Pergamon

FINDING THE MOTION, POSITION AND ORIENTATION OF A PLANAR PATCH IN 3D SPACE FROM SCALED-ORTHOGRAPHIC PROJECTION

Soo-Chang Pei† and Lin-Gwo Liou

Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

Abstract— For the time being, many methods about how to solve the motion and structure of a rigid object in 3D space by using point or line correspondences have been widely researched. In this paper, the projected 2D image shapes of planar patches in 3D space are used as our basic image features. We try to determine the motion parameters and plane equations of planar patches in 3D space from the relationships among their observed image shapes. Besides, scaled-orthographic projection is used instead of the original perspective projection because of its simpler and tractable mathematical formulas. Not only a closely-approximate model to perspective projection, it further solves the inherent depth indeterminacy which exists in orthographic projection. Moreover, after applying a pre-processing of rotation correction to the original image data, our estimation result can achieve a better numerical solution than that without this pre-processing. Simulation experiments show the excellence of our algorithms in estimating the motion parameters. Ambiguities, indeterminacy, and data degeneracy are also carefully discussed in this paper, which provides some important insights to the problems.

Planar patch   Affine transformation   Shape normalization   Perspective projection
Orthographic projection   Scaled-orthographic projection   Ortho-perspective projection

1. INTRODUCTION

Determination of the 3D structure and motion of a rigid object from its projected 2D image features (points, lines, or contours) is one of the most important tasks in computer vision. Most existing studies adopt *perspective projection* which is known to be one of the most suitable models for formulating the phenomenon of camera projection in real applications.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \left(\frac{1}{Z}\right) \begin{bmatrix} X \\ Y \end{bmatrix} \tag{1}$$

where (x, y) is the image of a point (X, Y, Z) in 3D space (see line 1 in Fig. 1). However, no matter how good it is, some inherent limits still exist when we want to *uniquely* recover the motion and structure of a 3D rigid object whose size is relatively small when compared with the object-to-camera distance.

As a well-known fact, the most important reason why the algorithms using perspective projection can often uniquely solve the motion parameters is based on the *depth-dependent* character of projective projection. Once this character cannot be clearly observed from the input image data (i.e. when the object size is comparatively small to its depth), these algorithms will fail because their solutions (although unique) are very sensitive to the observation errors. On the other hand, these algorithms are often complex and time-consuming because of the nonlinearity of perspective projection. It means some nonlinear searching method may

be introduced in these algorithms. Perhaps, for linearizing these nonlinear equations, several rigidity constraints are sometimes relaxed and the input data are thus over-used to solve some *intermediate* variables which are in fact dependent on each other. Both of these shortages should be avoided as much as possible.

In order to solve the problem of weak depth-dependent character of a far (or equivalently small) object, several simpler projection models are proposed.[1] Here $(X, Y, Z)^T$ is the 3D coordinate of a point, and $(x, y)^T$ is its projected point on the image plane.

1.1. Orthographic projection

$$\begin{bmatrix} x \\ y \end{bmatrix} = k \begin{bmatrix} X \\ Y \end{bmatrix}. \tag{2}$$

The value k is a scaling constant which will not be changed during the motion of the target (see line 2 in Fig. 1). This projection is used when the following two constraints are satisfied: (1) the target is placed far enough away; (2) the depth change resulting from motion will not substantially affect the apparent size of its images. That is to say, it does not consider the depth change during the motion of the target. Some related researches[2-13] were proposed in the past years.

1.2. Scaled-orthographic projection

This projection is proposed for further considering the change of apparent size of image during the motion. It is defined as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{Z_r} \begin{bmatrix} X \\ Y \end{bmatrix} \tag{3}$$

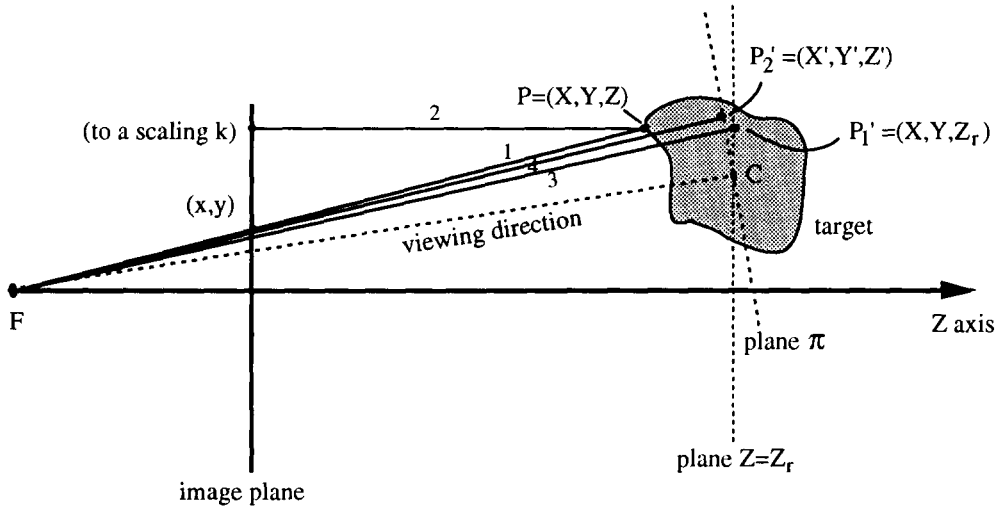† Author to whom all correspondence should be addressed.

