コンピュータビジョン 48－4
（1987. 5. 21）

# ノイズに強いオプティカルフローからの３次元復元

金谷 健一　　　　吉田 淳一郎
群馬大学工学部情報工学科　　　　沖ファームウェア・システムズ㈱

運動する多面体の３次元形状をそのオプティカルフローから復元する新しい手法を提案する。これは、物体が多面体であるという拘束条件を利用して、誤差の影響を最小限に抑えるものである。まず、物体の各面の「勾配」（グラジエント）を未知数とする方程式を立てる。これによって、各面の平面性が自動的に保証される。そして、解が「金谷の解析的公式」によって直ちに与えられる。これによって Marr の意味の「2.5Dスケッチ」が得られる。しかしこれは一般に誤差を含んでいるので、面の接続に不適合が生じて、多面体形状が復元できない。そこで、多面体の隣接構造に基づく「杉原の最適化手法」を改良したものを適用する。これにより、実用的な精度で形状復元ができることを、実際の画像を用いた例によって示す。

# NOISE ROBUST 3D RECOVERY FROM OPTICAL FLOW

Ken-ichi Kanatani* and Jun-ichiro Yoshida**

* Department of Computer Science, Gunma University, Kiryu, Gunma 376, Japan
** Oki Firmware System, Co., Takasaki, Gunma 370, Japan

A new scheme is presented to recover the 3D shapes of polyhedra from optical flow. The effect of noise is minimized by exploiting the fact that the object is a polyhedron. First, the *gradients* of the faces are taken as unknowns, which assures planarity of the reconstructed faces and also enables us to use *Kanatani's analytical solution* directly. Then, a *2.5D sketch* in the sense of Marr is obtained. However, a consistent polyhedron cannot be reconstructed from it due to incompatibility of face adjacency if error is present. In this paper, a variant of *Sugihara's optimization technique* based on the polyhedron adjacency structure is applied. As a result, the 3D shape can be reconstructed with sufficient accuracy for practical purposes. An example is shown by using real images.

## 1. INTRODUCTION

Reconstructing the 3D shape of an object moving rigidly in a scene from a sequence of camera images is known as the *shape-from-motion* problem. Since humans apparently have such ability, this problem has long attracted psychologists. The first extensive study from the viewpoint of artificial intelligence was done by Ullman <1> in relation to human visual perception, and his formulation has given a dominant influence over subsequent studies of computer vision.

The first step is *correspondence detection*, namely identification of *feature points* of the object, say particular vertices or prominent surface markings, in each image frame. If the feature points are moving rigidly in the scene as a whole, mathematical analysis shows that their 3D positions are determined from the image coordinates of these feature points (up to some degree of indeterminacy depending on the number of feature points and the number of image frames.

Later, more systematic approaches were explored <2, 3>. It turned out, however, that all these mathematical schemes are very sensitive to noise, and computed solutions are too inaccurate for practical use. There even exist very pessimistic views that noise vulnerability is *inherent* to the problem itself and there *can not exist* stable schemes at all. Are they really true?

Granted that the problem becomes inevitably unstable *in general* (i.e., for feature points in general position), a stable solution may be obtained if we make use of some *knowledge* about the object. Suppose the object is a polyhedron, and its vertices are chosen as feature points. If an existing scheme is applied directly in the presence of noise, the computed 3D positions of vertices adjacent to a face may not necessarily be coplanar. But we know that the object is a polyhedron and that they must be coplanar. Can we exploit this fact in some way?

One way to incorporate this knowledge is to choose, as unknowns, not the positions of vertices but the *gradients* of the object faces. As a result, computed solutions necessarily have planar faces. The advantage is not limited to this; *analytical closed solution* is known for planar surfaces in instantaneous motion <4 - 7>. In this paper, we use the results of Kanatani <7>.

The knowledge of the gradients of all visible faces results in a *2.5D sketch* (Marr <8>). However, if these gradient values are not exact, we may not be able to reconstruct a polyhedron which is correctly projected onto the observed image and yet whose faces exactly have the specified gradient values. In this paper, we obtain a consistent polyhedron whose faces have gradient values closest to the specified values *on the average* by applying Sugihara's optimization <9, 10>.

Sugihara <10> analyzed line drawings of polyhedra in detail and completely classified indeterminacies in interpretation. He presented a general combinatoric procedure to find the *degree of freedom* with which arbitrary values can be assigned without violating the condition that the object is a polyhedron correctly projected onto the observed image. He proposed least square optimization to determine the remaining independent indeterminate values from various other sources of information (e.g., light reflectance, texture density, etc.).

However, Sugihara's scheme as proposed is a non-linear optimization, and the convergence is not guaranteed. This is because his scheme is defined in general terms. In this paper, we show that if we have a *2.5D sketch*, i.e., estimate gradient values of the (visible) faces, the problem reduces to a set of simultaneous *linear* equations: no iterations and no worry about convergence. This procedure predicts a consistent polyhedron correctly projected onto the observed image. Hence, if the solution is not exact, it is useful in many practical situations, since it has no inconsistencies. We demonstrate this by an example, using real images.

