COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 38, 122–146 (1987)

Structure and Motion from Optical Flow under Perspective Projection

KEN-ICHI KANATANI*

Center for Automation Research, University of Maryland, College Park, Maryland 20742

Received January 7, 1986; accepted July 23, 1986

The 3D structure and motion of an object are determined from its optical flow under perspective projection. The solution is given in explicit analytical form in terms of the parameters characterizing the flow of planar motion. The solution is not unique, but the spurious solution disappears if two or more planar parts of the same object are observed, and the adjacency condition of optical flows is explicitly obtained. Unique interpretation also becomes possible by considering the transition to the "pseudo-orthographic approximation," since no spurious solution arises in this approximation. The choice of the coordinate system and parameterization of rigid motion are discussed in relation to robustness of computation.

1. INTRODUCTION

This is a continuation of Kanatani's paper [15], in which the 3D structure and motion of an object are recovered from its optical flow under orthographic projection. The solution is given explicitly in analytical terms, from which various geometrical interpretations are obtained in invariant forms. In this paper, exactly the same analysis is given for optical flows under perspective projection.

Recovery of the 3D structure and motion from perspectively projected images has already been studied by many people, and many different approaches have been presented. One approach is to reconstruct numerically the 3D structure from the 2D displacements of a finite number of points, making no assumptions about the object except rigidity [4, 5, 17, 21, 23, 25, 28, 29]. When motion is infinitesimal, the displacement is identified with the velocity, or the *optical flow*. The problem becomes somewhat simpler if we assume an object model, say a planar surface. From an optical flow, various clues are used for computation—the velocity values at particular points, global characteristics such as the vanishing points or lines, coefficients of the equations fitted to the observed flow, contour evolution, etc. [3, 16, 18, 19, 24, 26, 27, 30, 31].

In this paper, we fit global equations to the observed optical flow and express the 3D structure and motion analytically in terms of the parameters characterizing the flow. We consider objects whose surfaces consist of, or are approximated by, planar faces and focus on each of these planar portions. For the case where the object is a planar surface, a complete solution, including the existence of spurious solution, has already been given by Longuet-Higgins [18], and Subbarao and Waxman [27]. However, the present approach is different from theirs in some respects.

First, the choice of the coordinate system is different. In almost all existing studies of this type of problem, the origin of the *xyz*-coordinate system is chosen as the "viewpoint" or the "camera focus." Here, however, we choose the *xy*-plane as

* Current address: Department of Computer Science, Gunma University, Kiryu, Gunma 376, Japan.

0734-189X/87 $3.00

the image plane. If the viewpoint is moved away from the image plane (mathematically, in the limit of the focal length approaching infinity), the orthographic projection is obtained as the limit. Thus, both perspective and orthographic projections can be treated in the same framework. Special scaling is necessary if we want to obtain the orthographic limit in the conventional setting. Moreover, in our setting, we can obtain an intermediate approximation, which we call the *pseudo-orthographic approximation*, in the transition to the orthographic limit.

As was pointed out first by Hay [7] and later by Longuet-Higgins [18] and Subbarao and Waxman [27], two different solutions exist for the same optical flow if the object surface is a plane, and it has been believed that one cannot distinguish the true solution from the spurious one from a single optical flow field, since the two solutions yield exactly the same optical flow. However, we will show that the distinction is possible in some sense, because *no spurious solution exists for the pseudo-orthographic approximation*. Hence, we can pick out one of the perspective solutions which corresponds to the pseudo-orthographic solution. This means, as will be shown later, that we are picking out one solution for which the surface is far away and projective distortion is small.

As is well known, instantaneous rigid motion is specified by the translation velocity and rotation velocity at an arbitrarily chosen *reference point*. Almost all existing studies of this type of problem adopt the *camera-based interpretation*; i.e., the coordinate origin, or the viewpoint, is chosen as the reference point. This is equivalent to replacing the object motion by camera motion, or *egomotion*. In contrast, we adopt the *object-based interpretation*, taking a reference point on the object surface. The choices are equivalent from a mathematical point of view. However, these two interpretations are not equivalent *from an engineering point of view*.

Suppose the object is located far away from the camera. Then, in order to describe a very small spinning motion, a very large translation velocity must be assigned. This is necessary to compensate for the effect that a small rotation velocity at the origin causes a large displacement if the object is located far away. For robustness of computation, however, the velocity should be specified in such a way that *a small observed change is caused by a small velocity* (cf. Appendix C). Thus, taking the reference on the object surface is the best choice from an engineering point of view, and the subsequent mathematical treatment is no more complicated than for the camera-based interpretation.

The analysis of Longuet-Higgins [18] is embedded into *matrix calculus* based on the 3D geometry of projection. This surely makes it easy to produce a numerical solution. However, it is not so easy to interpret and understand the geometrical meaning of the observed image and its motion. In this paper, the solution of the 3D structure and motion is expressed analytically like that of Subbarao and Waxman [27]. However, the analysis is described in terms of *complex algebra* based on group representation theory. This treatment makes clear the invariant, and hence physical or geometrical, properties of the quantities involved. We construct parameters defining *irreducible representations* of the 2D rotation group $SO(2)$ [6, 32, 33]. These parameters describe invariant characteristics of the observed flow on the image plane.

Using our complex expressions, we can easily derive the *adjacency condition*, which tests whether or not two adjacent optical flows are images of different planar

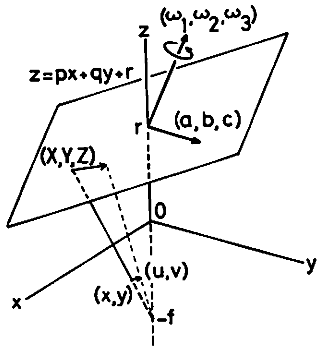

FIG. 1. A plane having equation $z = px + qy + r$ is moving with translation velocity (a, b, c) at $(0, 0, r)$ and rotation velocity $(\omega_1, \omega_2, \omega_3)$ around it. An optical flow is induced on the xy-plane by perspective projection, $(0, 0, -f)$ being the viewpoint.

portions of the same object in rigid motion. This condition also tells us where the boundaries of those planar portions are located, even if these boundaries are not observed on the image plane. The formulation presented in this paper is quite general, and many useful applications can be derived from our analysis, a typical one being the *correspondenceless* approach [9–13].

2. CHARACTERIZATION OF OPTICAL FLOW

As discussed in the previous section, we take an xyz-coordinate system in the scene and regard the xy-plane as the image plane, choosing $(0, 0, -f)$, the point on the z-axis at distance f from the image plane on the negative side, as the viewpoint or camera focus (Fig. 1). A point in the scene is projected to the intersection between the image plane and the ray connecting the point and the viewpoint. Hence, point (X, Y, Z) in the scene is projected to point (x, y) on the image plane, where

$$x = \frac{fX}{f + Z}, \qquad y = \frac{fY}{f + Z}. \tag{2.1}$$

This convention enables us to consider orthographic projection to be simply the limit of $f \to \infty$ without requiring special scaling.

Suppose a plane is moving in the scene, and let $z = px + qy + r$ be its equation.[1] Then p, q designate the gradient of the plane, and r the distance of the plane from the image plane (not the focal point) along the z-axis, which we call the *absolute depth*. The instantaneous rigid motion is specified by the translation velocity (a, b, c) at a *reference point* and the rotation velocity $(\omega_1, \omega_2, \omega_3)$ around it (i.e., with $(\omega_1, \omega_2, \omega_3)$ as the rotation axis and $\sqrt{\omega_1^2 + \omega_2^2 + \omega_3^2}$ (rad/sec) as the angular velocity screwwise around it). While any point can be chosen as the reference point, we choose $(0, 0, r)$, the intersection between the surface and the z-axis. As is mentioned in the previous section, this is the best choice from an engineering point of view (see also Appendix C). Thus, the instantaneous position and motion of the

[1] This parameterization excludes the case where the plane is perpendicular to the image plane. Since we are analyzing the optical flow projected onto the image plane, we are implicitly assuming that the surface gradient is not large.

plane are specified by nine parameters $p, q, r, a, b, c, \omega_1, \omega_2, \omega_3$, which we call the *structure and motion parameters*. The objective here is to recover them by observing the image motion on the image plane.[2]

If the motion is as described above, the *optical flow* $\dot{x} = u(x, y)$, $\dot{y} = v(x, y)$ induced on the image plane is given as follows (Appendix A):

$$\begin{aligned} u(x, y) &= u_0 + Ax + By + (Ex + Fy)x, \\ v(x, y) &= v_0 + Cx + Dy + (Ex + Fy)y. \end{aligned} \tag{2.2}$$

Here, the eight coefficients $u_0, v_0, A, B, C, D, E, F$, which we call the *flow parameters*,[3] are given by

$$\begin{aligned} u_0 &= \frac{fa}{f+r}, & v_0 &= \frac{fb}{f+r}, \\ A &= p\omega_2 - \frac{pa+c}{f+r}, & B &= q\omega_2 - \omega_3 - \frac{qa}{f+r}, \\ C &= -p\omega_1 + \omega_3 - \frac{pb}{f+r}, & D &= -q\omega_1 - \frac{qb+c}{f+r}, \\ E &= \frac{1}{f}\left(\omega_2 + \frac{pc}{f+r}\right), & F &= \frac{1}{f}\left(-\omega_1 + \frac{qc}{f+r}\right). \end{aligned} \tag{2.3}$$

Thus, we are viewing a very restricted form of motion specified by only eight parameters; if the flow parameters are the same, the motions seem identical to the viewer. Hence, *one* of the nine structure and motion parameters $p, q, r, a, b, c, \omega_1, \omega_2, \omega_3$ must remain indeterminate.[4]

If the *focal length* f is sufficiently large compared with the size of the object image, the projection must be regarded as orthographic. This is achieved by simply taking the limit $f \to \infty$, resulting in the *orthographic approximation*

$$\begin{gathered} u_0 = a, \qquad v_0 = b, \\ A = p\omega_2, \qquad B = q\omega_2 - \omega_3, \qquad C = -p\omega_1 + \omega_3, \qquad D = -q\omega_1, \\ E = 0, \qquad F = 0. \end{gathered} \tag{2.4}$$

A complete analysis of this case is given in Kanatani [15].[5]

[2] See footnote 4.