Fig. 1. Four kinds of projection models. P is a point on the object. c is an appropriately-defined center of the object. Lines 1–4 separately represent the perspective, orthographic, scaled-orthographic, and ortho-perspective projection.
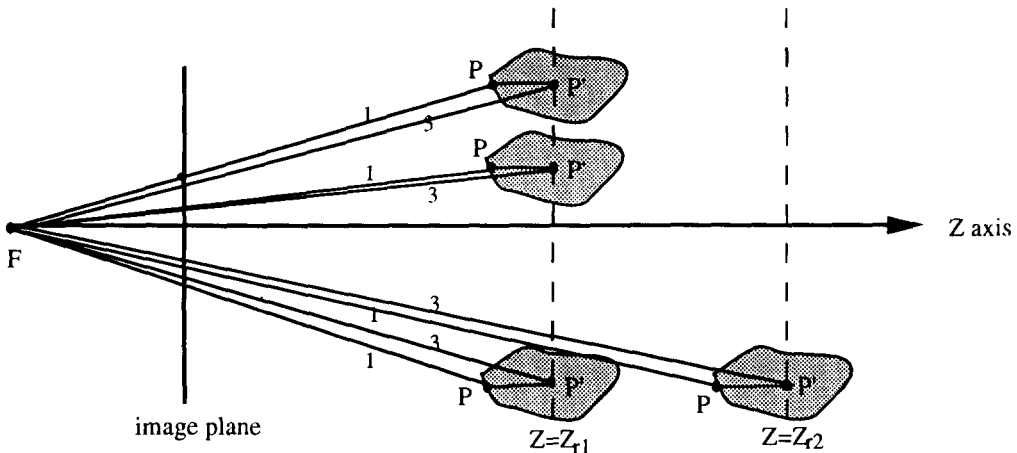

Fig. 2. Lines 1 and 3 separately represent the perspective projection and scaled-orthographic projection of a point P. It can be clearly observed that the perspective projection is better approximated by scaled-orthographic projection when the object is placed far from the camera or near to the Z-axis.

where Z_r is an *appropriate* positive value which depends on the depth of the observable parts of the target object in 3D space (see line 3 in Fig. 1). For example, the projective depth Z_r may be the Z-value of the object's center. Scaled-orthographic projection has a larger working depth range than that of orthographic projection. So, it is a better model. Besides, from Fig. 2, it is easy to observe that scaled-orthographic projection is closer to perspective projection when the object is placed farther away from the camera or nearer to the Z-axis.

1.3. *Ortho-perspective projection*

Before we introduce it, we define a plane Π which passes through the center c of the observable parts of the target object; the normal vector of this plane Π is

parallel to the viewing direction $\overline{FC}$ (see line 4 in Fig. 1). The projection is described as follows. First, along the direction of $\overline{FC}$, the point $P = (X, Y, Z)^T$ projects a point $P_2 = (X', Y', Z')^T$ on the plane Π; then we define the image point $(x, y)^T$ as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{Z'} \begin{bmatrix} X' \\ Y' \end{bmatrix}. \tag{4}$$

