COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 35, 181–199 (1986)

Structure and Motion from Optical Flow under Orthographic Projection

KEN-ICHI KANATANI*

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

Received December 30, 1985; revised March 24, 1986

The 3D structure and motion of an object are determined from its optical flow under orthographic projection. The image domain is divided into planar or almost planar regions by checking the flow. For each region, parameters specifying the flow are estimated, and the 3D structure and motion are determined in explicit analytical forms in terms of these parameters. The solution is not unique, containing an indeterminate scale factor and consisting of true and spurious solutions. The spurious solution disappears if two or more regions of the same object are observed. Geometrical invariance properties are emphasized throughout this paper. © 1986 Academic Press, Inc.

1. INTRODUCTION

Determination of the 3D structure and motion of an object from its projected 2D images is one of the most important tasks of computer vision. Most existing studies adopt *perspective projection*, or a pin hole camera model [1–3]. However, the effect of perspective projection decreases when the object size is small compared with the distance from the camera. As a result, many clues are lost. For example, the motion along the camera axis is indiscernible. In this case, the projection must be regarded as orthographic.

Although 3D recovery from images orthographically viewed from finitely many different orientations was studied in [4, 5], only a few attempts have been made at 3D recovery from optical flow under orthographic projection [6, 7]. (For perspective projection, see [8].) The reason seems to be that we cannot obtain sufficient clues for 3D recovery. However, since the effect of perspective projection sometimes becomes negligible, it is an important problem to examine to what extent the 3D information can be recovered. In other words, since 3D recovery from orthographic projection necessarily involves indeterminacy, we need to know what kind of indeterminacy it is or what the geometrical meaning of the indeterminacy is.

This attempt was partly made by Hoffman [6] from a human perception viewpoint. From a computer vision viewpoint, Sugihara and Sugie [7] presented a general reconstruction algorithm, pointing out that a scale factor is necessarily involved. However, their proof is incomplete. In fact, they did not prove that no other indeterminacy is involved. A purpose of this paper is to present a different approach to the same problem. The approach here is similar to that of Hoffman [6] but is fundamentally different from that of Sugihara and Sugie [7] in the following three aspects.

First, the computation of Sugihara and Sugie [7] is based on the *coordinates* of point-to-point correspondences, while the present method first extracts *global characteristics* of the flow field and subsequent computation is based on these quantities,

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

0734-189X/86 $3.00

which we call the "flow parameters" of the optical flow. In the "correspondence-based" approach of Sugihara and Sugie [7], the velocity measurement must be accurate, while in our "flow-based" approach the flow parameters are less sensitive to noise, since they are global quantities obtained through averaging a set of observed data. This is of great practical significance. At the same time, our method is more general than theirs in the sense that our flow-based approach also includes the correspondence-based approach of Sugihara and Sugie [7] as a special case, as is shown later.

Second, the algorithm of Sugihara and Sugie [7] only gives numerical results, while our method gives the 3D solution expressed explicitly *in analytical terms*. Hence, it provides a complete description of the involved indeterminacy, which the approach of Sugihara and Sugie [7] is unable to show. The solution is not unique if the object is a plane. However, since the solution has an analytically closed form, the geometrical meaning of the spurious solution can be given in general terms.

Third, since the algorithm of Sugihara and Sugie [7] is based on a necessary condition of rigid motion, their numerical solution also includes physically impossible solutions, while our method yields and exhausts all physically possible solutions.

2. IDENTIFICATION OF OPTICAL FLOW

Suppose a plane is moving in the scene and we are looking at its image orthographically projected onto the xy plane, which is identified with the image plane, along the z axis (Fig. 1). Let $z = px + qy + r$ be the equation of the plane. Let $(0, 0, r)$, the intersection between the surface and the z axis, be a reference point. The instantaneous rigid motion is specified by the translation velocity (a, b, c) at the reference point and the rotation velocity $(\omega_1, \omega_2, \omega_3)$ around it (i.e., with rotation axis $(\omega_1, \omega_2, \omega_3)$ and angular velocity $\sqrt{\omega_1^2 + \omega_2^2 + \omega_3^2}$ (rad/s) around it).

Thus, the instantaneous position and motion of the plane is described by nine parameters $p, q, r, a, b, c, \omega_1, \omega_2, \omega_3$, which we call the *structure and motion parameters*. Our objective here is to recover them by observing the image motion on the image plane. Obviously, two of them cannot be recovered due to the orthography of

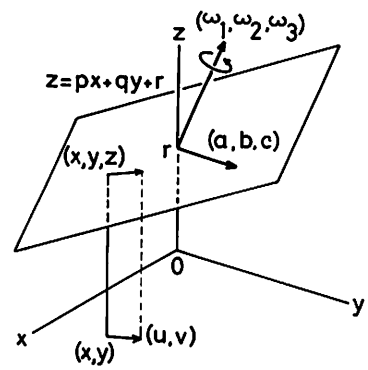

FIG. 1. A plane having equation $z = px + qy + r$ is moving with translation velocity (a, b, c) at $(0, 0, r)$ and rotation velocity $(\omega_1, \omega_2, \omega_3)$ around it. An optical flow is induced on the xy plane by orthographic projection along the z axis.

projection, namely c, the velocity in the z direction, and r, the distance from the origin along the z axis, which we call the *absolute depth* of the plane.