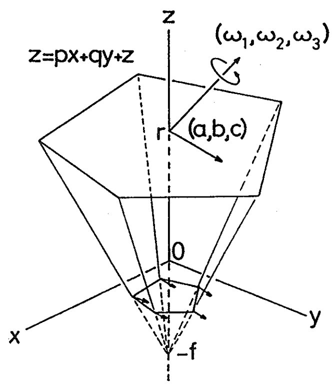



**Fig. 1** A camera image can be modeled as perspective projection from a viewpoint (0, 0, $-f$), where $xy$-plane is the image plane.

図１．カメラ画像は、点(0,0,-f)を視点とする、xy平面上への中心投影とみなせる。

## 2. OPTICAL FLOW OF A PLANAR FACE

Take an $xyz$-coordinate system in the scene, and regard the $xy$-plane as the image plane. A camera image can be modeled as *perspective projection* from a *viewpoint* (0, 0, $-f$) on the negative side of the $z$-axis (Fig. 1). The position of the viewpoint corresponds to the center of the lens, and the distance $f$ between the image plane and the viewpoint is regarded as the *focal length* of lens.[1] A point ($X$, $Y$, $Z$) in the scene is projected onto the intersection of the image plane with the *ray* connecting the point and the viewpoint. Its image coordinates ($x$, $y$) are

1 This is exact as long as the object is far away. If the object is near the camera, some correction is necessary.

given as follows:[2]

$$x=fX/(f+Z), \quad y=fY/(f+Z). \tag{2.1}$$

Suppose a polyhedron is moving in the scene, and let $z = px + qy + r$ be the equation of one particular face. Parameters $p$, $q$ designate the *gradient* of the face. Let us call $r$ the *depth* of the face.

An instantaneous rigid motion is resolved into *translation* $(a, b, c)$ at an arbitrarily fixed *reference point* and *rotation* $(\omega_1, \omega_2, \omega_3)$ around it (i.e., with $(\omega_1, \omega_2, \omega_3)$ as the axis orientation and $\sqrt{\omega_1^2+\omega_2^2+\omega_3^2}$ (rad/sec) as the angular velocity around it). Let $(0, 0, r)$, the intersection of the face (or its extension) with the $z$-axis, be the reference point. It is shown that the velocities of points on the face produces on the image plane the following *optical flow*:

$$u=u_0+Ax+By+(Ex+Fy)x,$$

$$v=v_0+Cx+Dy+(Ex+Fy)y. \tag{2.2}$$

Here, the coefficients, which we call the *flow parameters*, are given as follows <7>:

$$u_0=\frac{fa}{f+r}, \quad v_0=\frac{fb}{f+r},$$

$$A=p\omega_2-\frac{pa+c}{f+r}, \quad B=q\omega_2-\omega_3-\frac{qa}{f+r},$$

$$C=-p\omega_1+\omega_3-\frac{pb}{f+r}, \quad D=-q\omega_1-\frac{qb+c}{f+r},$$

$$E=\frac{1}{f}(\omega_2+\frac{pc}{f+r}), \quad F=\frac{1}{f}(-\omega_1+\frac{qc}{f+r}), \tag{2.3}$$

Suppose the face in question has four corners, and let $(x_i, y_i)$, $i = 1, \ldots, 4$, be their image coordinates. If we measure the velocities $(u_i, v_i)$, $i = 1, \ldots, 4$, at these points, the flow parameters $u_0$, $v_0$, $A$, $B$, $C$, $D$, $E$, $F$ are determined, from eqs. (2.2), by solving the following simultaneous linear equations:

$$\begin{bmatrix} 1 & & x_1 & y_1 & & & x_1^2 & x_1y_1 \\ 1 & & x_2 & y_2 & & & x_2^2 & x_2y_2 \\ 1 & & x_3 & y_3 & & & x_3^2 & x_3y_3 \\ 1 & & x_4 & y_4 & & & x_4^2 & x_4y_4 \\ & 1 & & & x_1 & y_1 & x_1y_1 & y_1^2 \\ & 1 & & & x_2 & y_2 & x_2y_2 & y_2^2 \\ & 1 & & & x_3 & y_3 & x_3y_3 & y_3^2 \\ & 1 & & & x_4 & y_4 & x_4y_4 & y_4^2 \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ A \\ B \\ C \\ D \\ E \\ F \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \tag{2.4}$$

The determinant of the above matrix, which is equal to

$$\begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix} \cdot \begin{vmatrix} 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{vmatrix} \cdot \begin{vmatrix} 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \\ 1 & x_1 & y_1 \end{vmatrix} \cdot \begin{vmatrix} 1 & x_4 & y_4 \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix}, \tag{2.5}$$

does not vanish unless three among the four points are collinear. Hence, if no three corners are collinear, the flow parameters are uniquely determined from the velocities at four corners. If the face has more than four corners, where velocities are observed, we can determine the flow parameters more accurately, say by the least-square method.