[3] They can be regarded as $u, v, \partial u/\partial x, \partial u/\partial y, \partial v/\partial x, \partial v/\partial y, (1/2)\partial^2 u/\partial x^2$ $(= \partial^2 v/\partial x\, \partial y)$, $(1/2)\partial^2 v/\partial y^2$ $(= \partial^2 u/\partial x\, \partial y)$, respectively, evaluated at the origin (0,0). The flow equations are quadratic in x, y for planar surfaces. For general curved surfaces, however, the flow equations are expressed in an infinite Taylor series (Waxman and Ullman [30]). For curved surfaces, see also Subbarao and Waxman [27] and Subbarao [26].

[4] It is not true that the absolute depth r must *necessarily* be regarded as indeterminate and hence should be removed from the structure and motion parameters. We later say that the absolute depth r is indeterminate. This is because that choice is the most convenient one, not because it is a logical consequence. Indeed, we can regard, say, a as indeterminate and express all the rest (including the absolute depth r) in terms of a. See footnote 8.

[5] We show the explicit analytical solution of Eqs. (2.3) for perspective projection in the form of Eqs. (5.11) and (5.12). However, we cannot obtain the orthographic solution by merely taking the limit $f \to \infty$ in Eqs. (5.11) and (5.12). Most terms become indeterminate in the form of ∞/∞. Hence, we must solve Eqs. (2.4) from the beginning, as is done in Kanatani [15]. In other words, the perspective solution is not well "connected" with the orthographic solution.

On the other hand, if the focal length f is large but not large enough for the orthographic approximation, we may omit terms of $O(1/f^2)$ but not terms of $O(1/f)$. Then, E and F in Eqs. (2.3) are replaced by

$$E = \frac{\omega_2}{f}, \qquad F = -\frac{\omega_1}{f}, \tag{2.5}$$

respectively. We call this approximation the *pseudo-orthographic approximation*.

Suppose the optical flow is already obtained on the image plane by some available means (e.g., [8, 22, 29]). Then, the flow parameters $u_0, v_0, A, B, C, D, E, F$ are estimated by fitting equations of the form (2.2), say by the least square method, minimizing

$$M = \sum_i \Big[\left(u_0 + Ax_i + By_i + (Ex_i + Fy_i)x_i - u(x_i, y_i) \right)^2 + \left(v_0 + Cx_i + Dy_i + (Ex_i + Fy_i)y_i - v(x_i, y_i) \right)^2 \Big], \tag{2.6}$$

where the summation is taken over all feature points belonging to the same planar surface where the velocity is observed.

By computing the "residual" of Eq. (2.6), we obtain a *planarity criterion*; if the resulting M is not less than a prescribed threshold value, the object cannot be regarded as a plane. This also suggests the following segmentation procedure: starting from a small number of feature points where the residual M is very small, add feature points from their vicinity one by one, each time recomputing the flow parameters and checking the residual M, until it reaches a prescribed threshold value. Then, we end up with a region which is regarded as an image of a planar or almost planar part of the object surface. We call such a region a *planar patch*. If this procedure is repeated, the image domain is theoretically segmented into planar patches. (Exact boundaries of these planar patches are not necessary. They are reconstructed by the procedure described in Section 6.)

In practice, however, this process might be unreliable for noisy data, and the resulting segmentation might be affected by the order of adding the points. Hence, other clues and measurements might be necessary to detect moving objects (for actual techniques, see Adiv [1]). Since what we want is the flow parameters, not the flow itself, there also exist methods of estimating the flow parameters directly from a sequence of images without computing the optical flow, i.e., without using the point-to-point correspondence. For example, Waxman and Wohn [30] presented a method of computing the flow parameters from the normal velocities of contour images, and Kanatani [12, 13] computed global "features" and estimated the flow parameters. (See also Kanatani [9–11].)

In this paper, we do not deal with detection techniques. The subsequent analysis is concentrated on recovery of the structure and motion parameters from the flow parameters which are assumed to be already determined.

3. INVARIANT PARAMETERS OF OPTICAL FLOW

Once the flow parameters $u_0, v_0, A, B, C, D, E, F$ are computed from a given optical flow, the structure and motion parameters $p, q, r, a, b, c, \omega_1, \omega_2, \omega_3$ are given as a solution of Eqs. (2.3), which are the only restrictions constraining the solution, for the motions seem identical to the viewer if the flow parameters are the same. An important fact is that an optical flow is described in the form of eqs. (2.3) in reference to an xy-coordinate system on the image plane and that the choice of the coordinate system is completely arbitrary. Suppose we use an $x'y'$-coordinate system obtained by rotating the xy-coordinate system around the z-axis by angle θ counterclockwise. Since we are observing the rigid motion of a plane, the optical flow *must* have the same form

$$\begin{aligned} u' &= u_0' + A'x + B'y' + (E'x + F'y')x', \\ v' &= v_0' + C'x' + D'y' + (E'x' + F'y')y', \end{aligned} \tag{3.1}$$

i.e., the form of the optical flow is *form invariant*.

The old coordinates x, y and the new coordinates x', y' are related by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{3.2}$$

Since the (2D) velocity components are transformed as a (2D) vector, the old components u, v and the new components u', v' are related by

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{3.3}$$

If Eqs. (3.2) and (3.3) are substituted in Eqs. (3.1) and compared with Eqs. (2.2), we find that u_0, v_0 and E, F are transformed *as vectors* and that A, B, C, D are transformed *as a tensor* (Appendix B):

$$\begin{bmatrix} u_0' \\ v_0' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}, \quad \begin{bmatrix} E' \\ F' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} E \\ F \end{bmatrix}, \tag{3.4}$$

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{3.5}$$

Equations (3.4) and (3.5) describe a linear mapping from $u_0, v_0, A, B, C, D, E, F$ onto $u_0', v_0', A', B', C', D', E', F'$, and this mapping is a *representation*, or a *homomorphism*, from the 2D rotation group $SO(2)$ [6, 32]. As is well known in group representation theory, any representation of $SO(2)$ is reduced to one-dimensional *irreducible representations* due to *Schur's lemma*, since $SO(2)$ is a compact Abelian group [6, 32]. In fact, if we define quantities

$$\begin{aligned} &U_0 = u_0 + iv_0, \qquad T = A + D, \qquad R = C - B, \\ &S = (A - D) + i(B + C), \qquad K = E + iF, \end{aligned} \tag{3.6}$$

where i is the imaginary unit, the transformation rule becomes

$$U_0' = e^{-i\theta}U_0, \qquad T' = T, \qquad R' = R, \qquad S' = e^{-2i\theta}S, \qquad K' = e^{-i\theta}K \tag{3.7}$$

(Appendix B). In other words, T, R are (absolute) invariants (of weight 0), U_0, K are (relative) invariants of weight -1, and S is a (relative) invariant of weight -2.

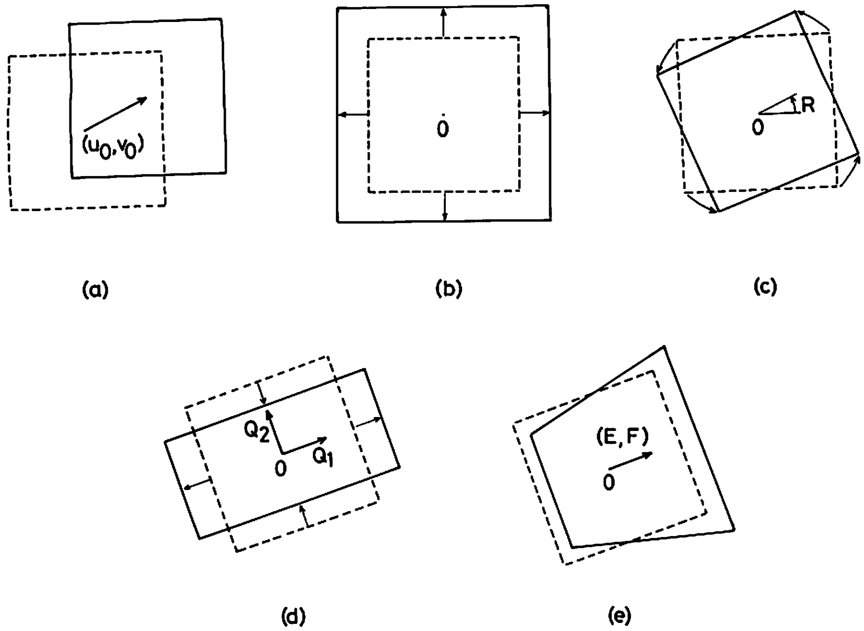

FIG. 2. (a) Translation by (u_0, v_0). (b) Divergence by T. (c) Rotation by R. (d) Shearing with Q_1 and Q_2 as the axes of maximum extension and maximum compression. (e) *Fanning* along (E, F).

