STRUCTURE FROM MOTION WITHOUT CORRESPONDENCE: GENERAL PRINCIPLE

Ken-ichi Kanatani*

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

ABSTRACT

A general principle is given for detection of the 3D structure and motion from an image sequence without using point-to-point correspondence. The procedure consists of two stages: (i) determination of the "flow parameters" from "features" without using correspondence and (ii) computation of the 3D structure and motion from these flow parameters. The first stage is done by solving equations of functionals, and the second stage is described in analytical expressions.

I INTRODUCTION

Schemes to recovery the 3D structure and motion from a 2D image sequence have been studied by many people, *e.g., [1 - 6]*, but most of them are based on the point-to-point correspondence, which requires a large amount of implementation effort. There do exist methods which do not require the correspondence [7 - 9], but they can be used only to trace the motion along time, starting from a given initial information, and the results are obtained only numerically.

In this paper, we present a general mathematical principle to detect the 3D structure and motion without using correspondence. Yet, the solution at particular time is given In analytical expressions, giving geometrical Interpretations and proving the existence of the spurious solution [6], *etc.* The procedure consists of two stages. First, we extract, without using correspondence, the "flow parameters" which completely characterize the viewed motion for each planar region of the object. This is done by measuring "features" of the image. The next stage is the computation of structure and motion from these flow parameters, and the solution is given in the form of analytical expressions.

II OPTICAL FLOW AND FLOW PARAMETERS

Take a Cartesian arzy-coordinate system on the image plane and the z-axis perpendicular to It. Consider a plane moving in the scene. Let $z = px + qy + r$ be its equation. Let $(0, 0, r)$, the intersection between the plane and the z-axis, be a reference point. The motion is instantaneously specified by translation velocity (a, b, c) at the reference point and rotation velocity $(\omega_1, \omega_2, \omega_3)$ around it (*i.e.*, with rotation axis orientation $(\omega_1, \omega_2, \omega_3)$ and angular velocity $\sqrt{(\omega_1)^2 + (\omega_2)^2 + (\omega_3)^2}$ (rad/sec) screwwise around it). Our goal is to compute p, q, r, a, b, c, ω_1, ω_2 and ω_3 from an image sequence without using point-to-point correspondence.

Let $(0, 0, -f)$, the point away from the image plane by f on the negative side, be the viewpoint. A point (X, Y, Z) in space is projected to $(fX/(f + Z), fY/(f + Z))$ on the image plane. If the point is on the plane $z = px + qy + r$ moving as described above, it induces the following "optical flow" at point (x, y) on the image plane (*cf.* Kanatani [8, 9]:

$$\begin{aligned} u &= u_0 + Ax + By + (Ex + Fy)x, \\ v &= v_0 + Cx + Dy + (Ex + Fy)y, \end{aligned} \tag{1}$$

where 8 parameters u_0, v_0, A, B, C, D, E and F are given by

$$\begin{aligned} u_0 &= fa/(f + r), \qquad v_0 = fb/(f + r), \\ A &= p\omega_2 - (pa + c)/(f + r), \\ B &= q\omega_2 - \omega_3 - qa/(f + r), \\ C &= - p\omega_1 + \omega_3 - pb/(f + r), \\ D &= - q\omega_1 - (qb + c)/(f + r), \\ E &= (\omega_2 + pc/(f + r))/f, \\ F &= (- \omega_1 + qc/(f + r))/f. \end{aligned} \tag{2}$$

In other words, what we are viewing is a very restricted form of motion whose velocities are specified only by 8 parameters u_0, v_0, A, B, C, D, E and F. If these parameters are the same, motions seem identical to the viewer. Hence, our procedure is divided into two stages. First, we will show how to detect the "flow parameters" u_0, v_0, A, B, C, D, E and F from an image sequence without using point-to-point correspondence. Next, we will show how to solve simultaneous non-linear equations (2) for p, q, r, a, b, c, ω_1, ω_2 and ω_3 not simply "numerically" but also "analytically." Here, the "focal length" f of the camera is assumed to be a known constant.