Obviously, the ortho-perspective projection is a *rotated* form of scaled-orthographic projection. It is a much better model because it further considers the *off-axis* property (see Fig. 2). Although its mathematical formula seems complex, it is not hard to use. For example, we can *suitably* rotate the coordinate system such that the viewing direction $\overline{FC}$ is lying on the new Z-axis, and the simpler formula of scaled-orthographic projection can be applied in the new coordinate system.
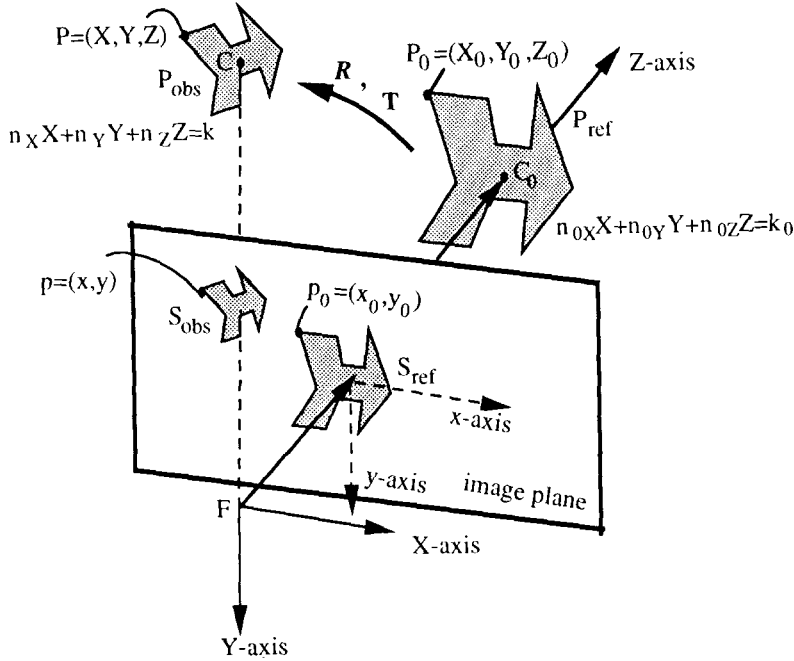
Fig. 3. Formulation of problem (I). Determine the position of P_{obs} by using the relationship between S_{ref} and S_{obs}. Here the position of the reference patch P_{ref} is known as prior information.

In this paper, we try to solve two closely-related problems ((I) and (II)) about how to determine the motion and position of a planar patch (or patches) in 3D space by using scaled-orthographic projection. References (14–20) are some related researches about using planar patches. Readers can see how the scaled-orthographic projection simplifies the mathematical derivations and makes a fairly-good estimation.

As a well-known fact, there exist some indeterminacy and ambiguities in orthographic projection. The scaled-orthographic projection has similar problems, too. In this paper, we will discuss how these indeterminacy and ambiguities happen.

Our paper is organized as follows: Section 2, the problem formulation and main algorithm of problem (I). Section 3, the problem formulation and main algorithm of problem (II). Section 4, discussions about the indeterminacy, ambiguities, and some degenerate cases. Section 5, simulation experiments. Section 6, final conclusion. Finally, the appendixes show (1) some basic relationships of an orthogonal matrix especially used in scaled-orthographic projection; (2) the methods of acquiring an affine transformation with or without correspondences.

Before leaving this section, we must emphasize one thing: We derive our algorithm by scaled-orthographic projection because we want to model the projective geometry in a simpler way. Of course, this adopted model should be as close to perspective projection as possible for fitting the practical use. Therefore, solving the general problems (i.e. done by perspective projection) by using a simpler and well-approximate model is our most important goal.

2. FORMULATION AND ALGORITHM FOR PROBLEM (I)

2.1. Problem formulation

Problem (I) is formulated as follows (see Fig. 3): Assume that a reference planar patch P_{ref}, placed at a *known* position and orientation, projects a shape S_{ref} on the image plane by *perspective* projection. After applying an unknown 3D rotation $\mathbf{R}$ and translation $\mathbf{T}$ to P_{ref}, the new patch P_{obs} projects another 2D shape S_{obs} on image plane by perspective projection, too. The new patch P_{obs} is placed far enough away such that the scaled-orthographic projection approximates the perspective projection quite well. Besides, we also assume the recognition problem has been solved previously. Now, our problem is: "Can we estimate the unknown position and orientation of the patch P_{obs} from the two projected shapes S_{ref} and S_{obs} by using the scaled-orthographic projection?"

2.2. Relationship between S_{ref} and S_{obs}

Without loss of generality, the camera coordinate system is considered as the global coordinate system. If a point P_0 on the planar patch P_{ref} is denoted by $\mathbf{P}_0 = (X_0, Y_0, Z_0)^T$, and its corresponding point P in P_{obs} by $\mathbf{P} = (X, Y, Z)^T$, we can define a coordinate transformation between these two points by

$$\mathbf{P} = \mathbf{R}\mathbf{P}_0 + \mathbf{T} \qquad (5)$$

where $\mathbf{R}$ is a 3×3 orthogonal matrix; $\mathbf{T}$ is a 3×1 translation vector.

For the special use of scaled-orthographic projection, it is convenient to separate the rotation matrix $\mathbf{R}$ and

translation T into the following forms:

$$R = \begin{bmatrix} R^* & r_1 \\ r_2^T & r_3 \end{bmatrix}; \quad T = \begin{bmatrix} T^* \\ T_3 \end{bmatrix} \quad (6)$$

where R^* is a 2×2 matrix; r_1, r_2, and T^* are 2×1 vectors; T_3 and r_3 are 1×1 constants.