Since these quantities are irreducible representations, each of them should have a distinctive geometrical meaning (Weyl [33]). In fact, U_0 represents *translation* (Fig. 2a), T *divergence* (Fig. 2b), R *rotation* (Fig. 2c), S *shearing* (Fig. 2d) and K what we call *fanning* (or *foreshortening* [7]) of the optical flow (Fig. 2e). These quantities are also derived by decomposing the flow according to tensor symmetry properties as was shown in Kanatani [15].[6]

Similarly, the gradient components p, q, the translation velocities a, b and the rotation velocities ω_1, ω_2 are transformed as vectors with respect to the coordinate rotation, while r, c, ω_3 are scalars. Namely,

$$\begin{bmatrix} p' \\ q' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix},$$
$$\begin{bmatrix} a' \\ b' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}, \quad \begin{bmatrix} \omega_1' \\ \omega_2' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}, \tag{3.8}$$
$$r' = r, \qquad c' = c, \qquad \omega_3' = \omega_3.$$

Hence, if we combine them into complex parameters

$$P = p + iq, \qquad V = a + ib, \qquad W = \omega_1 + i\omega_2, \tag{3.9}$$

[6]It should be emphasized that this is the *only* decomposition that has an *invariant* (hence *physical* or *geometrical*) meaning irrespective of the choice of the coordinate system. Other choices such as "stretching" and "shearing" by Hay [7] and $v_x, v_y, e_x, e_{xy}, e_{yy}, e_{xxx}, e_{xxy}, e_{xyy}, e_{yyy}$ by Waxman and Ullman [30] can be given physical or geometrical interpretations only in reference to a particular coordinate system. However, these coordinate-dependent quantities may be sometimes very useful in practical computation. (Invariant quantities are in general complex numbers, as we have shown.)

they are (relative) invariants of weight -1:

$$P' = e^{-i\theta}P, \qquad V' = e^{-i\theta}V, \qquad W' = e^{-i\theta}W. \tag{3.10}$$

4. ANALYTICAL SOLUTION FOR PSEUDO-ORTHOGRAPHIC PROJECTION

If Eqs. (2.3) are substituted in Eqs. (3.6), we obtain

$$\begin{aligned}
U_0 &= \frac{f(a+ib)}{f+r},\\
T &= p\omega_2 - q\omega_1 - \frac{pa+qb+2c}{f+r},\\
R &= -p\omega_1 - q\omega_2 + 2\omega_3 - \frac{pb-qa}{f+r},\\
S &= p\omega_2 + q\omega_1 - \frac{pa-qb}{f+r} + i\left(q\omega_2 - p\omega_1 - \frac{pb+qa}{f+r}\right),\\
K &= \frac{1}{f}\omega_2 + \frac{cp}{f(f+r)} + i\left(-\frac{1}{f}\omega_1 + \frac{cq}{f(f+r)}\right).
\end{aligned} \tag{4.1}$$

If these equations are rewritten in terms of the complex expressions P, V, W of Eqs. (3.10), then T and R in Eqs. (4.1) are combined into one complex equation for $R + iT$ in the form

$$R + iT = -PW^* + 2\omega_3 - \frac{i(PV^* + 2c)}{f+r}, \tag{4.2}$$

where * denotes the complex conjugate. Next, S in Eqs. (4.1) is expressed as

$$S = -iPW - \frac{PV}{f+r}. \tag{4.3}$$

Hence, Eqs. (4.1) are now equivalently rewritten as

$$\begin{aligned}
&U_0 = \frac{f}{f+r}V, \qquad P\left(W^* + \frac{i}{f}U_0^*\right) = (2\omega_3 - R) - i\left(\frac{2c}{f+r} + T\right),\\
&P\left(W - \frac{i}{f}U_0\right) = iS, \qquad K = -\frac{i}{f}W + \frac{cP}{f(f+r)}.
\end{aligned} \tag{4.4}$$

Thus, the eight real equations (2.3) are equivalently reduced to four complex equations in unknowns P, r, V, c, W, ω_3.

If the pseudo-orthographic approximation is applied, then K in Eqs. (4.4) is replaced by

$$K = -\frac{i}{f}W. \tag{4.5}$$

The solution is given as follows.

THEOREM 1. *For the pseudo-orthographic approximation, the structure and motion parameters are given by*

$$V = \frac{f+r}{f}U_0, \qquad P = \frac{S}{fK - U_0/f}, \qquad W = ifK,$$
$$\omega_3 = \frac{1}{2}\left(R + \mathrm{Im}[Se^{-2i\alpha}]\right), \qquad c = \frac{f+r}{2}\left(T - \mathrm{Re}[Se^{-2i\alpha}]\right), \tag{4.6}$$

where

$$\alpha \equiv \arg(fK - U_0/f), \tag{4.7}$$

and arg *denotes the argument of complex numbers. Hence, in general,*[7] (i) *the absolute depth r is indeterminate,*[8] (ii) *$a/(f+r), b/(f+r), c/(f+r)$ are uniquely determined,*[9] *and* (iii) *$p, q, \omega_1, \omega_2, \omega_3$ are uniquely determined. In particular,* no spurious solution exists.

Proof. Equation (4.4.1) gives V in the form of Eq. (4.6.1), and Eq. (4.5) gives W in the form of Eq. (4.6.3), which on substitution in Eq. (4.4.3) gives P in the form of Eq. (4.6.2). Taking the real and the imaginary parts of Eq. (4.4.2), we obtain

$$\omega_3 = \frac{1}{2}\left(R + \mathrm{Re}\left[P\left(W^* + \frac{i}{f}U_0^*\right)\right]\right),$$
$$c = \frac{f+r}{2}\left(T + \mathrm{Im}\left[P\left(W^* + \frac{i}{f}U_0^*\right)\right]\right). \tag{4.8}$$

If we note that

$$P\left(W^* + \frac{i}{f}U_0^*\right) = -iS\frac{fK^* - U_0^*/f}{fK - U_0/f} = -iSe^{-2i\alpha}, \tag{4.9}$$

Eqs. (4.6.4) and (4.6.5) are obtained.

EXAMPLE 1. Consider the flow of Fig. 3 for $f = 2$. The flow parameters are $u_0 = -0.04$, $v_0 = 0.04$, $A = -0.068$, $B = -0.196$, $C = 0.142$, $D = -0.079$, $E = 0.059$, $F = -0.054$. The invariant parameters become $U_0 = -0.04 + 0.04i$, $T = -0.146$, $R = 0.338$, $S = 0.011 - 0.054i$, $K = 0.059 - 0.054i$. Since $|T| > |S| = 0.055$, this flow cannot be regarded as an orthographic projection of a planar motion (Kanatani [15]). In the pseudo-orthographic approximation, we obtain $a/(r+2) = -0.02$, $b/(r+2) = 0.02$, $c/(r+2) = 0.10$, $p = 0.238$, $q = -0.171$, $\omega_1 = 6.19$ (deg/sec), $\omega_2 = 6.76$ (deg/sec), $\omega_3 = 9.88$ (deg/sec), and the absolute depth r is indeterminate. No spurious solution exists.

[7]Indeterminacy occurs when $S = 0$ and $fK = U_0/f$. See also footnote 10.
[8]See footnote 11.
[9]See footnote 12.

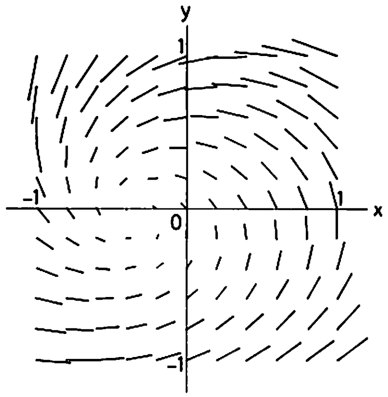

FIG. 3. An example of optical flow.

5. ANALYTICAL SOLUTION FOR PERSPECTIVE PROJECTION

Now, consider the case of perspective projection. If we put

$$c' = \frac{c}{f+r}, \qquad W' = W - \frac{i}{f}U_0, \tag{5.1}$$

Eqs. (4.4) are further reduced to

$$\begin{aligned} V = \frac{f+r}{f}U_0, \qquad PW'^* = (2\omega_3 - R) - i(2c' + T), \\ PW' = iS, \qquad c'P - iW' = fK - \frac{1}{f}U_0. \end{aligned} \tag{5.2}$$

Since V is already explicitly given by the first of Eqs. (5.2), we only need to solve the remining equations for c', P, W', ω_3. First, we must check if $c' = 0$ or not.

