

REFERENCES

[1] A. Bandyopadhyay and J. Aloimonos, "Perception of Rigid Motion from Spatio-Temporal Derivatives of Optical Flow", *Computer Science TR-157*, Univ. Rochester, March 1985.

[2] W. F. Clocksin, "Computer Prediction of Visual Thresholds for Surface Slant and Edge Detection from Optical Flow Fields", Ph.D. Dissertation, Univ. Edinburgh, 1980.

[3] D. D. Hoffman, "Inferring Local Surface Orientation from Motion Fields", *Journal Optical Soc. Am.* **72**, 888-892, 1982.

[4] H. Longuet-Higgins and K. Prazdny, "The Interpretation of a Moving Retinal Image", *Proc. Royal Soc. London* **B 208**, 385-397, 1980.

[5] A. Mitiche, "On Kineopsis and Computation of Structure and Motion", *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 109-112, 1986.

[6] R. Y. Tsai and T.S. Huang, "Uniqueness and Estimation of 3-D Motion Parameters of Rigid Objects with Curved Surfaces", *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 13-27, 1984.

[7] S. Ullman, "Maximizing Rigidity: The Incremental Recovery of 3-D Structure and Rubbery Motion", *A.I. Memo 721*, M.I.T., June 1983.

[8] A. Waxman, and S. Ullman, "Surface Structure and 3-D Motion From Image Flow: A Kinematic Analysis", *Intl. Journal of Robotics Research* **4**, 72-94, 1985.

[9] A. M. Waxman and K. Wohn, "Contour Evolution, Neighborhood Deformation and Global Image Flow: Planar Surfaces in Motion", *Intl. Journal of Robotics Research* **4**, 95-108, 1985.

[10] K. Wohn and A. M. Waxman, "Contour Evolution, Neighborhood Deformation and Local Image Flow: Curved Surfaces in Motion", *Computer Science Tech. Report TR-1531*, July, 1985.

[11] K. Wohn and J. Wu, "Recovery of 3-D Structure and Motion from 1'st-Order Image Deformation", *Robotics Lab. Tech. Report*, Harvard University, January 1986.

[12] K. Wohn, J. Wu and R. Brockett, "A Contour-based Recovery of Optical Flows by Iterative Improvement", *Robotics Lab. Tech. Report*, Harvard University, June 1986.

noise	e_T	e_R	e_S
0 %	0.000%	0.000%	0.000%
5 %	0.171%	1.493%	0.064%
10%	1.294%	4.532%	0.224%
15%	4.324%	8.218%	0.965%
20%	10.080%	11.115%	2.428%
25%	18.338%	12.523%	4.884%

Table 1. Relative error in 3-D solution with respect to the noise in normal flow.

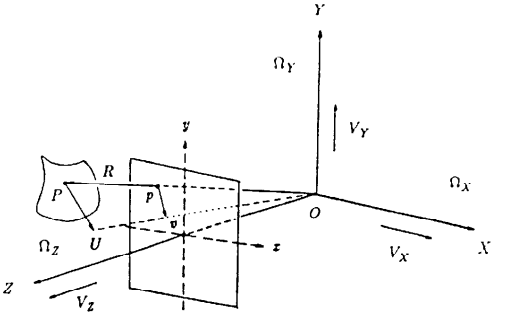

Figure 1. 3-D coordinate system and 2-D image coordinates with motion relative to an object.

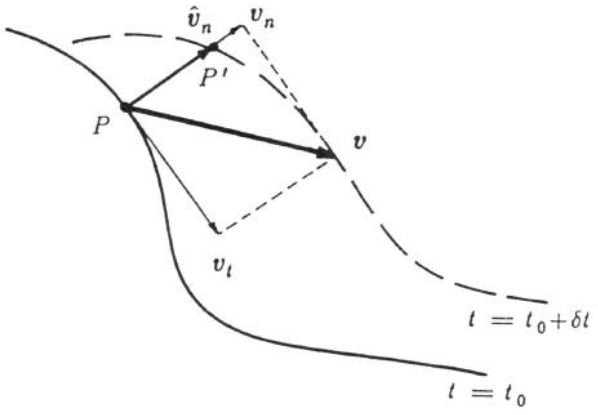

Figure 2. Estimating normal flow.
The normal flow cannot be obtained exactly due to the tangential component.