It is easy to see that the motion of the plane described above induces an *optical flow* on the image plane of the form

$$u(x, y) = a + p\omega_2 x + (q\omega_2 - \omega_3)y, \qquad v(x, y) = b - (p\omega_1 - \omega_3)x - q\omega_1 y. \tag{2.1}$$

In other words, the optical flow is described by linear equations in the form

$$u(x, y) = a + Ax + By, \qquad v(x, y) = b + Cx + Dy, \tag{2.2}$$

where

$$A = p\omega_2, \qquad B = q\omega_2 - \omega_3, \qquad C = -p\omega_1 + \omega_3, \qquad D = -q\omega_1. \tag{2.3}$$

Suppose the optical flow is already obtained on the image plane by some available means [9–12]. Then, coefficients a, b, A, B, C, D, which we call the *flow parameters*, can be estimated by fitting equations of the form (2.2), say by the least square method, minimizing

$$M = \sum_i \left[(a + Ax_i + By_i - u(x_i, y_i))^2 + (b + Cx_i + Dy_i - v(x_i, y_i))^2 \right], \tag{2.4}$$

where summation is taken over all feature points where the velocity is observed.

By computing the residual of Eq. (2.4), 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 procedure: Starting from three or more 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 decomposed into planar patches. (Exact boundaries of these planar patches are not necessary. They are reconstructed by the procedure described later.)

3. INVARIANCE PROPERTIES OF OPTICAL FLOW

Once the flow parameters a, b, A, B, C, D 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). They are the only restrictions constraining the solution, for motions seem identical to the viewer if the flow parameters are the same. Here, the translation velocities a, b need not be considered because they are already observed, while c and r are indeterminate because they are not contained in Eqs. (2.3).

Equations (2.3) provide four equations for five unknowns $p, q, \omega_1, \omega_2, \omega_3$, and hence one degree of freedom remains indeterminate. It seems, at first sight, that we can solve Eqs. (2.3) by choosing one unknown, say p, and expressing the rest of unknowns in terms of it. Hoffman [6] actually did this. In this way, however, we cannot understand the geometrical implication of the indeterminacy. This is a defect

inherent to "algorithmic solutions" in general, because the "invariance properties" are destroyed. It seems, however, that very few people in computer vision realize this fact. To illustrate this aspect of invariance is one of the main purposes of this paper.

The key fact is that an optical flow is described in the form of Eqs. (2.2) with respect to a 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 origin by angle θ counterclockwise. Since we are observing the rigid motion of a plane, the optical flow must have the same form

$$u' = a' + A'x' + B'y', \qquad v' = b' + C'x' + D'y', \tag{3.1}$$

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 velocity components are transformed as a 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 a, b are transformed as a vector and A, B, C, D are transformed as a tensor, i.e.,

$$\begin{bmatrix} a' \\ b' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}, \tag{3.4}$$

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

Equations (3.4) and (3.5) describe a linear mapping from a, b, A, B, C, D onto a', b', A', B', C', D', and this mapping is a *representation*, or a *homomorphism*, of the 2D rotation group $SO(2)$. As is well known in group representation theory, any representation of $SO(2)$ is reduced to 1-dimensional *irreducible representations* due to *Schur's lemma*, since $SO(2)$ is a compact Abelian group [13, 14]. In fact, if we define quantities

$$U_0 = a + ib, \qquad T = A + D, \qquad R = C - B, \qquad S = (A - D) + i(B + C), \tag{3.6}$$

where i is the imaginary unit, the transformation rule becomes

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

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

Since these quantities are irreducible representations, each of them has a distinctive geometrical meaning [15]. In fact, U_0 represents *translation*, T *divergence*, R *rotation*, and S *shearing* of the optical flow. In particular, the magnitude $|S|$ is the *shear strength*, and the *principal axes* defined by

$$Q_1 = e^{i \arg(S)/2}, \qquad Q_2 = ie^{i \arg(S)/2}, \tag{3.8}$$

indicate the orientations of the *maximum extension* and *maximum compression*, respectively (Appendix B). The quantities are also derived by decomposing the flow according to tensor symmetry properties as shown in Appendix B. This is not a coincidence; according to the general theorem of Weyl, *all* irreducible representations of any *tensor representation* of $SO(n)$ are obtained by this process [15].

Similarly, the gradient components p, q and the rotation velocities ω_1, ω_2 are transformed as vectors with respect to the coordinate rotation, while ω_3 is a scalar. Namely,

$$\begin{bmatrix} p' \\ q' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}, \qquad \begin{bmatrix} \omega_1' \\ \omega_2' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}, \tag{3.9}$$

and $\omega_3' = \omega_3$. Hence, if we combine them into complex parameters

$$P = p + iq, \qquad W = \omega_1 + i\omega_2, \tag{3.10}$$

they are (relative) invariants of weight -1:

$$P' = e^{-i\theta}P, \qquad W' = e^{-i\theta}W. \tag{3.11}$$

4. ANALYTICAL SOLUTION OF STRUCTURE AND MOTION

From Eqs. (2.3), the invariants introduced in the previous section become as follows:

$$T = p\omega_2 - q\omega_1, \qquad R = 2\omega_3 - p\omega_1 - q\omega_2, \qquad S = p\omega_2 + q\omega_1 + i(q\omega_2 - p\omega_1). \tag{4.1}$$

The first two real equations are combined into one complex equation

$$R + iT = 2\omega_3 - p\omega_1 - q\omega_2 + i(p\omega_2 - q\omega_1). \tag{4.2}$$

In terms of the complex notation (3.10), the last of Eqs. (4.1) and Eq. (4.2) become

$$PW^* = 2\omega_3 - (R + iT), \qquad PW = iS, \tag{4.3}$$

respectively. Thus, four real equations (2.3) for five unknowns $p, q, \omega_1, \omega_2, \omega_3$ are converted into two complex equations (4.3) for three unknowns P, W, ω_3. Note that P and W are invariants of weight -1, and hence W^*, the complex conjugate of W, is of weight 1. Hence, both sides of the first of Eqs. (4.3) are of weight 0 (scalars), while both sides of the second are of weight -2.