In the following, we assume that any (visible) corner vertex of the object is adjacent to at least one face which has four corners or more. The purpose of this paper is to show that estimation accuracy can be increased by taking advantage of polyhedron consistency constraints. As the number of corners of a face increases, the constraints become more and more strong, and hence estimation accuracy is expected to increase.

## 3. ANALYTICAL 3D RECOVERY OF FACES

Suppose the flow parameters $u_0$, $v_0$, $A$, $B$, $C$, $D$, $E$, $F$ are determined for each face by the procedure in the previous section. Then, the rotation and the gradient of the face are determined by solving eqs. (2.3) with known values of $u_0$, $v_0$, $A$, $B$, $C$, $D$, $E$, $F$. The solution is given in analytical forms <4 - 7>. Here, we follow Kanatani <7>.

First, compute the following quantities.[3]

$$T=A+D, \quad R=C-B,$$

$$S=(A-D)+i(B+C), \quad L=(E-u_0/f)+i(F-v_0/f). \tag{3.1}$$

Here, $i$ is the imaginary unit, so that $U_0$, $S$, $L$ are complex numbers in general.

It can be proved that the following cubic equation has three real roots in general <7>:

$$X^3+TX^2+\frac{1}{4}(T^2-|S|^2-|L|^2)X$$

$$+\frac{1}{8}(\mathrm{Re}[L^2S]-T|L|^2)=0. \tag{3.2}$$

Here, $|\cdot|$ denotes the absolute value of a complex number, and $\mathrm{Re}[\cdot]$, $\mathrm{Im}[\cdot]$ the real part and the imaginary part, respectively. Let $\alpha$ be the middle of the three real roots of eq. (3.2).

2 The limit $f \to \infty$ of an infinite focal length corresponds to *orthographic projection*.

3 These quantities are derived as *invariants* with respect to coordinate rotation on the image plane according to *group representation theory*, cf. Kanatani <24> for details.

The gradient components p, q, and the the rotation velocities $\omega_1$, $\omega_2$, $\omega_3$ are given as follows <7>:[4]

$$p=\frac{1}{2\alpha}\mathrm{Re}\,[L\pm\sqrt{L^2-4\alpha S}]\,,$$

$$q=\frac{1}{2\alpha}\mathrm{Im}\,[L\pm\sqrt{L^2-4\alpha S}]\,,$$

$$\omega_1=-\frac{1}{2}\mathrm{Im}\,[L\mp\sqrt{L^2-4\alpha S}]-\frac{v_0}{f},$$

$$\omega_2=\frac{1}{2}\mathrm{Re}\,[L\mp\sqrt{L^2-4\alpha S}]+\frac{u_0}{f},$$

$$\omega_3=\frac{1}{2}R\pm\mathrm{Im}\,[L^*\sqrt{L^2-4\alpha S}]\,. \tag{3.3}$$

Here, $*$ denotes complex conjugate, and one particular branch is chosen for the complex square root <7>.

Eqs. (3.3) show that there exist two sets of solutions. However, since the rotation velocities $\omega_1$, $\omega_2$, $\omega_3$ are common to all the faces, we can pick up only the true solutions if two or more faces are observed. In the presence of noise, however, we need a *clustering technique* in the three-dimensional parameter space of $\omega_1$, $\omega_2$, $\omega_3$.

Thus, the gradient of each face is uniquely determined. However, we should note that the computed gradient values may *not be exact*. Even if the observed images are precise and exact measurement is possible, the "velocity" must be approximated by the "displacement" (divided by the time lapse), which may not be regarded as an instantaneous velocity no matter how short the time interval is. Besides, noise on images and error in measurements are usually inevitable. As a result, the computed gradient values may not be able to define a consistent polyhedron. We consider this problem in detail in the next section.

## 4. 3D RECONSTRUCTION FROM A 2.5D SKETCH

Consider an object image. Suppose preprocessings such as edge detection, segmentation and object boundary detection are already done, resulting in a line drawing of the object. Let us call such a line drawing a *2.5D sketch* if the gradient of (or equivalently the normal vector to) the object surface is estimated at each point according to Marr <8> (Fig. 2).

Estimation of the surface gradient can be done in many ways, and the use of optical flow (i.e. *shape from motion*) is one of them. Other attempts frequently made include observation of surface texture (i.e. *shape from texture*), measuring distortion of texture elements <11> or texture density <12> and measurement of light reflectance (i.e. *shape from shading* <13>. Projection of regular patterns onto the object surface is another possibility <14>.