LEMMA 1. *We can conclude $c' = 0$ if and only if*

$$\mathrm{Re}[Se^{-2i\alpha}] = T \tag{5.3}$$

is satisfied (within a certain threshold), where α is defined by Eq. (4.7). In this case, the structure and motion parameters are given as follows[10]:

$$\begin{aligned} V = \frac{f+r}{f}U_0, \qquad c = 0, \\ P = \frac{S}{fK - U_0/f}, \qquad W = ifK, \qquad \omega_3 = \frac{1}{2}\left(R + \mathrm{Im}[Se^{-2i\alpha}]\right). \end{aligned} \tag{5.4}$$

[10] Indeterminacy occurs when $T = 0$, $S = 0$ and $fK = U_0/f$. In this case, P is completely indeterminate. In fact, it can be shown (Kanatani [14]) that this occurs if and only if the surface is *orbiting* around the camera focus, always keeping the configuration relative to the camera focus rigidly fixed. In other words, the configuration of the *rays* from the camera focus to the feature points is always the same. Hence, no information is obtained about the position and orientation of the surface, i.e., the optical flow is *uninformative*.

Proof. If $c' = 0$, Eq. (5.2.4) gives W as Eq. (5.4.4), and then Eq. (5.2.3) gives P as Eq. (5.4.3). These are solutions if and only if they satisfy the imaginary part of Eq. (5.2.3), or equivalently Eq. (5.3), and Eq. (5.4.5) is obtained by taking the real part of Eq. (5.2.3).

Now, suppose we have already checked that c' is not zero. Noting that Eq. (5.2.3) is rewritten as $(c'P)(-iW') = c'S$, we see that Eqs. (5.2.3) and (5.2.4) imply that $c'P,\ -iW'$ are the two roots of the quadratic equation

$$X^2 - LX + c'S = 0, \tag{5.5}$$

where

$$L \equiv fK - U_0/f. \tag{5.6}$$

Hence, P, W' are given as functions of c' by

$$P(c') = \frac{1}{2c'}\left(L \pm \sqrt{L^2 - 4c'S}\right), \qquad W'(c') = \frac{i}{2}\left(L \mp \sqrt{L^2 - 4c'S}\right). \tag{5.7}$$

Taking the real and the imaginary parts of Eq. (5.2.2), we obtain

$$\omega_3 = \frac{1}{2}\left(R + \mathrm{Re}[P(c')W'(c')^*]\right),\ c' = -\frac{1}{2}\left(T + \mathrm{Im}[P(c')W'(c')^*]\right). \tag{5.8}$$

If Eqs. (5.7) are substituted in Eqs. (5.8), they become

$$\begin{gathered}\omega_3 = \frac{1}{2}R \pm \frac{1}{4c'}\mathrm{Im}\left[L^*\sqrt{L^2 - 4c'S}\right], \\ \sqrt{16|S|^2c'^2 - 8\,\mathrm{Re}[L^2S^*]c' + |L|^4} = -8c'^2 - 4Tc' + |L|^2.\end{gathered} \tag{5.9}$$

The second of Eqs. (5.9) is the equation to determine c'. If c' is determined, then P, W', ω_3 are given by Eqs. (5.7.1), (5.7.2) and the first of Eq. (5.9), respectively.

Now, the left-hand side of the second of Eqs. (5.9) is a smooth concave function (or constant if $S = 0$) passing through $(0, |L|^2)$, while the right-hand side is a smooth convex quadratic function also passing through $(0, |L|^2)$ (Fig. 4). Since we know that the solution is nonzero, we can see from Fig. 4 that there exists a single unique nonzero solution c'. Taking the squares of both sides and dropping off c' from both sides, we obtain

$$c'^3 + Tc'^2 + \frac{1}{4}\left(T^2 - |S|^2 - |L|^2\right)c' + \frac{1}{8}\left(\mathrm{Re}[L^2S] - T|L|^2\right) = 0. \tag{5.10}$$

From Fig. 4, we can easily see that Eq. (5.10) has three real roots and that the middle one is the desired root. The other two roots were introduced by squaring of both sides. Thus, we obtain the following result.

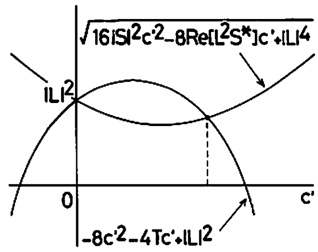

FIG. 4. Existence and uniqueness of nonzero c'.

THEOREM 2. *If c is known to be nonzero, the solution is given as follows. Let c' be the middle root of the cubic equation*

$$X^3 + TX^2 + \frac{1}{4}(T^2 - |S|^2 - |L|^2)X + \frac{1}{8}(\mathrm{Re}[L^2S] - T|L|^2) = 0. \quad (5.11)$$

Then the structure and motion parameters are given by

$$V = \frac{f+r}{f}U_0, \qquad c = (f+r)c', \qquad P = \frac{1}{2c'}\left(L \pm \sqrt{L^2 - 4c'S}\right),$$
$$W = \frac{i}{2}\left(L \mp \sqrt{L^2 - 4c'S}\right) + \frac{i}{f}U_0, \quad (5.12)$$
$$\omega_3 = \frac{1}{2}R \pm \frac{1}{4c'}\mathrm{Im}\left[L^*\sqrt{L^2 - 4c'S}\right].$$

Thus, (i) *the absolute depth r is indeterminate,*[11] (ii) $a/(f+r)$, $b/(f+r)$, $c/(f+r)$ *are uniquely determined,*[12] *and* (iii) *two sets of solutions exist for* p, q, ω_1, ω_2, ω_3.

Since an explicit form of the solution of a cubic equation exists, we can express the solution c' of Eq. (5.11) explicitly, as is done by Subbarao and Waxman [27]. However, from a practical point of view, the use of iteration, say the Newton–Raphson method, to solve the second of Eqs. (5.9) seems more feasible, because the equation is known to behave as shown in Fig. 4. Rough bounds of c', ω_3 are given as follows.

Remark. The solutions for c', ω_3 are bounded by

$$-\frac{1}{2}(|S| + T) \le c' \le \frac{1}{2}(|S| - T), \qquad \frac{1}{2}(R - |S|) \le \omega_3 \le \frac{1}{2}(R + |S|). \quad (5.13)$$

[11] Indeterminacy of the absolute depth r has been well known, but, strictly speaking, the absolute depth r is treated as an indeterminate just because this is the most convenient choice, for r is an absolute invariant and this choice makes the forms of the translation velocities a, b, c the same. However, this is *not* a unique logical consequence. For example, we can also choose a as an indeterminate, determining r in the form $r = f(a/u_0 - 1)$, and we can assert that b/a, c/a are uniquely determined. In this case, a is a single indeterminate. Similarly, we can choose b or c as an indeterminate.

[12] This may seem contradictory, since in the camera-based interpretation the translation velocity scaled by the absolute depth is not uniquely determined (e.g., [5, 18, 20, 27, 30]). However, as stated in Section 2, any 3D rigid motion is represented by the translation velocity at a *reference point* and the rotation velocity around it, and this decomposition of a motion into the translation part and the rotation part is different if the reference point is different. In fact, the translation in the camera-based interpretation is a linear combination of our translation and rotation in the object-based interpretation. Thus, the distinction between the camera-based interpretation and the object-based interpretation is very important. See also Appendix C.

Proof. Since $|PW'^*| = |PW'|$, we see from Eqs. (5.2.2) and (5.2.3) that

$$(2\omega_3 - R)^2 + (2c' + T)^2 = |S|^2. \tag{5.14}$$

In other words, point (ω_3, c') is on the circle of center $(R/2, -T/2)$ and of radius $|S|/2$ in the two-dimensional plane.

EXAMPLE 2. Consider again the flow of Fig. 3. We obtain $a/(r+2) = -0.02$, $b/(r+2) = 0.02$, $c/(r+2) = 0.10$, and two sets of solutions (i) $p = 0.300$, $q = -0.200$, $\omega_1 = 5.00$ (deg/sec), $\omega_2 = 5.00$ (deg/sec), $\omega_3 = 10.00$ (deg/sec) and (ii) $p = 1.073$, $q = -1.073$, $\omega_1 = 0.00$ (deg/sec), $\omega_2 = 0.57$ (deg/sec), $\omega_3 = 9.39$ (deg/sec). The absolute depth r is indeterminate.

Note that solution (i) is close to the pseudo-orthographic solution of Example 1. The geometrical interpretation of these two solutions is as follows. As is seen in Fig. 2e, the *fanning* K is caused by rotation W around an axis parallel to the image plane. However, as is seen in the last of Eqs. (4.4), this deformation is also caused by the velocity c along the z-axis when the focal length f and the absolute depth r are not large compared with the gradient P. Intuitively speaking, the translation along the line of sight *mimics* the effect of rotation when the surface is too near to the viewer and the projective distortion is too large. The pseudo-orthographic approximation of Eq. (4.5) removes this effect, eliminating the spurious solution.

Thus, if the surface is known to be far away from the viewer compared with the camera focal length f and the projective distortion is not so large, we can identify with the true solution the one which corresponds to the pseudo-orthographic solution. In the above example, solution (i) is the desired one in the sense above mentioned.

The fact that an optical flow of planar motion yields, other than the true solution, a spurious (or *confusable* [7]) solution was pointed out first by Hay [7] and later by Longuet-Higgins [18] and Subbarao and Waxman [27]. Hence, it seems that one cannot distinguish the true solution from the spurious one, for both solutions yield exactly the same optical flow. However, our analysis makes clear the geometrical meaning of the spurious solution, making it possible to distinguish the true one from the spurious one in the sense described above.