Since $|PW^*| = |PW|$, the right-hand sides of Eqs. (4.3) have the same modulus, i.e.,

$$(2\omega_3 - (R + iT))(2\omega_3 - (R - iT)) = SS^*, \tag{4.4}$$

from which ω_3 is given by

$$\omega_3 = \tfrac{1}{2}(R \pm \sqrt{SS^* - T^2}). \tag{4.5}$$

Here, we obtain the following *rigidity criterion*:

PROPOSITION 1. *The magnitude of divergence should not be greater than the shear strength*:

$$|T| \leq |S|. \tag{4.6}$$

Namely, if inequality (4.6) is not satisfied, the flow cannot be regarded as caused by rigid planar motion.

Another thing to note is that Eq. (4.5) gives two solutions for ω_3. If Eqs. (3.6) and (2.3) are substituted, it can be checked that one solution is indeed the true ω_3 and the other is $\omega_3 - (p\omega_1 + q\omega_2)$. However, we cannot tell which of the two solutions of Eq. (4.4) is the true solution and which is the spurious one. The spurious solution does not appear if and only if $p\omega_1 + q\omega_2 = 0$. In the following, complex numbers are also identified with 2D vectors on the complex plane. Then,

PROPOSITION 2. *The spurious solution does not appear if and only if the orientations of P and W are orthogonal.*

For each of the two solutions of ω_3, Eqs. (4.3) determine P and W. However, it is immediately observed that the magnitude of W is indeterminate, since W multiplied by a real number together with P divided by that number also satisfy Eqs. (4.3). Thus, the magnitude $k = |W|$ can be taken as an indeterminate scale factor, which is a scalar and hence is invariant with respect to the coordinate rotation. This means that k describes a certain geometrical property. In fact, it will be shown that k is a measure of elongation (or compression) of the object along the z axis. If an indeterminate parameter were arbitrarily introduced, it would not necessarily have a geometrical meaning like that. This is one of the key points of the present formulation.

Elimination of P from Eqs. (4.3) by taking the ratio yields

$$\frac{W}{W^*} = \frac{iS}{2\omega_3 - (R + iT)}. \tag{4.7}$$

Taking the argument of both sides yields

$$2\arg(W) = \frac{\pi}{2} + \arg(S) - \arg(2\omega_3 - (R + iT)) \pmod{2\pi}, \tag{4.8}$$

and hence

$$\arg(W) = \frac{\pi}{4} + \frac{1}{2}\arg(S) - \frac{1}{2}\arg(2\omega_3 - (R + iT)) \pmod{\pi}. \tag{4.9}$$

However, the mod π can be ignored by allowing the scale factor k to be negative. Then, W is obtained in the form of $ke^{i\arg(W)}$, and, from the second of Eqs. (4.3), P is given by $P = iS/W$. In summary,

THEOREM 1.

$$\omega_3 = \tfrac{1}{2}(R \pm \sqrt{SS^* - T^2}), \tag{4.5}$$

$$W = k \exp i\left(\frac{\pi}{4} + \frac{1}{2}\arg(S) - \frac{1}{2}\arg(2\omega_3 - (R + iT))\right), \tag{4.10}$$

$$P = \frac{S}{k}\exp i\left(\frac{\pi}{4} - \frac{1}{2}\arg(S) + \frac{1}{2}\arg(2\omega_3 - (R + iT))\right). \tag{4.11}$$

From Eqs. (4.10) and (4.11), we find that

$$\frac{1}{2}(\arg(P) + \arg(W)) = \frac{1}{2}\arg(S) \pm \frac{\pi}{4}, \tag{4.12}$$

where $\pm$ corresponds to the sign of the scale factor k. As shown in Appendix B, $\arg(S)/2$ is the orientation of maximum extension. Hence, the orientation of the right-hand side is the orientation bisecting the orientations of maximum extension and maximum compression, known in fluid mechanics as the orientation of *maximum shearing*, since viscosity becomes maximum along this orientation. Thus, we conclude

COROLLARY 1. *The orientations of P and W are always symmetric with respect to the orientation of maximum shearing.*

So far, we have assumed that ω_3 is the true solution. If that is replaced by $\omega_3 - (p\omega_1 + q\omega_2)$ in the previous derivation, we also find the following (Appendix C):

COROLLARY 2. *The orientations of true P and spurious W are always mutually orthogonal, and so are the orientations of true W and spurious P. Hence, the orientations of true and spurious W are always symmetric with respect to the principal axes, and so are the orientations of true and spurious P.*

These general properties result from our analytical solution; they cannot be obtained by numerical algorithms. The next example illustrates these observations.