4 We assume that $\alpha$ is not zero. If it happens to be zero, another set of formulae must be used. See Kanatani <7> for details.

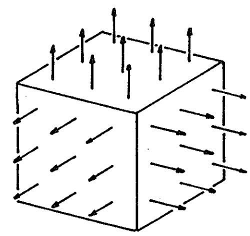


**Fig. 2** If estimates of the surface gradients or the surface normal vectors of the object are given, the image is called a *2.5D Sketch* (Marr <8>).

図2．物体面の勾配または法線ベクトルの推定値が与えられている画像を「2.5Dスケッチ」と呼ぶ（Marr <8>）。

Now, a 2.5D sketch can be regarded as a polyhedron image; the image domain can be decomposed, according to the estimated gradient values, into small polygonal regions (or *planar patches*) in such a way that the gradient is almost the same in each of them. (If the object is a polyhedron, its faces themselves are planar patches), Theoretically, the 3D object shape can be reconstructed by by connecting these planar patches one by one in the scene according to the specified orientations in such a way that each of them is projected onto the corresponding region on the image plane. However, if the estimated gradient values are not accurate, the reconstructed object shape depends on the order of patch there may arise *incompatibility* of face adjacency; two faces may not meet with a common boundary (Fig. 3). Since most techniques of surface gradient measurement almost invariably introduce noise and error, we cannot always obtain the 3D shape of the object from a given 2.5D sketch.

The above consideration can be rephrazed as follows: *Being a polyhedron is a strong constraint, and its faces cannot be assigned the gradient values arbitrarily.* Implications of this fact were studied in detail by Sugihara <9, 10> as a corollary of his algebraic description of constraints on polyhedron images, and he proposed to obtain, starting from a given 2.5D sketch, a polyhedron which satisfies necessary constraints and yet has face gradient values closest to the estimated values *on the average*. His method goes as follows.

Given a polyhedron image, the *adjacency structure* must be determined. Let $V_i$, $i = 1, \ldots, n$, be the vertices and $F_\alpha$, $\alpha = 1, \ldots, m$, be the faces. The adjacency structure is specified by a set of *adjacency pairs* $(F_\alpha, V_i)$ (meaning that point $V_i$ is adjacent to face $F_\alpha$). The first thing to do is check if the adjacency structure is *regular* or not. If it is non-regular, the adjacency pairs are "overspecification" and 3D reconstruction is possible only for special cases <10>. The necessary and sufficient condition for regularity is[5]

$$V(\mathfrak{F})+3|\mathfrak{F}| \geq R(\mathfrak{F})+4, \quad (4.1)$$

for any subset $\mathfrak{F}$ of the $m$ faces $\{F_\alpha | \alpha=1,...,m\}$, such that $|\mathfrak{F}| \geq 2$. Here, $|\mathfrak{F}|$ is the number of faces in subset $\mathfrak{F}$, $V(\mathfrak{F})$ is the number of vertices adjacent to at least one of the faces in subset $\mathfrak{F}$, and $R(\mathfrak{F})$ is the number of adjacency pairs involving the faces in subset $\mathfrak{F}$ <10>. If eq. (4.1) is not satisfied, we ignore some (minimal number of) adjacency pairs so that the adjacency structure becomes regular.

If the adjacency structure is regular, it can be shown that the *degree of freedom* the *rank* is

$$r=n+m-l, \quad (4.2)$$

where $l$ is the number of adjacency pairs, meaning that we can assign at most $r$ arbitrary values to the positions of the vertices and gradients of the faces <10>. In other words, if we choose an appropriate set of $r$ variables (or *basis*), the polyhedron is uniquely reconstructed and the positions of its vertices and the gradients of its faces are expressed in terms of these $r$ variables.

Sugihara <9, 10> proposed to use these independent variables for optimization, searching the $r$ dimensional space for values for which the corresponding polyhedron best fits additional requirements resulted from other sources of information (e.g., shape from texture or shape from shading). However, this is a non-linear optimization, requiring iterative search, and convergence is not guaranteed. In the following, we show that the optimization reduces to a set of simultaneous *linear* equations, so that an optimal solution can be obtained without any iterations.

Sugihara first described his formulation under orthographic projection and then suggested a way to reduce perspective projection to orthographic projection. However, his formalism under perspective projection is not convenient in our setting, i.e., optimization of a 2.5D sketch. In the following, we present a convenient formulation for perspective projection.

## 5. OPTIMIZATION OF A 2.5D SKETCH

Let $V_i$: $(X_i, Y_i, Z_i)$, $i = 1, ..., n$, be the 3D positions of the vertices of the polyhedron under consideration, and let $F_\alpha$: $z = p_\alpha x + q_\alpha y + r_\alpha$, $\alpha = 1, ..., m$, be the equations of its faces. The condition that vertex $V_i$ is adjacent to face $F_\alpha$ is expressed by

$$Z_i=p_\alpha X_i+q_\alpha Y_i+r_\alpha. \quad (5.1)$$

The image coordinates $(x_i, y_i)$ are related to the 3D coordinates $(X_i, Y_i, Z_i)$, $i = 1, ..., n$, corresponding point is given by eqs. (2.1).