Since the flow is characterized by eight flow parameters, the minimum number of feature points whose velocities must be observed in order to determine the flow uniquely is four. It is easily checked that the flow parameters are uniquely determined if velocities are measured at four coplanar points no three of which are collinear.

EXAMPLE 3. Suppose velocities are measured at four points (0.4, 0.2), (−0.4, 0.4), (−0.2, −0.4), (0.6, −0.2) for $f = 2$, resulting in (−0.105, −0.015), (0.036, 0.067), (−0.085, 0.112), (−0.196, −0.011), respectively (Fig. 5). The corresponding flow is given by flow parameters $u_0 = -0.037$, $v_0 = 0.023$, $A = -0.162$, $B = 0.124$, $C = -0.120$, $D = -0.080$, $E = 0.019$, $F = 0.059$ (Fig. 6). The procedure shown previously yields $a/(r+2) = -0.14$, $b/(r+2) = 0.09$, $c/(r+2) = 0.40$ and two sets of solutions: (i) $p = 1.24$, $q = 0.65$, $\omega_1 = -3$ (deg/sec), $\omega_2 = -5$ (deg/sec), $\omega_3 = -9$ (deg/sec) and (ii) $p = -0.50$, $q = 0.30$, $\omega_1 = -5$ (deg/sec), $\omega_2 = 5$

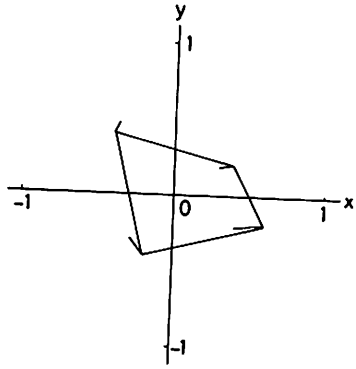

FIG. 5. The velocity is observed at four coplanar points of a rigid object.

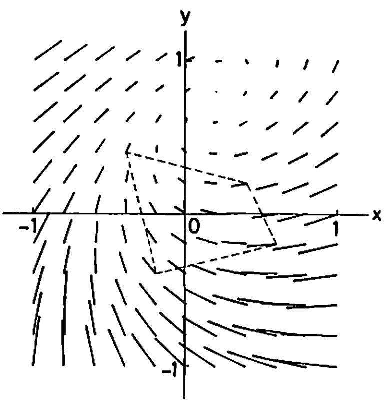

FIG. 6. The hypothetical optical flow computed from the velocities at the four points in Fig. 5.

(deg/sec), $\omega_3 = -5$ (deg/sec). The z-coordinates of the four vertices are obtained by substituting the xy-coordinates in the last of Eqs. (A.2), resulting in (i) $z_1 = 0.91 + 1.46r$, $z_2 = -0.21 + 0.89r$, $z_3 = -0.41 + 0.80r$, $z_4 = 0.89 + 1.44r$, or (ii) $z_1 = -0.13 + 0.93r$, $z_2 = 0.38 + 1.19r$, $z_3 = -0.02 + 0.99r$, $z_4 = -0.31 + 0.85r$. Thus, the absolute depth r is indeterminate.

However, this process may be very sensitive to noise and misdetection of correspondence. If we want to obtain accurate estimates of the flow parameters, the velocity must be observed at many feature points of one planar patch.

6. ADJACENCY OF OPTICAL FLOWS

So far, only one planar patch has been considered on the image plane. Here, let us consider the relationship among planar patches when they are images of one and the same rigid object. First, note that Eqs. (2.2) are combined into a single equation in complex variables of the form

$$U = U_0 + \frac{1}{2}\left((T + iR)Z + SZ^* + KZZ^* + K^*Z^2\right), \tag{6.1}$$

where $U = u + iv$, $Z = x + iy$. Let U_0', T', R', S', K' be the parameters for another patch. If the two patches correspond to two planes of the same object, the induced optical flows must be continuous over the *intersection line* (or, to be precise, the image of the intersection line of the two planes). In other words, at any point (x, y) on the intersection line, which may or may not appear on the image plane, we have the relation

$$(2[U] =)2[U_0] + [T + iR]Z + [S]Z^* + [K]ZZ^* + [K]^*Z^2 = 0, \quad (6.2)$$

where [ ] designates the difference, e.g., $[U_0] = U_0' - U_0$. Since Eq. (6.2) holds at any point on the intersection line, it must be the actual equation of the intersection line.

First, consider the case of $[K] \neq 0$. Equation (6.2) is rewritten as

$$([K]Z + [S])Z^* + ([K]^*Z^2 + [T + iR]Z + 2[U_0]) = 0. \quad (6.3)$$

This equation reduces to two linear equations if and only if polynomial $[K]^*Z^2 + [T + iR]Z + 2[U_0]$ in Z is divisible by $[K]Z + [S]$. By the well-known "remainder theorem," the necessary and sufficient condition for that is

$$[K]^*[S]^2 - [T + iR][S][K] + 2[U_0][K]^2 = 0. \quad (6.4)$$

If this condition is satisfied, Eq. (6.3) is factored into

$$([K]Z + [S])\left(Z^* + \frac{[K]^*}{[K]}Z + 2\frac{[U_0]}{[S]}\right) = 0. \quad (6.5)$$

Since $[K]Z + [S] = 0$ describes one point, the intersection line must be

$$Z^* + \frac{[K]^*}{[K]}Z + 2\frac{[U_0]}{[S]} = 0. \quad (6.6)$$

In terms of x, y, this becomes

$$[E]x + [F]y + \frac{[U_0][K]}{[S]} = 0. \quad (6.7)$$

This is an equation of a line if and only if $[U_0][K]/[S]$ is a real number, namely if and only if

$$\mathrm{Im}([U_0][K][S]^*) = 0 \quad \text{or} \quad \arg([U_0]) + \arg([K]) = \arg([S]) \pmod{\pi}. \quad (6.8)$$

We say that two optical flows are *adjacent* if the corresponding surfaces (extended if necessary) have a fixed intersection (or are "hinged together"). Then, from the above result, we obtain

LEMMA 2. *For $[K] \neq 0$, the necessary and sufficient condition that two optical flows are adjacent is*

$$[K]^*[S]^2 - [T + iR][S][K] + 2[U_0][K]^2 = 0,$$
$$\text{Im}([U_0][K][S]^*) = 0 \text{ or } \arg([U_0]) + \arg([K]) = \arg([S]) \ (\text{mod } \pi). \quad (6.9)$$

If this condition is satisfied, the equation of the intersection line is given by

$$[E]x + [F]y + \frac{[U_0][K]}{[S]} = 0. \quad (6.10)$$

If $[K] = 0$, then Eq. (6.3) is linear in x, y and is rewritten as

$$([u_0] + i[v_0]) + ([A] + i[C])x + ([B] + i[D])y = 0. \quad (6.11)$$

Hence, we have

LEMMA 3. *For $[K] = 0$, the necessary and sufficient condition that two optical flows are adjacent is*

$$[u_0] : [v_0] = [A] : [C] = [B] : [D]. \quad (6.12)$$

If this condition is satisfied, the equation of the intersection line is given by

$$[u_0] + [A]x + [B]y = 0 \quad \text{or} \quad [v_0] + [C]x + [D]y = 0. \quad (6.13)$$

In other words, if Eq. (6.9) or (6.12) is not satisfied, the two patches are images of two different independently moving objects, while if they are satisfied, the intersection line is immediately obtained even if it does not appear on the image plane. Thus, once portions of planar patches are detected on the image plane, the exact boundaries between them are completely determined as long as all the patches belong to the same object, and there is no need for edge detection.

However, adjacency does not necessarily mean "rigid connection," which must be determined by analyzing each planar patch and recovering $a/(f+r)$, $b/(f+r)$, $c/(f+r)$, and two sets of $p, q, \omega_1, \omega_2, \omega_3$. If the motion is rigid, then $\omega_1, \omega_2, \omega_3$ must be common to both patches, and hence we can pick out the true solution for each patch. If common $\omega_1, \omega_2, \omega_3$ are not found, the corresponding planar surface cannot be rigidly connected.

For the connection of two planes, we obtain the following important observation:

LEMMA 4. *If the intersection line on the image plane is $y = mx + n$ (obtained by Lemmas 2 and 3) and the equations of the corresponding planar surfaces are $z = px + qy + r$, $z = p'x + q'y + r'$, the* relative depth $[r] = r' - r$ *is given by*

$$[r] = \frac{(f+r)n}{fm + pn}[p] = \frac{(f+r)n}{nq - f}[q]. \quad (6.14)$$

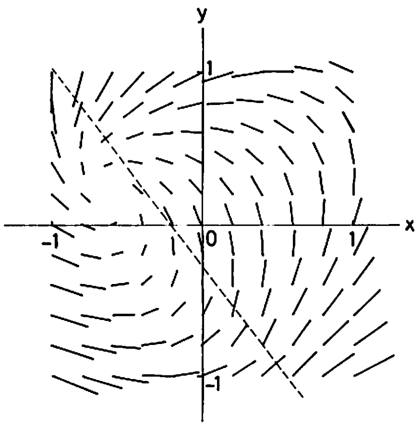

FIG. 7. This flow cannot be regarded as a single flow of a planar patch. There exist two planar patches, and the dashed line is the intersection computed from the flow.