EXAMPLE 1. Consider the flow of Fig. 2. The flow parameters are $a = 0.1$, $b = 0.1$, $A = 0.087$, $B = -0.227$, $C = 0.087$, $D = 0.0524$, and hence the invariants are $T = 0.140$, $R = 0.314$, $S = 0.035 - 0.140i$. Thus, the rigidity criterion $|T| < |S|$ $(= 0.144)$ is satisfied. Equation (4.5) yields $\omega_3 = 10, 8$ (deg/s), and Eqs. (4.11) and (4.12) yield two solutions $W_1 = (0.706 + 0.708i)k$ (rad/s), $P_1 = (0.123 - 0.074i)/k$ and $W_2 = (0.516 + 0.857i)k$ (rad/s), $P_2 = (0.102 - 0.102i)/k$, where k is an indeterminate scale factor. Figure 3 shows the case of $k = 0.5$ on the complex plane, where the orientations of maximum extension, maximum compression and maximum shearing are also indicated. The statements in Corollaries 1 and 2 are easily confirmed.

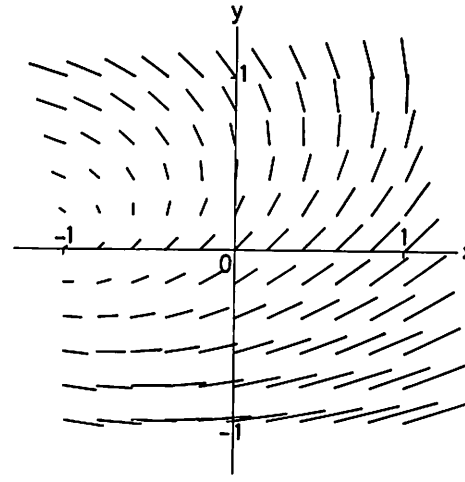

FIG. 2. An example of optical flow.

5. ADJACENCY OF OPTICAL FLOWS

So far, only one planar patch has been considered on the image plane. Here, let us consider the relationship among them when the planar patches are images of one and the same rigid object. Let a, b, A, B, C, D are the observed flow parameters for one patch and a', b', A', B', C', D' those for another. If these two patches correspond to two planar surface of the same object, the induced optical flows must be continuous over the *intersection line* (or, to be precise, the image of the intersection

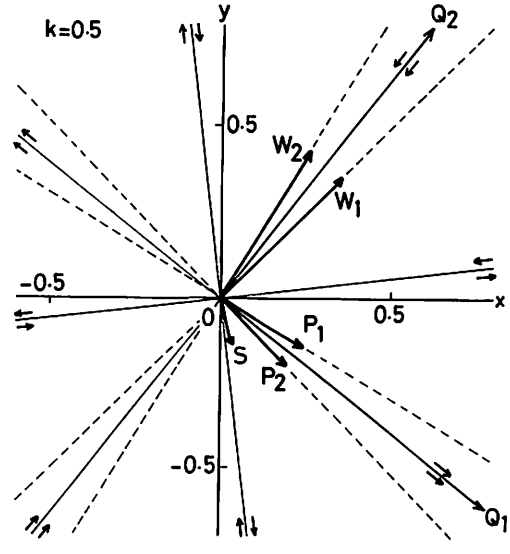

FIG. 3. The result of analysis of the flow of Fig. 2. Two solutions exist, one being true and the other spurious. The principal axes and the orientations of maximum shearing are also indicated.

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 relations

$$([u] =)[a] + [A]x + [B]y = 0, \qquad ([v] =)[b] + [C]x + [D]y = 0, \quad (5.1)$$

where [ ] designates the difference, e.g., $[a] = a' - a$. Since Eqs. (5.1) hold at any point on the intersection line, they must be the actual equation of the intersection line. The necessary and sufficient condition that Eqs. (5.1) represent one and the same line is that the ratios between the corresponding coefficients are equal. Hence, we obtain the following *adjacency criterion*:

PROPOSITION 3. *If two planar patches belong to the same object,*

$$[a] : [b] = [A] : [C] = [B] : [D] \quad (5.2)$$

must be satisfied. The equation of the intersection line is given by either of Eqs. (5.1).

In other words, if Eq. (5.2) is not satisfied, the two patches are images of two different independently moving objects, while if it is 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; there is no need for edge detection.

Once the image is partitioned into adjacent planar patches by the above procedure, the previous analysis is done for each planar patch, recovering P, W, ω_3 up to a scale factor k. If $z = px + qy + r$ and $z' = p'x + qy' + r'$ are two planes in the scene, the image of the intersection line has the form

$$[p]x + [q]y + [r] = 0. \quad (5.3)$$

If we regard the complex number P as a 2D vector on the image plane, which is identified with the complex plane, we obtain the well known fact (often described in terms of the *gradient space*):

PROPOSITION 4. $[P]$ *is perpendicular to the intersection line.*

As was shown in the previous section, two sets of solutions for P, W, ω_3 exist for each planar patch. However, if the motion is rigid, W and ω_3 must be common to all, and hence we can pick up the true solution for each patch. (If both the true and spurious W and ω_3 happen to be common, we have $p\omega_1 + q\omega_2 = p'\omega_1 + q'\omega_2$, or $[p]\omega_1 + [q]\omega_2 = 0$. This means that W is perpendicular to $[P]$, and hence parallel to the intersection line by Proposition 4. Thus, the true W can be detected.) Moreover, since the indeterminate scale factor k is defined as the magnitude of W, it can be chosen so that it is common over the entire object image.