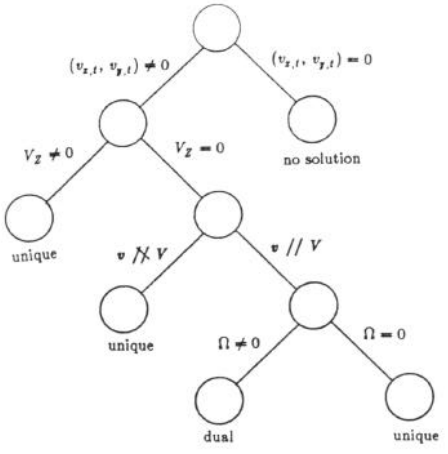

Figure 3. Solution tree for 3-D structure and motion algorithm.

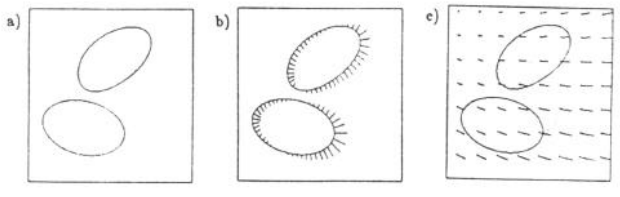

Figure 4. Recovering full flow from normal flow.
a) contours on the image plane.
b) normal flow along the contours as input (no noise added).
c) optical flow field recovered.

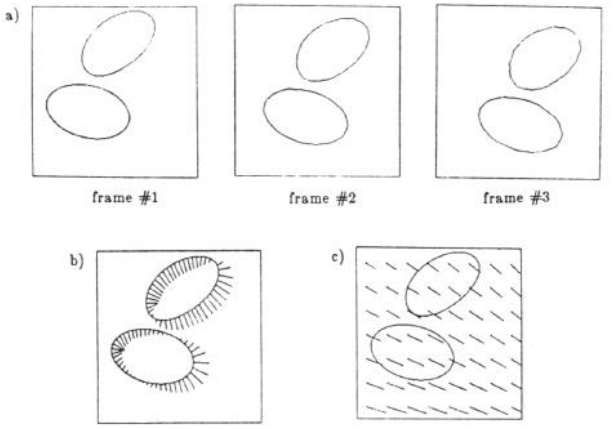

Figure 5. Recovering optical flow from contours.
a) contours on the image planes as input. (3 frames are shown.)
b) measured normal flow along the contours (from frame #1 and #2).
c) optical flow field recovered.

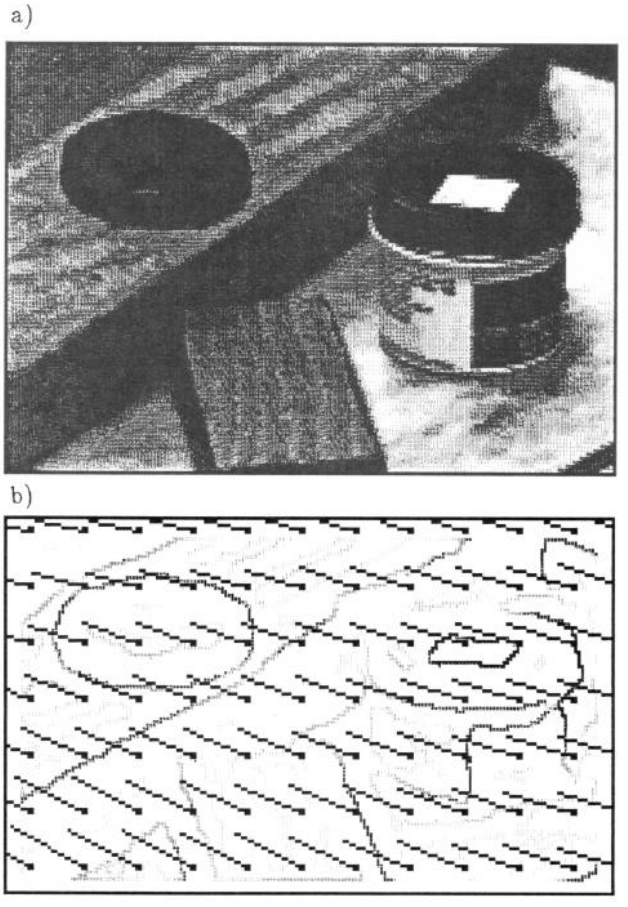

Figure 6. Recovering optical flow from natural images.
a) input images. (One frame is shown).
b) flow field recovered (with zerocrossings).