Proof. If a point (X, Y, Z) is on the intersection line of the two planes $z = px + qy + r$, $z = p'x + q'y + r'$, it satisfies both of these equations, so that $[p]X + [q]Y + [r] = 0$ is satisfied. Substituting the first and the second of Eqs. (A.2) in Appendix A, we obtain the equation of the intersection line on the image plane in the form

$$((f + r)[p] - p[r])x + ((f + r)[q] - q[r])y + f[r] = 0. \qquad (6.15)$$

Comparison of this equation with $y = mx + n$ yields Eq. (6.14).

Lemma 4 states that if the absolute depth is known or assumed for one patch, the depths for all other patches are uniquely determined. In conclusion, we have

THEOREM 3. *The structure and motion of an object whose surface consists of, or is approximated by, planar faces are uniquely determined from its optical flow under perspective projection only up to a single indeterminate absolute depth r.*

This fact[13] was first found by Hay [7] and later rediscovered by many people in many different forms, though the analytical result of Lemmas 2 and 3 is not found elsewhere.

EXAMPLE 4. Consider the flow of Fig. 7 for $f = 2$. The flow as whole does not satisfy the planarity criterion discussed in Section 2, and hence it cannot be regarded as a single flow. For the upper right part, the flow parameters are estimated to be $u_0 = -0.061$, $v_0 = 0.126$, $A = 0.003$, $B = -0.134$, $C = 0.056$, $D = -0.148$, $E = 0.112$, $F = -0.077$ and for the lower left part $u_0' = -0.097$, $v_0' = 0.167$, $A' = -0.176$, $B' = -0.264$, $C' = 0.252$, $D' = -0.006$, $E' = 0.071$,

[13]Strictly speaking, the solution may not be unique if the optical flows of *all* the planar patches happen to have two identical sets of ω_1, ω_2, ω_3. However, it is known that this type of ambiguity occurs only for a special type of quadric surface (cf. Fang and Huang [5], Tsai and Huang [28], Maybank [20] and Subbarao [26]).

$F' = -0.109$. The adjacency condition (Eqs. (6.9)) is satisfied within rounding error, and the intersection line is estimated to be $y = -1.30x - 0.27$, which is indicated by a broken line in the figure. For the upper right part, we obtain $\omega_1 = 10.0, -4.7$ (deg/sec), $\omega_2 = 10.0, 1.1$ (deg/sec), $\omega_3 = 10.0, 0.9$ (deg/sec), and for the lower left part $\omega_1 = -2.3, 10.0$ (deg/sec), $\omega_2 = -4.6, 10.0$ (deg/sec), $\omega_3 = 19.5, 10.0$ (deg/sec). Hence, the true solution is $\omega_1 = 10.0$, $\omega_2 = 10.0$, $\omega_3 = 10.0$ (deg/sec). The gradients are $p = 0.5$, $q = 0.2$ for the upper right part and $p' = -0.3$, $q' = -0.4$ for the lower left part. The equations of the two planes are $z = 0.5x + 0.2y + r$, $z = -0.3x - 0.4y + (0.92r - 0.16)$, respectively.

7. CONCLUDING REMARKS

We have presented a complete analysis of the optical flow resulting from planar surface motion under perspective projection and expressed the solution in analytical closed form including the spurious solution. The underlying principle is invariance with respect to coordinate changes on the image plane, and the equations are written in terms of invariants, which are in general complex numbers, based on group representation theory. The adjacency condition of optical flows and the equation of intersection are obtained in simple forms by means of complex algebra. Hence, analyzing each planar patch, we can reconstruct the structure and shape of the object whose surface consists of, or is approximated by, planar faces.

It should be pointed out that explicit analytical solutions are obtained for both perspective and pseudo-orthographic projections not because complex quantities are introduced. In fact, all complex equations can be rewritten in terms of real quantities alone, separating the real and the imaginary parts, or the same analytical solutions can be attained without using complex algebra or matrix calculus from the beginning, as was done by Waxman and Ullman [30] and Subbarao and Waxman [27]. The merit of complex expressions is that they make clear the invariant properties with respect to coordinate changes and that various types of analysis, including error sensitivity and the adjacency condition, become very easy.[14]

We have also emphasized the importance of a good parameterization of the underlying geometry such as the choice of the reference point and the position of the viewpoint. This is closely related to the *robustness* of parameter estimation. Adiv [2], for example, claims that *any* algorithm of 3D recovery is "inherently ambiguous" because, among other things, "translational flows" and "rotational flows" have very similar patterns, and hence it is very difficult to distinguish them. However, translational flows and rotational flows look quite different in our object-based interpretation. Thus, his assertion is not "algorithm independent." There are many factors that come into play (Appendix C). One of them is a *bad* parameterization; one cannot expect robustness if the problem is parameterized in such a way that large changes induce small observed effects.

Our parameterization also makes it possible to treat orthographic projection and perspective projection in the same framework. Moreover, we can obtain the *pseudo-orthographic approximation*, for which no spurious solution exists. This enables us to distinguish the true solution from the spurious one for perspective projection of planar motion. The geometrical interpretation for this is also given.

[14] See footnote 15.

In addition, the pseudo-orthographic solution is expected to be more robust than the exact solution. As Adiv [1, 2] points out, estimation of E, F (hence $K = E + iF$) of Eqs. (2.2) may be difficult and vulnerable to errors. However, as is seen from Eqs. (4.6), K enters the solution of the pseudo-orthographic approximation in a very straightforward way, so that it is expected that a small error in K will cause only small errors in the solution.[15]

In our analysis, all the computations are done on the flow parameters extracted from the optical flow. Therefore, the optical flow is not necessary if the flow parameters can be estimated. This idea leads to detection of structure and motion without correspondence and is fully studied by Kanatani [12, 13], who computed "features" to estimate the flow parameters.

APPENDIX A: Equations of Optical Flow

If a point (X, Y, Z) in the scene is on the plane $z = px + qy + r$, there is a one-to-one correspondence between the point (X, Y, Z) in the scene and its projection (x, y) on the image plane. In fact, solving Eqs. (2.1) and

$$Z = pX + qY + r \tag{A.1}$$

simultaneously for X, Y, Z, we obtain

$$X = \frac{(f+r)x}{f - px - qy}, \qquad Y = \frac{(f+r)y}{f - px - qy}, \qquad Z = \frac{f(px + qy + r)}{f - px - qy}. \tag{A.2}$$

which, together with Eqs. (2.1), establish the one-to-one correspondence between (X, Y, Z) and (x, y).

In our object-based interpretation (Fig. 1), the velocity of point (X, Y, Z) in the scene is given by

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z - r \end{bmatrix}. \tag{A.3}$$

Substituting Eq. (A.1) for Z, we obtain

$$\begin{aligned} \dot{X} &= a + p\omega_2 X + (q\omega_2 - \omega_3)Y, \\ \dot{Y} &= b + (\omega_3 - p\omega_1)X - q\omega_1 Y, \\ \dot{Z} &= c - \omega_2 X + \omega_1 Y. \end{aligned} \tag{A.4}$$

Differentiating both sides of Eqs. (2.1), we obtain the velocity of the image point

[15] From Eq. (4.6.1), we see that V is not affected by the error in K. From Eqs. (4.6.2), we see that P is stably computed unless $fK - U_0/f$ is very close to zero. From Eq. (4.6.3), we see that W is not so greatly affected by the error in K. From Eq. (4.7), we see that α is not so greatly affected by the error in K unless $fK - U_0/f$ is very close to zero. From Eqs. (4.6.4) and (4.6.5), we see that the errors in ω_3 and c due to the error in K are proportional to the magnitude $|S|$ of the shearing. In sum, the computation is stable except when $fK - U_0/f$ is very close to zero or S has a very large magnitude. This exceptional case is exactly when the surface gradient P has a very large magnitude, which is intuitively understandable as well.

(x, y) as follows:

$$\dot{x} = \frac{f\dot{X}}{f+Z} - \frac{fX\dot{Z}}{(f+Z)^2} = \frac{f\dot{X}}{f+Z} - \frac{x\dot{Z}}{f+Z},$$
$$\dot{y} = \frac{f\dot{Y}}{f+Z} - \frac{fY\dot{Z}}{(f+Z)^2} = \frac{f\dot{Y}}{f+Z} - \frac{y\dot{Z}}{f+Z}. \tag{A.5}$$

From Eqs. (A.4) and Eqs. (2.1), we see that

$$\frac{f\dot{X}}{f+Z} = \frac{fa}{f+Z} + p\omega_2 x + (q\omega_2 - \omega_3)y,$$
$$\frac{f\dot{Y}}{f+Z} = \frac{fb}{f+Z} + (\omega_3 - p\omega_1)x - q\omega_1 y, \tag{A.6}$$
$$\frac{f\dot{Z}}{f+Z} = \frac{fc}{f+Z} - \omega_2 x + \omega_1 y.$$

Substituting these in Eqs. (A.5) and eliminating $1/(f+Z)$ by

$$\frac{1}{f+Z} = \frac{f - px - qy}{f(f+r)}, \tag{A.7}$$

which is obtained from the last of Eqs. (A.2), we obtain the result of Eqs. (2.2) and (2.3).