Let $y = mx + n$ be the equation of the intersection line between two patches, which is immediately obtained as described before. Comparison of this with Eq. (5.3) yields $[p] : [q] : [r] = m : -1 : n$. Hence,

PROPOSITION 5. *If the intersection line is* $y = mx + n$ *and the equations of the planes for the adjacent planar patches are* $px + qy + r = 0$ *and* $p'x + q'y + r' = 0$,

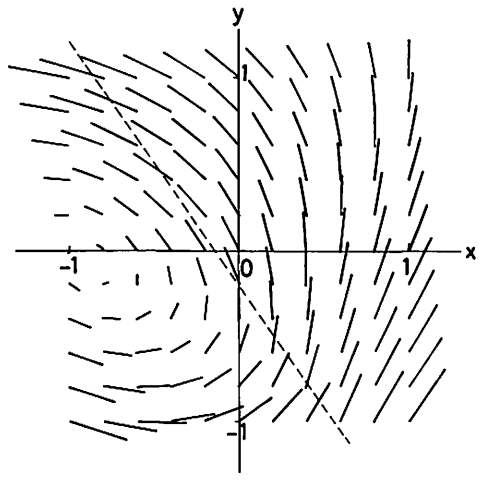

FIG. 4. This flow cannot be regarded as a single flow of a planar patch. There exist two planar patches, and the dashed line is the intersection line computed from the flow.

the relative depth *is given by*

$$[r] = \frac{n}{m}[p] = -n[q]. \tag{5.4}$$

This means that if the absolute depth is assumed for one patch, the depths for all other patches are uniquely determined. In conclusion, we have

THEOREM 2. *The structure and motion of an object are determined from its optical flow under orthographic projection only up to a single indeterminate absolute depth r and a single indeterminate scale factor k.*

The recovery is done in analytical terms in closed form. This theorem covers both objects with smooth surfaces, which are approximated by planar patches, and polyhedra, for which each face corresponds to a planar patch.

EXAMPLE 2. Consider the flow of Fig. 4. The flow as a 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 $a = -0.1$, $b = 0.2$, $A = 0.209$, $B = -0.105$, $C = 0.070$, $D = -0.035$, and for the lower left part $a' = -0.149$, $b' = 0.224$, $A' = -0.140$, $B' = -0.349$, $C' = 0.244$, $D' = 0.087$. Equation (5.2) is satisfied within rounding error, and the intersection line is estimated to be $y = -1.429x - 0.2$, which is indicated by a broken line in the figure. For the upper right part, we obtain $\omega_3 = 10, 0$ (deg/s), and for the lower left part $\omega_3 = 24, 10$ (deg/s). Hence, the true solution is $\omega_3 = 10$ (deg/s) and $W = (0.447 + 0.894i)k$. The gradients are $P = (0.234 + 0.078i)/k$ and $P' = (-0.1561 - 0.195i)/k$, respectively. The case of $k = 0.5$ is indicated in Fig. 5. As expected, $[P]$ is perpendicular to the intersection line. The equations of the two planes are $z = (0.234x + 0.078y)/k + r$, $z = (-0.156x - 0.195y)/k + (r - 0.055/k)$, respectively.

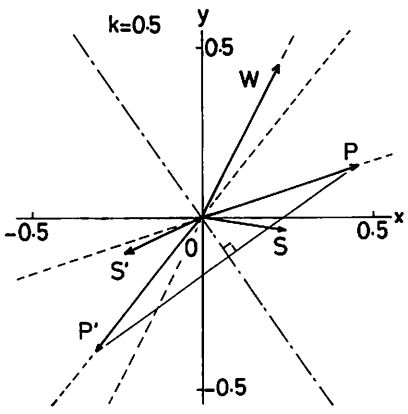

FIG. 5. The result of the analysis of the flow of Fig. 4. No spurious solution exists.

6. REDUCTION TO THE CORRESPONDENCE-BASED APPROACH

Our flow-based approach reduces to the correspondence-based approach in the extreme case where each planar patch is a triangle containing only three feature points at the three vertices. Let (x_i, y_i), $i = 1, 2, 3$, be the coordinates of these three vertices, and let (u_i, v_i), $i = 1, 2, 3$, be the velocities observed there. According to Eqs. (2.2), the flow parameters are determined by solving the simultaneous equations

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} a \\ A \\ B \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \quad \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} b \\ C \\ D \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}. \tag{6.1}$$

The solution is unique unless the determinant

$$\begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix} \tag{6.2}$$

vanishes, which is a condition for collinearity of the three points. Hence, if velocities are measured at three non-collinear points, the flow parameters are uniquely determined.

EXAMPLE 3. Suppose velocities are measured at three points $(0.6, 0.2)$, $(-0.2, -0.4)$, $(-0.4, 0.8)$, resulting in $(-0.042, 0.105)$, $(-0.098, 0.177)$, $(0.077, 0.159)$, respectively (Fig. 6). Eqs. (6.1) give $a = -0.049$, $b = 0.152$, $A = -0.035$, $B = 0.140$, $C = -0.070$, $D = -0.026$. The corresponding hypothetical flow is drawn in Fig. 7. The procedure shown previously yields two sets of solutions; $\omega_3 = -5$ (deg/s), $W = (0.448 + 0.894i)k$ (rad/s), $P = (-0.039 + 0.059i)/k$, and $\omega_3 = -7$ (deg/s), $W = (0.832 + 0.555i)k$ (rad/s), $P = (-0.063 + 0.032i)/k$. Hence, there are two possibilities for the equation of the plane. One is $z =$

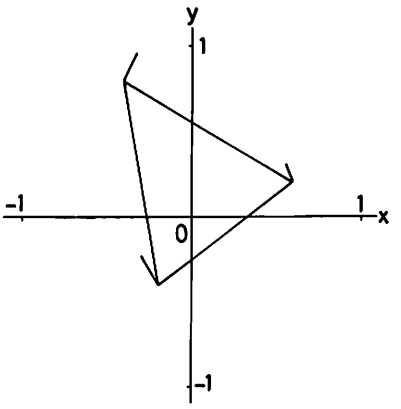

FIG. 6. The velocity is observed at three points of a rigid object.