5 This result is true unless the corresponding polyhedron has an exceptional structure (e.g., three faces sharing a common edge). For details, see Sugihara <10>.

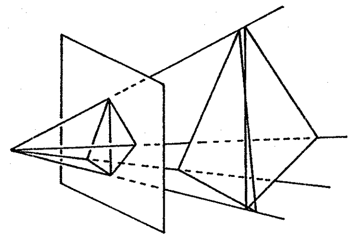


Fig. 3 The 3D shape cannot be reconstructed due to the incompatibility of face adjacency.

図3．面の接続の不適合により、3D復元ができない。

Now, we introduce a new quantity defined by $z_i = fZ_i/(f+Z_i)$ in analogy of eqs. (2.1), and use

$$x_i=\frac{fX_i}{f+Z_i}, \quad y_i=\frac{fY_i}{f+Z_i}, \quad z_i=\frac{fZ_i}{f+Z_i}, \quad (5.2)$$

instead of $X_i$, $Y_i$, $Z_i$. Solving eqs. (5.2) for $X_i$, $Y_i$, $Z_i$, we obtain

$$X_i=\frac{fx_i}{f-z_i}, \quad Y_i=\frac{fy_i}{f-z_i}, \quad Z_i=\frac{fz_i}{f-z_i}. \quad (5.3)$$

Substituting eqs. (5.3) into eq. (5.1), we obtain

$$z_i=\frac{fp_\alpha}{f+r_\alpha}x_i+\frac{fq_\alpha}{f+r_\alpha}y_i+\frac{fr_\alpha}{f+r_\alpha}. \quad (5.4)$$

Next, we define the following new quantities:

$$P_\alpha=\frac{fp_\alpha}{f+r_\alpha}, \quad Q_\alpha=\frac{fq_\alpha}{f+r_\alpha}, \quad R_\alpha=\frac{r_\alpha}{f+r_\alpha}. \quad (5.5)$$

They are non-dimensional quantities. The inverse of eqs. (5.5) is

$$p_\alpha=\frac{P_\alpha}{1-R_\alpha}, \quad q_\alpha=\frac{Q_\alpha}{1-R_\alpha}, \quad r_\alpha=\frac{fR_\alpha}{1-R_\alpha}. \quad (5.6)$$

Thus, eq. (5.1) is equivalently rewritten, from eq. (5.4), as

$$z_i=P_\alpha x_i+Q_\alpha y_i+fR_\alpha. \quad (5.7)$$

Since $x_i$, $y_i$, $i = 1, ..., n$, are the known image coordinates of the vertices, their 3D positions are determined by eqs. (5.3) if $z_i$, $i = 1, ..., n$, are known. Consequently, $z_i$, $i= 1, ..., n$, can be taken as unknowns about the 3D vertex positions. On the other hand, $P_\alpha$, $Q_\alpha$, $R_\alpha$, $\alpha = 1, ..., m$, can be regarded as unknowns, equivalent to $p_\alpha$, $q_\alpha$, $r_\alpha$, for the faces. Thus, we obtain $l$ (= the number of adjacency pairs) equations of the form of eq. (5.7), which are the only constraints on 3D reconstruction. (We can assume that the adjacency structure is already regular by the argument in the previous section.)

For optimal 3D reconstruction, let us use the least square method, minimizing

$$J=\frac{1}{2}\sum_{\alpha=1}^{m} w_\alpha [(p_\alpha-\hat{p}_\alpha)^2+(q_\alpha-\hat{q}_\alpha)^2], \quad (5.8)$$

where $\hat{p}_\alpha$, $\hat{q}_\alpha$ are the estimated gradient components of face $F_\alpha$, and $w_\alpha$ is the weight of face $F_\alpha$. If eqs. (5.6) are substituted, eq. (5.8) becomes

$$J=\frac{1}{2}\sum_{\alpha=1}^{m} w_\alpha \left(\frac{f+r_\alpha}{f}\right)^2 \left[ (P_\alpha+\hat{p}_\alpha R_\alpha-\hat{p}_\alpha)^2 + (Q_\alpha+\hat{q}_\alpha R_\alpha-\hat{q}_\alpha)^2 \right]. \tag{5.9}$$

Now, if face $F_\alpha$ has large depth $r_\alpha$, the gradient estimation may not be so accurate, and hence a small weight $w_\alpha$ should be assigned. At the same time, if $\hat{p}_\alpha$ and $\hat{q}_\alpha$ have very large magnitudes, the measurement may be inaccurate, and hence a small weight $w_\alpha$ should be assigned. In view of these considerations, it seems appropriate to choose, as the weight,

$$w_\alpha=\frac{1}{\hat{p}_\alpha^2+\hat{q}_\alpha^2}\left(\frac{f}{f+r_\alpha}\right)^2. \tag{5.10}$$