If we take the limit $f \to \infty$ of a large focal length f, we obtain the following "orthographic approximation"

$$u_0 = a, \qquad v_0 = b,$$

* On leave from the Department of Computer Science, Gunma University, Kiryu, Gunma 376, Japan.

$$\begin{aligned} A &= p\omega_2, & B &= q\omega_2 - \omega_3, \\ C &= -p\omega_1 + \omega_3, & Q &= -q\omega_1, \\ E &= 0, & F &= 0, \end{aligned} \tag{3}$$

and if we omit terms of $O(1/f^2)$ but retain terms of $O(1/f)$, we obtain the following "pseudo-orthographic approximation"

$$\begin{aligned} u_0 &= fa/(f + r), \qquad v_0 = fb/(f + r), \\ A &= p\omega_2 - (pa + c)/(f + r), \\ B &= q\omega_2 - \omega_3 - qa/(f + r), \\ C &= -p\omega_1 + \omega_3 - pb/(f + r), \\ D &= -q\omega_1 - (qb + c)/(f + r), \\ E &= \omega_2/f, \qquad F = -\omega_1/f. \end{aligned} \tag{4}$$

Here, we consider only the planar motion. However, if the object is not planar, we can decompose the object surface image into small planar or almost planar regions by fitting the form of eqns (1) to the observed flow, say by the least square error method. Thus, the subsequent analysis applies to objects of arbitrary shape.

III ESTIMATION OF FLOW PARAMETERS FROM FEATURES

Let $X(x, y)$ represent an image. For example, if the image consists of gray-levels, $X(x, y)$ denotes its intensity at point (x, y). If the image consists of colors, $X(x, y)$ may be a vector value function corresponding to R, G and B. If the image consists of points and lines, $X(x, y)$ has delta-function-like singularities. In any case, we define a "feature" $F[X]$ of image $X(x, y)$ as a "functional," *i.e.*, a map $F[.]$ from the set of images $X(x, y)$ to real numbers.

Consider the time change when there exists an optical flow $(u(x, y), v(x, y))$ on the image plane. If $X(x, y)$ is an image at time t, it changes at time $t + \delta t$ after a short time interval into

$$X(x - u(x, y)\delta t,\ y - v(x, y)\delta t) = X(x, y) - X_x(x, y)u(x, y)\delta t - X_y(x, y)v(x, y)\delta t + \ldots, \tag{5}$$

where X_x and X_y are partial derivatives. Then, the corresponding feature $F[X]$ becomes $F[X] + \mathrm{D}F[X]\delta t + \ldots$, where in general the "change rate" $\mathrm{D}F[.]$ is a "linear functional" in $u(x, y)$ and $v(x, y)$. In view of the optical flow of eqns (1), this means that we have a "linear" equation of the form

$$\mathrm{D}F[X] = C_1[X]u_0 + C_2[X]v_0 + C_3[X]A + C_4[X]B + C_5[X]C + C_6[X]D + C_7[X]E + C_8[X]F. \tag{6}$$

Here, $C_1[.]$, ..., $C_8[.]$ are functionals derived from the given feature $F[.]$, so that they are known functionals. The change rate $\mathrm{D}F[.]$ can be estimated by difference schemes. For example, observe an image at time t and compute its feature $F(t)$. Next, observe the image at time $t + \delta t$ after a short time interval and compute its feature $F(t + \delta t)$. Then, $\mathrm{D}F[X]$ is approximated by $(F(t + \delta t) - F(t))/\delta t$ or by other higher order difference schemes. Thus, all quantities except u_0, v_0, A, B, C, D, E and F in eqn (6) are directly computed from the image sequence without using point-to-point correspondence. Hence, if we prepare 8 or more independent functionals $F_1[.]$, $F_2[.]$, ..., we obtain a set of simultaneous "linear" equations in u_0, v_0, A, B, C, D, E and F of the form of eqn (6) to determine them.

The idea of using features was already introduced by Amari [10, 11] and Kanatani [7 - 9]. However, they did not divide the process into two stages as described here but tried to compute p, q, r, a, b, c, ω_1, ω_2 and ω_3 directly. This leads to a set of simultaneous "non-linear" equations, which is difficult to solve. Kanatani [7 - 9] proposed an iterative method which traces the motion along time, starting from known initial values of p, q and r. Here, however, we divide the process into two stages and first determine the "flow parameters," which can be computed by solving a set of "linear" equations. This poses no computational problem. The desired p, q, r, a, b, c, ω_1, ω_2 and ω_3 are given in terms of the flow parameters as is shown subsequently.