APPENDIX B: Transformation of Optical Flow

Let us define vectors

$$\mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix}, \quad \mathbf{u}_0 = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} E \\ F \end{bmatrix}, \tag{B.1}$$

and a matrix

$$\mathbf{A} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}. \tag{B.2}$$

Let us also denote the rotation matrix in Eqs. (3.2) and (3.3) by $\mathbf{R}(\theta)$. Then, the optical flows of Eqs. (2.2) and (3.1) are expressed as

$$\mathbf{u} = \mathbf{u}_0 + \mathbf{A}\mathbf{x} + (\mathbf{k}, \mathbf{x})\mathbf{x}, \qquad \mathbf{u}' = \mathbf{u}'_0 + \mathbf{A}'\mathbf{x}' + (\mathbf{k}', \mathbf{x}')\mathbf{x}', \tag{B.3}$$

respectively, where $(\cdot, \cdot)$ denotes the inner product.

Equations (3.2) and (3.3), respectively, become

$$\mathbf{x}' = \mathbf{R}(\theta)\mathbf{x}, \qquad \mathbf{u}' = \mathbf{R}(\theta)\mathbf{u}. \tag{B.4}$$

Hence, from u of Eqs. (B.3), we obtain

$$\begin{aligned}\mathbf{u}' &= \mathbf{R}(\theta)(\mathbf{u}_0 + \mathbf{A}\mathbf{x} + (\mathbf{k},\mathbf{x})\mathbf{x}) \\ &= \mathbf{R}(\theta)\left(\mathbf{u}_0 + \mathbf{A}\mathbf{R}(\theta)^{\mathrm{T}}\mathbf{x}' + \left(\mathbf{k},\mathbf{R}(\theta)^{\mathrm{T}}\mathbf{x}'\right)\mathbf{R}(\theta)^{\mathrm{T}}\mathbf{x}'\right) \\ &= \mathbf{R}(\theta)\mathbf{u}_0 + \mathbf{R}(\theta)\mathbf{A}\mathbf{R}(\theta)^{\mathrm{T}}\mathbf{x}' + (\mathbf{R}(\theta)\mathbf{k},\mathbf{x}')\mathbf{x}'.\end{aligned} \tag{B.5}$$

Note that $\mathbf{R}(\theta)^{-1} = \mathbf{R}(\theta)^{\mathrm{T}}$ and $(\mathbf{R}(\theta)\,\cdot\,,\,\cdot) = (\cdot\,,\mathbf{R}(\theta)^{\mathrm{T}}\,\cdot\,)$. Comparing this with $\mathbf{u}'$ of Eqs. (B.3), we find

$$\mathbf{u}_0' = \mathbf{R}(\theta)\mathbf{u}_0, \qquad \mathbf{A}' = \mathbf{R}(\theta)\mathbf{A}\mathbf{R}(\theta)^{\mathrm{T}}, \qquad \mathbf{k}' = \mathbf{R}(\theta)\mathbf{k}. \tag{B.6}$$

Thus, Eqs. (3.4) and (3.5) are obtained.

Equations (B.6) describe a linear mapping from $u_0, v_0, A, B, C, D, E, F$ onto $u_0', v_0', A', B', C', D', E', F'$. If we pick out the vector and tensor components, Eqs. (B.6) are rearranged into the following form:

$$\begin{bmatrix} u_0' \\ v_0' \\ A' \\ B' \\ C' \\ D' \\ E' \\ F' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & & & & & & \\ -\sin\theta & \cos\theta & & & & & & \\ & & \cos^2\theta & \cos\theta\sin\theta & \cos\theta\sin\theta & \sin^2\theta & & \\ & & -\cos\theta\sin\theta & \cos^2\theta & -\sin^2\theta & \cos\theta\sin\theta & & \\ & & -\cos\theta\sin\theta & -\sin^2\theta & \cos^2\theta & \cos\theta\sin\theta & & \\ & & \sin^2\theta & -\cos\theta\sin\theta & -\cos\theta\sin\theta & \cos^2\theta & & \\ & & & & & & \cos\theta & \sin\theta \\ & & & & & & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ A \\ B \\ C \\ D \\ E \\ F \end{bmatrix}. \tag{B.7}$$

This linear transformation is a *representation*, or a *homomorphism*, from the 2D rotation group $SO(2)$, but it is reducible. The matrix is diagonalized if the flow parameters are rearranged as follows:

$$\begin{bmatrix} u_0' + iv_0' \\ u_0' - iv_0' \\ A' + D' \\ B' - C' \\ (A' - D') + i(B' + C') \\ (A' - D') - i(B' + C') \\ E' + iF' \\ E' - iF' \end{bmatrix} = \begin{bmatrix} e^{-i\theta} & & & & & & & \\ & e^{i\theta} & & & & & & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & e^{-2i\theta} & & & \\ & & & & & e^{2i\theta} & & \\ & & & & & & e^{-i\theta} & \\ & & & & & & & e^{i\theta} \end{bmatrix} \begin{bmatrix} u_0 + iv_0 \\ u_0 - iv_0 \\ A + D \\ B - C \\ (A - D) + i(B + C) \\ (A - D) - i(B + C) \\ E + iF \\ E - iF \end{bmatrix}. \tag{B.8}$$

Thus, the representation of Eq. (B.6) is reduced to the direct sum of one-dimensional irreducible representations. This is possible due to Schur's lemma, because

$SO(2)$ is a compact Abelian group [6, 32]. Hence, each of these new parameters should have an invariant, hence *physical* or *geometrical*, meaning (Weyl [33]). This reduction is also related to the decomposition of a tensor with respect to symmetry familiar in fluid dynamics (cf. Kanatani [15]).

APPENDIX C: ROBUSTNESS OF ESTIMATION

We generalize the problem a little so that the underlying mathematical structure becomes clear. As stated earlier, the rigid motion is represented by the (3D) translation velocity at a *reference point* and the (3D) rotation velocity around it. The decomposition is different if the reference point is different. We called it the *camera-based interpretation* when the reference point is at the viewpoint or camera focus, and the *object-based interpretation* when it is taken in the object.

Let us take a reference point (x_0, y_0, z_0) arbitrarily. Suppose, then, the rigid motion is represented by translation velocities a, b, c and rotation velocities $\omega_1, \omega_2, \omega_3$. This is one interpretation of the object motion. If we take another reference point (x_0', y_0', z_0') and extract translation velocities a', b', c', they are linear combinations of $a, b, c, \omega_1, \omega_2, \omega_3$ in the form

$$\begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{bmatrix} \times \begin{bmatrix} x_0' - x_0 \\ y_0' - y_0 \\ z_0' - z_0 \end{bmatrix}. \tag{C.1}$$

(The rotation velocities $\omega_1, \omega_2, \omega_3$ are the same.) As is well known, any two interpretations are mathematically equivalent, because we can move from one interpretation to another freely by Eq. (C.1).

Let us choose one interpretation arbitrarily. Consider the motion where $a = 1$ and all the other velocity components are zero. Let us symbolically denote by $U_a(x, y)$ the optical flow induced on the image plane by this motion. Similarly, define the optical flows $U_b(x, y)$, $U_c(x, y)$, $U_{\omega_1}(x, y)$, $U_{\omega_2}(x, y)$, $U_{\omega_3}(x, y)$, each representing a "pure" motion *in this interpretation*. (Flows $U_a(x, y)$, $U_b(x, y)$, $U_c(x, y)$ are independent of interpretation.)

Let $U(x, y)$ be the observed optical flow of the object in unknown motion. The motion parameters can be determined if we can express the optical flow $U(x, y)$ as a linear combination of the "pure" component flows:

$$\begin{aligned} U(x, y) = {} & aU_a(x, y) + bU_b(x, y) + cU_c(x, y) \\ & + \omega_1 U_{\omega_1}(x, y) + \omega_2 U_{\omega_2}(x, y) + \omega_3 U_{\omega_3}(x, y). \end{aligned} \tag{C.2}$$

The coefficients $a, b, c, \omega_1, \omega_2, \omega_3$ of this linear combination give the motion parameters in this interpretation.

If the problem is stated as above, we can observe the following. From a mathematical point of view, any interpretation is equivalent as long as the component flows are *linearly independent*. However, there are other things that must be taken into account *from an engineering point of view*. As is well known in many similar circumstances, in order that the coefficients of a linear decomposition are robustly computed, the component flows $U_a(x, y), \ldots, U_{\omega_3}(x, y)$ must be not only linearly independent but also "very different" from each other. If some of them are "very similar," the coefficients cannot be robustly computed. Also, the component flows must have almost the same "magnitude" in some sense. Alternatively, we can also say that a *small observed change must be caused by a small (3D) velocity.* In other words, computation is not robust if some large velocity components result in a small observed flow by cancellation due to subtraction or by contraction. This occurs if the component flows are "nearly linearly dependent" or some component flows are "disproportionately small". In mathematical terms, the most desirable case is when the component flows form an *orthonormal system*, each having "unit magnitude" and being "orthogonal" to each other, by introducing the *inner product* between two flows in some natural manner.