The problem is optimization of $J$ constrained by eqs. (5.7) for all adjacency pairs $(P_i, F_\alpha)$. Hence, by introducing Lagrangean multipliers $\Lambda_{\alpha i}$ corresponding to all existing adjacency pairs $(P_i, F_\alpha)$, the problem is converted to unconstrained optimization of

$$\frac{1}{2}\sum_{\alpha=1}^{m}\frac{1}{\hat{p}_\alpha^2+\hat{q}_\alpha^2}\left[ (P_\alpha+\hat{p}_\alpha R_\alpha-\hat{p}_\alpha)^2+ (Q_\alpha+\hat{q}_\alpha R_\alpha-\hat{q}_\alpha)^2\right] + \sum_{\alpha,i}\Lambda_{\alpha i}(P_\alpha x_i+Q_\alpha y_i+fR_\alpha-z_i). \tag{5.11}$$

Taking derivatives with respect to $P_\alpha$, $Q_\alpha$, $R_\alpha$, and $z_i$, we obtain the following equations.[6]

$$P_\alpha+\hat{p}_\alpha R_\alpha+\sum_i(\hat{p}_\alpha^2+\hat{q}_\alpha^2)x_i\Lambda_{\alpha i}=\hat{p}_\alpha, \quad \alpha=1,\ldots,m, \tag{5.12}$$

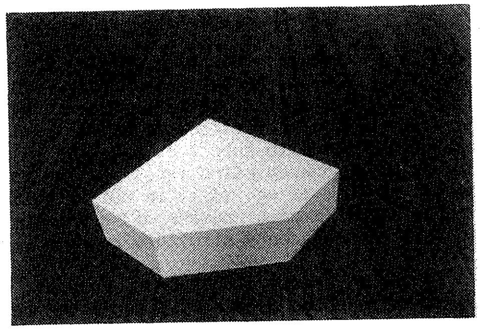
$$Q_\alpha+\hat{q}_\alpha R_\alpha+\sum_i(\hat{p}_\alpha^2+\hat{q}_\alpha^2)y_i\Lambda_{\alpha i}=\hat{q}_\alpha, \quad \alpha=1,\ldots,m, \tag{5.13}$$

$$\sum_i(\hat{p}_\alpha x_i+\hat{q}_\alpha y_i-f)\Lambda_{\alpha i}=0, \quad \alpha=1,\ldots,m, \tag{5.14}$$

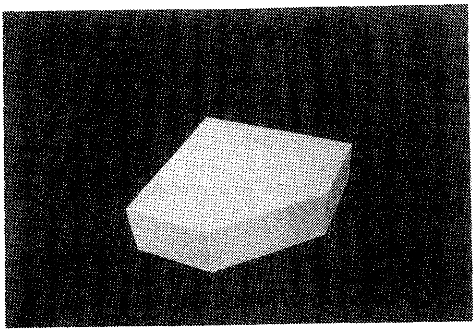
$$\sum_\alpha\Lambda_{\alpha i}=0, \quad i=1,\ldots,n. \tag{5.15}$$

Here, the summations range over existing adjacency pairs. Eqs. (5.12) - (5.15) together with eqs. (5.7) provide $n + 3m + l$ equations for unknowns $z_i$ $(i = 1, \ldots, n)$, $P_\alpha$, $Q_\alpha$, $R_\alpha$, $(\alpha = 1, \ldots, m)$, and the $l$ Lagrange multipliers $\Lambda_{\alpha i}$. However, since a 2.5D sketch, by definition, defines only the object shape, and the absolute distance from the viewer cannot be determined, we must give, as input, the value of $Z_i$ (and hence of $z_i$) of one particular vertex, removing the corresponding equation from eqs. (5.15).

Thus, the optimization of a 2.5D sketch can be done without any iterations; all we need is to solve a set of simultaneous *linear* equations. Once a solution for $z_i$, $i = 1, \ldots, n$, is obtained, the 3D positions $(X_i, Y_i, Z_i)$, $i = 1, \ldots, n$, of the vertices are recovered by eqs.



(a)



(b)

**Fig. 4** Two consecutive images of an object.

図４．連続した２枚の物体画像。

(5.3).[7]

## 6. EXAMPLE

Consider the two images of Fig. 4. Let us label the vertices and the faces as shown in Fig. 5. The adjacency structure is given by adjacency pairs

$(F_1, V_1)$, $(F_1, V_4)$, $(F_1, V_5)$, $(F_1, V_7)$, $(F_1, V_8)$,
$(F_2, V_3)$, $(F_2, V_4)$, $(F_2, V_5)$, $(F_2, V_6)$,
$(F_3, V_1)$, $(F_3, V_2)$, $(F_3, V_3)$, $(F_3, V_4)$,
$(F_4, V_1)$, $(F_4, V_2)$, $(F_4, V_8)$, $(F_4, V_9)$.