As for the feature functionals, we can choose those used by Amari [10, 11] and Kanatani [7 - 9]. Amari [10, 11] used weighted integral (or "filter") $F[X] = \iint m(x, y)X(x, y)dxdy$ of various $m(x, y)$ over a fixed window for gray-level images. Invoking a mathematics called "stereology," Kanatani [7] used, for textured surfaces, Fourier coefficients of function $N(\theta)$, where $N(\theta)$ is the number of intersections, per unit length, between parallel lines of orientation θ and the texture on the image plane. When no texture exists and only circumference contours are available, Kanatani [8] used Fourier coefficients of $D(\theta)$, which is the caliper diameter of the contour measured by two parallel lines of orientation θ. Kanatani [9] also used line integral $\int m(x, y)ds$ of various $m(x, y)$ along the contour and surface integral $\iint m(x, y)dxdy$ of various $m(x, y)$ inside the contour. In any case, a set of linear equations of the form of eqn (6) is obtained, and the flow parameters are determined immediately. However, the accuracy and reliability heavily depends on the choice of the features.

IV STRUCTURE AND MOTION FROM FLOW PARAMETERS

Suppose we have already computed the flow parameters u_0, v_0, A, B, C, D, E and F by the method described in the previous section. Or, if the point-to-point correspondence happens to be available, they are immediately determined, say by the least-square-error fitting of eqns (1). What we want is p, q, r, a, b, c, ω_1, ω_2 and ω_3. First, compute

$$\begin{aligned} U_0 &= u_0 + iv_0, \qquad T = A + D, \qquad R = C - B, \\ S &= (A - D) + i(B + C), \qquad K = E + iF, \end{aligned} \tag{7}$$

where i is the imaginary unit, and hence U_0, K and S are complex numbers. Define complex variables $V = a + ib$, $P = p + iq$ and $W = \omega_1 + i\omega_2$. Then, V, c, P, r, W and ω_3 are given as follows.

In the case of the orthographic approximation, we get

$$V = U_0, \qquad \omega_3 = (R \pm \sqrt{SS^* - T^2})/2,$$
$$W = ke(\pi/4+\arg(S)/2-\arg(2\omega_3-(R+iT))/2), \qquad (8)$$
$$P = Se(\pi/4-\arg(S)/2+\arg(2\omega_3-(R+iT))/2)/k,$$

where $e(.)$ denotes $\exp(i.)$, arg the argument and $*$ the complex conjugate. Here, k is an indeterminate scale factor. Thus, (i) the absolute depth r and the velocity c in the z-direction are indeterminate, (ii) an indeterminate scale factor k is involved, and (iii) there exist two types of solutions, one is the true one and the other a spurious one. They are indistinguishable because they yield identical flow parameters. However, if we observe two or more planar regions of the same rigidly moving object, we can pick up the true one because ω_1, ω_2 and ω_3 must be common to all. The fact that an indeterminate scale factor k is necessarily involved was already pointed out by Sugihara and Sugie [5], but the existence of the spurious solution and the explicit forms of eqns (8) have not been known.

In the case of the pseudo-orthographic approximation, we get

$$V/(f + r) = U_0/f, \quad W = ifK, \quad P = S/(fK - U_0/f),$$
$$\omega_3 = (R + \mathrm{Re}[P(W^* + iU_0^*/f)])/2, \qquad (9)$$
$$c/(f + r) = -(T + \mathrm{Im}[P(W^* + iU_0^*/f)])/2,$$

where Re[.] and Im[.] designate the real and the imaginary part, respectively. Hence, (i) the absolute depth r is indeterminate, but (ii) $a/(f + r)$, $b/(f + r)$, $c/(f + r)$, p, q, ω_1, ω_2 and ω_3 are uniquely determined.