$-0.039x/k + 0.059y/k + r$, and the z coordinates at the above three vertices are $z_1 = -0.012/k + r$, $z_2 = -0.016/k + r$, $z_3 = 0.062/k + r$, respectively. The other is $z = -0.063x/k + 0.032y/k + r$, and the z coordinates at the vertices are $z_1 = -0.031/k + r$, $z_2 = r$, $z_3 = 0.050/k + r$, respectively. The spurious solution disappears if two or more planar patches are observed.

By now, the role of the scale factor k is clear. The relative depth, i.e., the difference between the z coordinates, of points is inversely proportional to k, so that k is a measure of *elongation* (or *contraction*) of the object along the z axis. This interpretation was already pointed out by Sugihara and Sugie [7].

On the other hand, the correspondence-based approach of Sugihara and Sugie [7] is based on the rigidity condition $(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 = \text{const.}$, or on differentiation

$$(x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j) = 0, \quad (6.3)$$

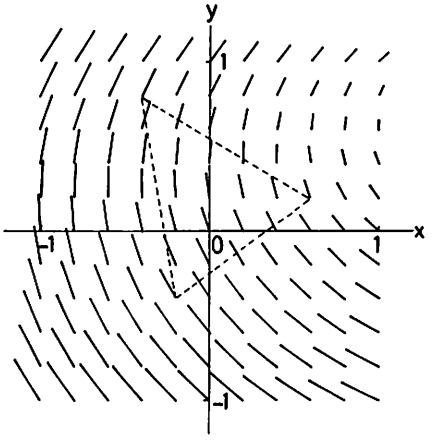

FIG. 7. The hypothetical optical flow computed from the velocities at the three points in Fig. 6.

for all i, j pairs. They regarded these constraints as equations for unknowns z_i and $\dot{z}_i$, analyzed the existence and indeterminacy and presented an algorithm for numerical computation. However, Eq. (6.3) provides only a necessary condition for rigid motion. This is also satisfied by non-rigid (i.e., self-deforming) motion when four or more points become coplanar. Indeed, such a non-rigid solution is actually found in the numerical solution of their algorithm, although they did not mention this in their paper. Our analytical formulation does not yield such an infeasible solution because it is based on the rigidity of planar surfaces.

7. CONCLUDING REMARKS

We have presented a complete analysis of optical flow under orthographic projection and expressed the solution in analytical closed form including the spurious solution. The underlying principle is the *invariance* with respect to coordinate changes on the image plane, and equations are written in terms of invariants, which are in general complex numbers, resulting from group representation theory. The analysis is based on the optical flow of plane motion but it can be reduced to the correspondence-based approach as a special case. There are two more things to be mentioned.

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 [16, 17], who computed what he called "features" to estimate the flow parameters.

On the other hand, the goal of the analysis shown so far is to detect 3D structure and motion from a given optical flow at one instant. There are several indeterminacies; the velocity along the z axis and the distance from the viewer are indeterminate, and one additional indeterminate scale factor is involved. Moreover, if the object is a plane, there exist two sets of solutions. However, these indeterminacies can be removed if multiple optical flows are available. For example, if two optical flows of the same object are detected at different times or from different viewpoints, the solution becomes unique, as was pointed out by Sugihara and Sugie [7].

The same principle applies if a time sequence of optical flows taken at a small time interval is available because the structure and motion parameters cannot evolve arbitrarily. If the values of $p, q, r, a, b, c, \omega_1, \omega_2, \omega_3$ are given at one instant, the translation velocities a, b, c and the rotation velocities $\omega_1, \omega_2, \omega_3$ change the position of the plane is such a way that

$$\dot{p} = pq\omega_1 - (p^2 + 1)\omega_2 - q\omega_3, \qquad \dot{q} = (q^2 + 1)\omega_1 - pq\omega_2 + p\omega_3,$$
$$\dot{r} = c - pa - qb. \tag{7.1}$$

Hence, the values of p, q, r after short time δt are predicted to be $p + \dot{p}\delta t$, $q + \dot{q}\delta t$, $r + \dot{r}\delta t$, respectively (or a higher order approximation scheme can be applied). For example, if the values of $p, q, r, a, b, c, \omega_1, \omega_2, \omega_3$ are known initially, Eqs. (7.1) provides a constraint sufficient to determine the subsequent values of these parameters uniquely. This principle, combined with the method of feature detection, is used by Kanatani [18–20] to trace the motion of a plane from a known initial condition.