We can easily check that the condition (4.1) is satisfied, so that the adjacency structure is regular. Consequently, the degree of freedom is four, and hence the object can be reconstructed in terms of four basis variables. However, we do not select any basis variables; we apply the optimization process to all the variables as shown below.

First, we measured the 2D positions $(x_i, y_i)$ of the vertices on Fig. 4 and obtained the displacements of the vertices shown in Fig. 6. In this experiment, no precise calibration was

6 Eq. (5.14) is obtained by differentiating eq. (5.11) with respect to $R_\alpha$, and substituting eqs. (5.12) and (5.13).

7 Sugihara <9, 10> also added *inequality constraints* which indicate that some vertices should be nearer to the viewer than others. Here, we can add some appropriate inequality constraints, too.

done, so that the focal length $f$ may contain a few percent error and the above measured values may contain about 10 percent error. Moreover, we approximated the theoretically instantaneous velocities by the observed finite displacements, which contains considerable amount of error.

We can obtain the flow parameters of the faces by regarding the displacements as instantaneous velocities (taking the time lapse between the two images as unit time), and applying eq. (2.4). The gradient of each face can be computed from eqs. (3.1) - (3.3). First, two gradients and two rotation velocities are obtained for each face, but the common rotation velocity is easily identified, though the value is somewhat different from face to face. We discarded, for each rotation velocity components, two extreme values (i.e., the maximum and the minimum) and took average of the rest. Assuming that this average gives the correct rotation velocity, we recomputed the gradient of each face by using the fourth and the fifth of eqs. (2.3).

Applying the optimization of the previous section, we obtain the 3D coordinates of the vertices. Fig. 7 shows the "top view" (orthographic projection onto the $yz$-plane) and the "side view" (orthographic projection onto the $zx$-plane. In spite of the presence of noise and inaccuracy of the estimated gradient values, the final result is fairly correct. Many experiments with synthetic data show that we can reconstruct the 3D object shape with sufficient accuracy for many practical applications if the motion is not infinitesimal, (as long as it is small and confined within a certain range).

## 7. CONCLUDING REMARKS

We have shown a new method to reconstruct 3D object shape from motion in the presence of noise, assuming that the object is a polyhedron. The process consists of three stages. First, given two consecutive images, we compute the "flow parameters" of each face from the vertex-to-vertex correspondences. Then, the "gradient" of each face is computed by Kanatani's analytical formulae <7>. Finally, from thus obtained "2.5D sketch", the 3D shape is reconstructed by optimization. The last stage is performed by solving a set of simultaneous "linear" equations. A fairly reasonable object shape is reconstructed in spite of noise and error involved in observation and measurement. We demonstrated this by an example, using real images.

The optimization technique presented here is also useful in many problems other than "shape from motion"; it can be applied whenever the surface gradient can be estimated by some means. In many 3D recovery problems from light reflectance and photometric stereo, it is assumed that the object has a smooth surface (e.g., spherical or cylindrical), and the projection is orthographic. Then, the image domain is decomposed into many small meshes and the 3D shape is reconstructed by employing an appropriate *smoothness constraint* saying that changes between meshes are very small (e.g., minimization of Laplacian), cf. <13>. However, appropriateness of a particular smoothness constraint cannot be checked a priori, and the method does not work well if the decomposition is very

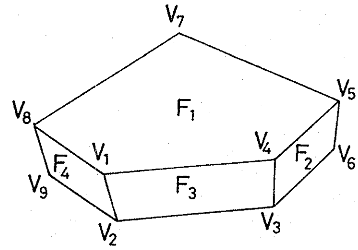



**Fig. 5** The adjacency structure of the line drawing of the object in Fig. 4.

図5. 図4の物体の線画の隣接構造。

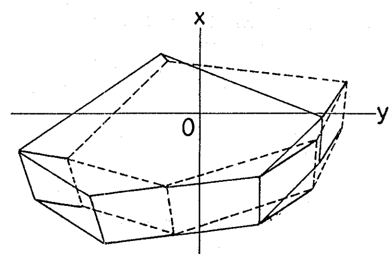



**Fig. 6** Displacements (*optical flow*) of the vertices of the object in Fig. 4.

図6. 図4の物体の頂点の移動("オプティカルフロー")。

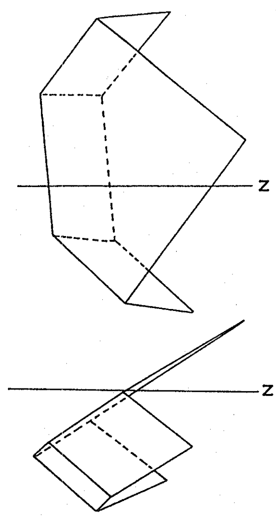



**Fig. 7** The "top view" (orthographic projection onto the $yz$-plane) and "side view" (orthographic projection onto the $zx$-plane) of the object reconstructed from the optical flow of Fig. 6.