In the case of the pure central projection, we obtain

$$V/(f + r) = U_0/f, \qquad c/(f + r) = c',$$
$$P(c') = (fK-U_0/f\pm\sqrt{(fK-U_0/f)^2-4c'S})/2c',$$
$$W(c') = i(fK-U_0/f\mp\sqrt{(fK-U_0/f)^2-4c'S})/2+iU_0/f, \qquad (10)$$
$$\omega_3 = (R + \mathrm{Re}[P(c')(W(c')^* + iU_0^*/f)])/2,$$
$$c' = -(T + \mathrm{Im}[P(c')(W(c')^* + iU_0^*/f)])/2. \qquad (11)$$

Here, P and W are given as functions of c', and c' is determined from eqn (11). Eqn (11) is proved to have only one non-zero solution. Since the uniqueness of the solution is guaranteed, a simpliest solution method is to assume an appropriate value of $c' = c/(f + r)$, say by the pseudo-orthographic approximation (9), compute the right-hand side of eqn (11) and repeat the process, using the new value of c', until convergence. We see that (i) the absolute depth r is indeterminate, (ii) $a/(f + r)$, $b/(f + r)$ and $c/(f + r)$ are uniquely determined and (iii) there exist two sets of solutions for p, q, ω_1, ω_2 and ω_3, one is the true one and the other a spurious one, and they are indistinguishable because they yield the same flow parameters. The spurious solution is eliminated by observing two or more planar regions of the same rigidly moving object because ω_1, ω_2 and ω_3 must to be common to them.

Numerical schemes of recovering 3D structure and motion from point-to-point correspondence pairs have been known [1 - 4], and the existence of the spurious solution was pointed out by Longuet-Higgins [6]. However, analytical expressions like eqns (10) and (11) have not been known. The parameters of eqns (7) have physical meanings: U_0 "translation," T "divergence," R "rotation," S "shearing," and K "fanning." They are transformed by a coordinate rotation by θ on the image plane as

$$U_0 \to U_0e(-\theta), \qquad T \to T, \qquad R \to R,$$
$$S \to Se(-2\theta), \qquad K \to Ke(-\theta), \qquad (12)$$

i.e., U_0 and K (as well as V, P and W) are (relative) invariants of "weight" - 1 (or "vectors"), S is an (relative) invariant of weight - 2 (or a "tensor"), and T and R (as well as r, c and ω_3) are (absolute) invariants of weight 0 (or "scalars").

REFERENCES

[1] Ullman, S. *The Interpretation of Visual Motion.* Cambridge, Mass.: MIT Press, 1979.

[2] Nagel, H.-H. "Representation of moving rigid objects based on visual observations." *Computer* 14:8 (1981) 29 - 39.

[3] Longuet-Higgins, H. C. "A computer algorithm for reconstructiong a scene from two projections." *Nature* 239:10 (1981) 133 - 135.

[4] Tsai, R. Y. and T. S. Huang, "Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces." *IEEE Trans.* PAMI-6 (1984) 13-27.

[5] Sugihara, K. and N. Sugie, "Recovery of rigid structure from orthographically projected optical flow." *Computer[1] Vision, Graphics, and Image Processing* 27 (1984) 309 - 320.

[6] Longuet-Higgins, H. C. "The visual ambiguity of a moving plane." *Proc. R. Soc. Lond.* B-223 (1984) 165 - 175.

[7] Kanatani, K. "Detection of surface orientation and motion from texture by a stereological technique." *Artificial Intelligence* 23 (1984) 213 - 237.

[8] Kanatani, K. "Tracing planar surface motion from projection without knowing correspondence." *Computer Vision, Graphics, and Image Processing* 29 (1985) 1-12.

[9] Kanatani, K. "Detecting the motion of a planar surface by line and surface integrals." *Computer Vision, Graphics, and Image Processing* 29 (1985) 13 - 22.

[10] Amari, S. "Invariant structures of signal and feature spaces in pattern recognition problems." *RAAG Memoirs* 4 (1968) 553 - 566.

[11] Amari, S. "Feature spaces which admit and detect invariant signal transformations" In *Proc. 4th Int. Joint Conf. Pattern Recognition, Tokyo, 1978,* pp. 452 - 456.