APPENDIX A

Equations (3.4) and (3.5) describe a linear mapping from a, b, A, B, C, D onto a', b', A', B', C', D' and is rearranged into the following form:

$$\begin{bmatrix} a' \\ b' \\ A' \\ B' \\ C' \\ D' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & & & & \\ -\sin\theta & \cos\theta & & & & \\ & & \cos^2\theta & \cos\theta\sin\theta & \cos\theta\sin\theta & \sin^2\theta \\ & & -\cos\theta\sin\theta & \cos^2\theta & -\sin^2\theta & \cos\theta\sin\theta \\ & & -\cos\theta\sin\theta & -\sin^2\theta & \cos^2\theta & \cos\theta\sin\theta \\ & & \sin^2\theta & -\cos\theta\sin\theta & -\cos\theta\sin\theta & \cos^2\theta \end{bmatrix} \begin{bmatrix} a \\ b \\ A \\ B \\ C \\ D \end{bmatrix}. \tag{A.1}$$

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

$$\begin{bmatrix} a' + ib' \\ a' - ib' \\ A' + D' \\ B' - C' \\ (A' - D') + i(B' + C') \\ (A' - D') - i(B' + C') \end{bmatrix} = \begin{bmatrix} e^{-i\theta} & & & & & \\ & e^{i\theta} & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & e^{-2i\theta} & \\ & & & & & e^{2i\theta} \end{bmatrix} \times \begin{bmatrix} a + ib \\ a - ib \\ A + D \\ B - C \\ (A - D) + i(B + C) \\ (A - D) - i(B + C) \end{bmatrix}. \tag{A.2}$$

Thus, the representation of Eq. (A.1) is reduced to the direct sum of 1-dimensional irreducible representations. This is possible due to Schur's lemma, because $SO(2)$ is a compact Abelian group [13, 14].

APPENDIX B

The optical flow of Eqs. (2.2) is written in matrix notation as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{B.1}$$

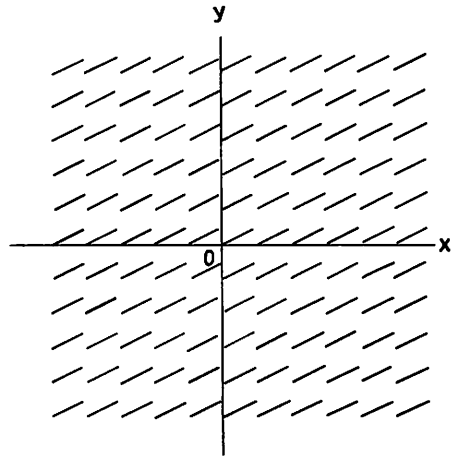

FIG. B1. Translational flow.

The geometrical meaning of a, b is clear. If the other flow parameters are zero, the flow takes the form of

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}, \tag{B.2}$$

which describes *translational flow* (Fig. B1).

The matrix in Eq. (B.1) is decomposed into its *symmetric part* and *antisymmetric* (or *skewsymmetric*) part:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & (B+C)/2 \\ (B+C)/2 & D \end{bmatrix} + \begin{bmatrix} 0 & -(C-B)/2 \\ (C-B)/2 & 0 \end{bmatrix}. \tag{B.3}$$

This decomposition is *invariant*, i.e., the two matrices on the right-hand side are transformed independently as tensors. The geometrical meaning of parameter $R = C - B$ is seen if the rest of the flow parameters are set to zero. The flow takes the form of

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{R}{2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \tag{B.4}$$

which describes *rotational flow* (Fig. B2), R being the *rotation* or *vorticity*.

The symmetric part is further decomposed into its *scalar part* and *deviator* (or *traceless*) *part*:

$$\begin{bmatrix} A & (B+C)/2 \\ (B+C)/2 & D \end{bmatrix} = \frac{A+D}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} (A-D)/2 & (B+C)/2 \\ (B+C)/2 & -(A-D)/2 \end{bmatrix}. \tag{B.5}$$

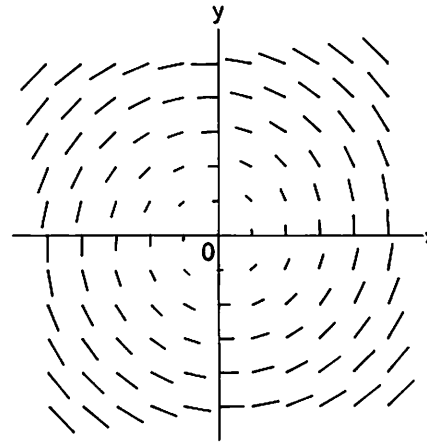

FIG. B2. Rotational flow.

Again, this decomposition is invariant, and the two parts are transformed independently. The geometrical meaning of parameter $T = A + D$ is seen if the rest of the flow parameters are set to zero. The flow takes the form of

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{T}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{B.6}$$

which describes *divergent flow* (Fig. B3), T being the *divergence*.

Since $S = (A - D) + i(B + C)$ is transformed as an invariant of weight -2, Q_1 and Q_2 defined by Eqs. (3.8) are transformed as invariants of weight -1. Hence, they correspond to vectors, and their orientations have an invariant meaning, if these two complex numbers are identified as vectors on the image plane, which is

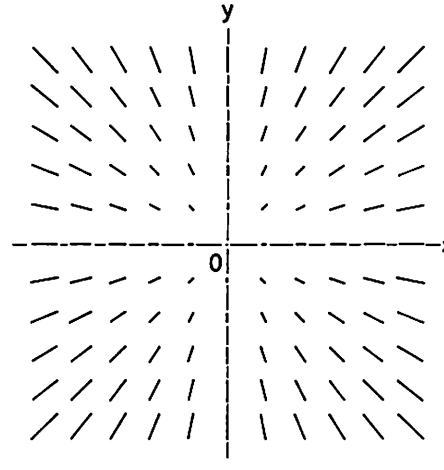

FIG. B3. Divergent flow.