図7. 図6のオプティカルフローから再構成した物体の上面図 (yz面への平行投影) および側面図 (zx面への平行投影)。

coarse, if the object has corners and edges, or if the projection is perspective. Our method works well in such cases.

Although a 2.5D sketch is capable of determining the object shape only up to a scale factor, the reconstructed shape can be used for identification (Which of the assumed models does it correspond to?) and classification (Which of the assumed categories does it belong to?). In such a circumstance, the optimization technique presented in this paper is very useful.

There still remain some problems to be solved which we did not consider in this paper. One is the detection of optical flow, i.e., finding the correspondences between feature points. Many schemes have been tested for this purpose <15 - 18>, but this is a very difficult task. Recently, schemes of 3D recovery from motion without requiring the correspondence have also been proposed <19 - 21>.

Another is the technique of accurate camera calibration; since the shape-from-motion problem is sensitive to noise, accuracy must be pursued in every stage of the process if we apply the optimization in the final stage. Necessity of accurate calibration has been strongly recognized recently <22, 23>.

*Acknowledgments*: The authors thank Prof. Kokichi Sugihara of the University of Tokyo for discussions on the polyhedron optimization problem.

## REFERENCES


1. S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, MA, 1979.
2. H. C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, *Nature*, 293(10) (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. Machine Intell.*, PAMI-6(1) (1984), 13 - 27.
4. H. C. Longuet-Higgins, The visual ambiguity of a moving plane, *Proc. R. Soc. Lond.*, B-233 (1984), 165 - 175.
5. M. Subbarao and A. M. Waxman, Closed form solution to image flow equations for planar surfaces in motion, *Comput. Vision Graphics Image Process.*, 36 (1986), 208 - 228.
6. K. Kanatani, Structure and motion from optical flow under orthographic projection, *Comput. Vision Graphics Image Process.* 35 (1986), 181 - 199.
7. K. Kanatani, Structure and motion from optical flow under perspective projection, *Comput. Vision Graphics Image Process.* to appear.
8. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Fransisco, 1982.
9. K. Sugihara, An algebraic approach to shape-from-image problems, *Artif. Intell.*, 23 (1984), pp. 59 - 95.
10. K. Sugihara, *Machine Interpretation of Line Drawings*, MIT Press, Cambridge, MA, 1986.
11. K. Kanatani, Detection of surface orientation and motion from texture by a stereological technique, *Artif. Intell.*, 23 (1984), 213 - 237.
12. K. Kanatani and T.-C. Chou, Shape from texture: General principle, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Miami Beach, FL, June 1986, pp. 578 - 583.
13. B. K. P. Horn, *Robot Vision*, MIT Press, Cambridge, MA, 1986.
14. K. Sugihara, K. Okazaki, K.-H. Feng and N. Sugie, Regular pattern projection for surface measurement, *Proc. 2nd Int. Symp. Robotics Research*, Kyoto, Japan, August 1984, pp 231 - 238.
15. B. K. P. Horn and B. G. Schunck, Determining optical flow, *Artif. Intell.*, 17 (1981), 185 - 203.
16. J. M. Prager and M. A. Arbib, Computing the optic flow: the MATCH algorithm and prediction, *Comput. Vision Graphics Image Process.*, 24 (1982), 271 - 304.
17. H. Miike, Y. Kurihara, H. Hashimoto and K. Koga, Velocity-field measurement by pixel-based temporal mutual-correlation analysis of dynamic image, *Trans. IECE Japan*, E-69(8) (1986), 877 - 882.
18. H.-H. Nagel and W. Enkelmann, An investigation of smoothness constraints for the estimation of vector fields from image sequences, *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-8(5) (1986), 565 - 593.
19. K. Kanatani, Tracing planar surface motion from a 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. K. Kanatani and T.-C. Chou, Tracing finite motions without correspondence, *Proc. Int. Workshop Indust. Appl. Machine Vision Machine Intell.*, Tokyo, February 1987, pp. 118 - 123.
22. R. Y. Tsai, An efficient and accurate camera calibration technique for 3D machine vision, *Proc. IEEE Conf. Comput Vision Pattern Recog.*, Miami Beach, FL, June 1986, pp. 364 - 374.
23. O. D. Faugeras and G. Toscani, Camera calibration for 3D computer vision, *Proc. Int. Workshop Indust. Appl. Machine Vision Machine Intell.*, Tokyo, Japan, February 1987, pp. 240 - 247.
24. K. Kanatani, *Group Theoretical Methods on Image Understanding*, Springer (to appear).