On the other hand, if the object is a planar surface, the optical flow always has the form of Eq. (2.2), no matter what interpretation we may adopt. This flow is naturally decomposed into "(2D) translation," "(2D) divergence," "(2D) rotation," "(2D) shearing," "(2D) fanning," as described in Section 3. This decomposition is based on the invariance properties obtained through group theoretical consideration, and is quite natural. Indeed, if we define the inner product of two flows $u_1(x, y)$, $v_1(x, y)$ and $u_2(x, y)$, $v_2(x, y)$ by

$$(U_1, U_2) \equiv \iint_W [u_1(x, y)u_2(x, y) + v_1(x, y)v_2(x, y)]\, dx\, dy, \qquad \text{(C.3)}$$

where integration is performed over a region W symmetric with respect to both the x- and the y-axis on the image plane (this definition clearly satisfies the *axiom of inner product*), we can easily check that the translational flow, the divergent flow, the rotational flow and the shearing flow are mutually orthogonal. The fanning flow is also orthogonal to the divergent flow, the rotational flow and the shearing flow. (The translational flow and the fanning flow are not orthogonal.)

From this nice property of the (2D) linear decomposition of an optical flow, it is desirable to choose an interpretation of the motion in such a way that the component flows $U_a(x, y), \dots, U_{\omega_3}(x, y)$ of Eq. (C.2) coincide with this (2D) decomposition of an optical flow on the image plane. However, this (2D) decomposition does not correspond to any interpretation. (The coefficients of this (2D) decomposition cannot be regarded exactly as the (3D) velocity components of the object.) However, if we adopt the object-based interpretation, the (2D) translational flow is dominantly affected by the (3D) translation along the x- and y-axes, the (2D) divergent flow by the (3D) translation along the z-axis, the (2D) rotational flow by the (3D) rotation around the z-axis, and the (2D) fanning flow by the (3D) rotation around the x- and y-axes. Hence, we can conclude that the object-based interpretation is quite appropriate, although it may not be optimal.

If we adopt the camera-based interpretation, the components $U_a(x, y)$ and U_{ω_2} are very "similar" (except for the sign). So are $U_b(x, y)$ and $U_{\omega_1}(x, y)$. These similar flows may cancel each other, resulting in a small observed flow on the image plane. Hence, this choice is undesirable.

A question naturally arises. What if we are only interested in determining the *egomotion*, i.e., the motion of the camera or the viewer, not the object motion? Is the object-based interpretation any better than the camera-based interpretation in computing egomotion? The answer is as follows. If we use the object-based interpretation, we can compute the object motion in reference to the camera with some error, but as stated above this is the best we can do. Then, we move to the camera-based interpretation by Eq. (C.1). This transition process amplifies the error when the object is far away from the camera. If the camera rotation is small, cancellation due to subtraction may take place on the right-hand side of Eq. (C.1), resulting in a large relative error. Overall, determination of egomotion from optical flow is a very difficult problem.

In short, determination of the object motion from optical flow is not in itself ill-posed or "inherently ambiguous" as claimed by Adiv [2]. Various other factors reduce the problem to being ill-posed. The transition from an object motion into an equivalent camera motion is one of them. Another case is when the object is very far away or f is very large. In this case, the component flow $U_c(x, y)$ becomes very small, making the decomposition of Eq. (C.2) very difficult. Still another possibility is when the surface is small and the gradient is close to zero. In this case, the component flows $U_{\omega_1}(x, y)$, $U_{\omega_2}(x, y)$ become very small, again making the decomposition of Eq. (C.2) difficult.

Error sensitivity is greatly reduced if some a priori knowledge about the motion is available, e.g., the object may be known to be translating but not rotating, or may be constrained on a horizontal plane, or the *ground*. Then, the number of degrees of freedom of the motion becomes small. Thus, in order to cope with the computational robustness problem, we must accurately identify the *true* sources of difficulty, analyze them and devise supplementary means to overcome them. After all, this is what a theoretical study like this one is all about.

ACKNOWLEDGMENTS

The author thanks Professor Azriel Rosenfeld and Professor Larry S. Davis of the University of Maryland and Professor Shun-ichi Amari of the University of Tokyo for helpful discussions and comments. This work was supported in part by the Defense Advanced Research Projects Agency and the U.S. Army Night Vision and Electro-Optics Laboratory under Contract DAAK70-K-0018 (DARPA Order 3206).

REFERENCES

1. G. Adiv, Determining 3-D motion and structure from optical flow generated by several moving objects, *IEEE Trans. Pattern. Anal. Mach. Intell.* **PAMI 7**, No. 4, 1985, 384–401.
2. G. Adiv, Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, Proc. Darpa Image Understanding Workshop, Miami Beach, Fla., 1985, pp. 399–412.
3. A. R. Bruss and B. K. P. Horn, Passive navigation, *Comput. Vision Graphics Image Process.* **21**, 1983, 3–20.
4. B. F. Buxton, D. W. Murray, H. Buxton, and N. S. Williams, Structure-from-motion algorithms for computer vision on an SIMD architecture, *Comput. Phys. Comm.* **37**, 1985, 273–280.
5. J.-Q. Fang and T. S. Huang, Solving three-dimensional small-rotation motion equations: Uniqueness, algorithms, and numerical results, *Comput. Vision Graphics Image Process.* **26**, 1984, 183–206.
6. M. Hammermesh, *Group Theory and Its Application to Physical Problems*, Addison–Wesley, Reading, Mass., 1964.
7. J. C. Hay, Optical motions and space perception: An extension of Gibson's analysis, *Psychol. Rev.* **73**, No. 6, 1966, 550–563.
8. B. K. P. Horn and B. G. Schunk, Determining optical flow, *Artif. Intell.* **17**, 1981, 185–203.
9. K. Kanatani, Detection of surface orientation and motion from texture by a stereological technique, *Artif. Intell.* **23**, 1984, 213–237.
10. K. Kanatani, Tracing planar surface motion from projection without knowing the correspondence, *Comput. Vision Image Process.* **29**, 1984, 1–12.
11. K. Kanatani, Detecting the motion of a planar surface by line and surface integrals, *Comput. Vision Graphics Image Process.* **29**, 1984, 13–22.

12. K. Kanatani, Structure and motion without correspondence: General principle, Proc. 9th Int. Joint. Conf. Artif. Intell., Los Angeles, Calif., 1985, pp. 886–888.
13. K. Kanatani, Structure and motion without correspondence: General principle, Proc. Darpa Image Understanding Workshop, Miami Beach, Fla., 1985, pp. 107–116.
14. K. Kanatani, Transformation of optical flow by camera rotation, Proc. IEEE Workshop on Motion: Representation and Analysis, Charleston, S.C., 1986, pp. 113–118.
15. K. Kanatani, Structure and motion from optical flow under orthographic projection, *Comput. Vision Graphics Image Process.* **35**, 1986, 181–199.
16. J. J. Koenderink and A. J. van Doorn, Exterospecific component of the motion parallax field, *J. Opt. Soc. Amer.* **71**, 1981, 953–957.
17. H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* **239**, 1981, 133–135.
18. H. C. Longuet-Higgins, The visual ambiguity of a moving plane, *Proc. Roy. Soc. London. Ser. B* **223**, 1984, 165–175.
19. H. C. Longuet-Higgins and K. Prazdny, The interpretation of a moving retinal image, *Proc. Roy. Soc. London. Ser. B* **208**, 1980, 385–397.
20. S. J. Maybank, The angular velocity associated with the optical flowfield arising from motion through a rigid environment, *Proc. Roy Soc. London Ser. A* **401**, 1985, 317–326.
21. H.-H Nagel, Representation of moving rigid objects based on visual observations, *Computer* **14**, No. 8, 1981, 29–39.
22. J. M. Prager and M. A. Arbib, Computing the optic flow: The MATCH algorithm and prediction, *Comput. Vision Graphics Image Process.* **24**, 1983, 271–304.
23. K. Prazdny, Egomotion and relative depth map from optical flow, *Biol. Cybernet.* **36**, 1980, 87–102.
24. K. Prazdny, Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer, *Comput. Graphics Image Process.* **17**, 1981, 238–248.
25. J. W. Roach and J. K. Aggarwal, Determining the movement of objects from a sequence of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI 2**, No. 6, 1980, 87–102.
26. M. Subbarao, *Interpretation of Image Motion Fields: Rigid Curved Surfaces in Motion*, Technical Report CAR-TR-199, Center for Automation Research, University of Maryland, 1986.
27. M. Subbarao and A. M. Waxman, Closed form solution to image flow equations for planar surface in motion, *Comput. Vision Graphics Image Process.* **36**, 1986, 208–228.
28. R. Y. Tsai and T. S. Huang, Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI 6**, No. 1, 1984, 13–27.
29. S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, Mass., 1979.
30. A. M. Waxman and S. Ullman, Surface structure and three-dimensional motion from image flow kinematics, *Int. J. Robotics Res.* **4**, No. 3, 1985, 72–94.
31. A. M. Waxman and K. Wohn, Contour evolution, neighborhood deformation, and global image flow: planar surfaces in motion, *Int. J. Robotics Res.* **4**, No. 3, 1985, 95–108.
32. J. Weyl, *The Theory of Groups and Quantum Mechanics*, Dover, New York, 1950.
33. H. Weyl, *The Classical Groups, Their Invariants and Representations*, 2nd ed., Princeton Univ. Press, Princeton, N.J., 1946.