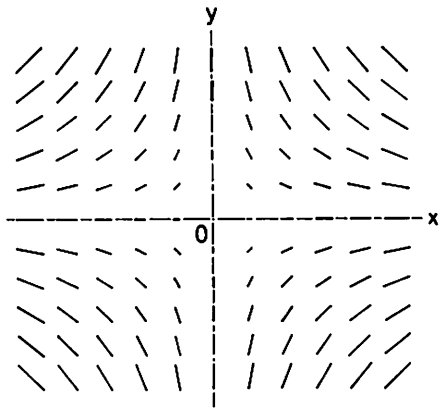

FIG. B4. Shear flow.

regarded as the complex plane. In fact, they correspond to the principal axes of the matrix of the deviator part, indicating the orientations of the *maximum extension* and *maximum compression*, respectively. To see this, take a new xy-coordinate system in such a way that Q_1 and Q_2 coincide with the x and y axes, respectively. If the other flow parameters are set to zero, the matrix is diagonalized and the flow takes the form

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{|S|}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{B.7}$$

which describes *shear flow* (Fig. B4), $|S|$ being the *shear strength*.

The invariant nature of the above decomposition of a flow is well known in fluid mechanics and is used for fluid motion analysis. This fact is also used for analysis of human visual perception resulting from optical flow induced on retina (cf. [21]).

APPENDIX C

If we replace ω_3 by $\omega_3 - (p\omega_1 + q\omega_2)$ in Eqs. (4.3) and applying $p\omega_1 + q\omega_2 = 2\omega_3 - R$ (from the second of Eqs. (4.1)), we obtain a set of equations to determine spurious P and W. In order to distinguish them from the true ones, let us write $\tilde{P}$, $\tilde{W}$. They are the solution of

$$\tilde{P}\tilde{W}^* = -2\omega_3 + (R - iT), \qquad \tilde{P}\tilde{W} = iS. \tag{C.1}$$

Taking the ratio, we see

$$\frac{\tilde{W}}{\tilde{W}^*} = -\frac{iS}{2\omega_3 - (R - iT)}. \tag{C.2}$$

If we take the complex conjugate of the first of Eqs. (4.3) and take the ratio, we obtain

$$\frac{P}{P^*} = \frac{iS}{2\omega_3 - (R - iT)}. \tag{C.3}$$

From Eqs. (C.2) and (C.3), we see that $2\arg(\tilde{W}) = 2\arg(P) + \pi \pmod{2\pi}$, or

$$\arg(\tilde{W}) = \arg(P) + \frac{\pi}{2} \pmod{\pi}. \tag{C.4}$$

Similarly, if we take the complex conjugate of the first of Eqs. (C.1) and take the ratio, we obtain

$$\frac{\tilde{P}}{\tilde{P}^*} = -\frac{iS}{2\omega_3 - (R + iT)}. \tag{C.5}$$

Comparing this with Eq. (4.7), we obtain

$$\arg(\tilde{P}) = \arg(W) + \frac{\pi}{2} \pmod{\pi}. \tag{C.6}$$

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. H.-H. Nagel, Representation of moving rigid objects based on visual observations, *Computer* **14**, 1981, 29–39.
2. H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature* **239**, 1981, 133–135.
3. 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**, 1984, 13–27.
4. S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, Mass., 1979.
5. D. D. Hoffman and B. E. Flinchbaugh, The interpretation of biological motion, *Biol. Cybernet.* **42**, 1982, 195–204.
6. D. D. Hoffman, Inferring local surface orientation from motion fields, *J. Opt. Soc. Amer.* **72**, 1982, 888–892.
7. K. Sugihara and N. Sugie, Recovery of rigid structure from orthographically projected optical flow, *Comput. Vision Graphics Image Process.* **27**, 1984, 309–320.
8. H. C. Longuet-Higgins, The visual ambiguity of a moving plane, *Proc. Roy. Soc. London Ser. B* **223**, 1984, 165–175.
9. R. Jain, Dynamic scene analysis using pixel-based processes, *Computer* **14**, No. 8, 1981, 12–18.
10. W. B. Thompson, Lower-level estimation and interpretation of visual motion, *Computer*, **14**, No. 8, 1981, 20–28.
11. B. K. P. Horn and B. G. Schunk, Determining optical flow, *Artif. Intell.* **17**, 1981, 185–203.
12. 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.
13. M. Hammermesh, *Group Theory and its Application to Physical Problems*, Addison-Wesley, Reading, Mass., 1964.
14. H. Weyl, *The Theory of Groups and Quantum Mechanics*, Dover, New York, 1950.
15. H. Weyl, *The Classical Groups, Their Invariants and Representations*, 2nd ed., Princeton Univ. Press, Princeton, N.J., 1946.
16. K. Kanatani, Structure and motion without correspondence: General principle, in *Proc. 9th Int. Joint. Conf. Artif. Intell.*, Los Angeles, 1985, pp. 886–888.

17. K. Kanatani, Structure and motion without correspondence: General principle, in *Proc. DARPA Image Understanding Workshop*, Miami Beach, 1985, pp. 107–116.
18. K. Kanatani, Detection of surface orientation and motion from texture by a stereological technique, *Artif. Intell.* **23**, 1984, 213–237.
19. K. Kanatani, Tracing planar surface motion from projection without knowing the correspondence, *Comput. Vision Graphics Image Process.* **29**, 1985, 1–12.
20. K. Kanatani, Detecting the motion of a planar surface by line and surface integrals, *Comput. Vision Graphics Image Process.* **29**, 1985, 13–22.
21. J. J. Koenderink and A. J. van Doorn, Exterospecific component of the motion parallax field, *J. Opt. Soc. Amer.* **71**, 1981, 953–957.