If the reference patch P_{ref} is lying on a plane characterized by

$$n_{0X}X_0 + n_{0Y}Y_0 + n_{0Z}Z_0 = k_0 = \mathbf{n}_0 \cdot \mathbf{P}_0 \quad (7)$$

then the target patch P_{obs} must be lying on another plane characterized by

$$n_X X + n_Y Y + n_Z Z = k = \mathbf{n} \cdot \mathbf{P} = (\mathbf{R} \mathbf{n}_0) \cdot (\mathbf{R} \mathbf{P}_0 + \mathbf{T}) \quad (8)$$

$(|\mathbf{n}_0| = |\mathbf{n}| = 1)$.

Although the *exact* relationship (i.e. by perspective projection) between the two projected shapes S_{ref} and S_{obs} can be easily obtained by substituting equation (7) and the formula of perspective projection equation (1) into equation (6), its mathematical formula is too complicated for us (ignored here). Because the target patch P_{obs} is placed far enough away, we may use scaled-orthographic projection model instead of perspective projection model to approximate the relationship between S_{ref} and S_{obs}. After substituting the definition in equation (3) and the plane equation in equation (7), we have

$$\begin{bmatrix} x \\ y \end{bmatrix} = \left(\frac{Z_{r0}}{Z_r}\right)(R^* - r_1 q_0^T)\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \left(\frac{T^* + k_0' r_1}{Z_r}\right)$$

$$\equiv A\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + b \quad (9)$$

where $q_0^T = (n_{0X}, n_{0Y})/n_{0Z}$ and $k_0' = k_0/n_{0Z}$; $(x, y) = (X, Y)/Z_r$, $(x_0, y_0) = (X_0, Y_0)/Z_{r0}$; Z_r and Z_{r0} are two appropriate positive values. We may define the projective depth Z_r (or Z_{r0}) to be the Z-component of the center of the planar patch P_{obs} (or P_{ref}).

From equation (9), we know there approximately exists an affine transformation between the two projected shapes S_{ref} and S_{obs}. Without loss of generality, let us consider a special case when the known reference patch P_{ref} is lying on a plane $Z_0 = k_0 = Z_{r0}$ and its center is located at $(0, 0, Z_{r0})^T$. Then the affine parameters A and b in equation (9) can be written in a simpler form:

$$R^* = \left(\frac{Z_r}{Z_{r0}}\right)A; \quad \frac{T^*}{Z_{r0}} = \left(\frac{Z_r}{Z_{r0}}\right)b - \left(\frac{k_0'}{Z_{r0}}\right)r_1. \quad (10)$$

In Appendix B, we propose a method which can solve the affine parameters, A and b, without using point correspondences. In the following derivations, we assume the affine parameters have been obtained previously.

2.3. Determine the rotation matrix

For a more convenient description, we define several unit vectors as follows:

$$\hat{\mathbf{r}}_1 = \frac{\mathbf{r}_1}{\|\mathbf{r}_1\|}; \quad \hat{\mathbf{r}}_2 = \frac{\mathbf{r}_2}{\|\mathbf{r}_2\|}; \quad \hat{\mathbf{r}}_1^\perp = J\hat{\mathbf{r}}_1; \quad \hat{\mathbf{r}}_2^\perp = J\hat{\mathbf{r}}_2 \quad (11)$$

where

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

According to the orthogonality of the matrix R, we can decompose R^* into the following form (the details of derivations are listed in Appendix A):

$$R^* = [\hat{\mathbf{r}}_1 | \hat{\mathbf{r}}_1^\perp]\begin{bmatrix} -r_3 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \hat{\mathbf{r}}_2^T \\ \hat{\mathbf{r}}_2^\perp \end{bmatrix} = USV^T. \quad (12)$$

Therefore, from equation (10), AA^T and A^TA can be decomposed by

$$AA^T = U\left[\left(\frac{Z_{r0}}{Z_r}\right)^2 S^2\right]U^T; \quad A^TA = V\left[\left(\frac{Z_{r0}}{Z_r}\right)^2 S^2\right]V^T. \quad (13)$$

If we assume that the two eigenvalues of AA^T (or A^TA) are λ_1 and λ_2, and they satisfy $0 \le \lambda_1 \le \lambda_2$, it is easy to obtain

$$r_3 = \text{sign}(\det(A))\sqrt{\left(\frac{\lambda_1}{\lambda_2}\right)}; \quad c = \sqrt{\left(1 - \left(\frac{\lambda_1}{\lambda_2}\right)\right)} \ge 0 \quad (14)$$

and

$$\mathbf{r}_1 = c\hat{\mathbf{r}}_1 = c(\delta\mathbf{u}_1); \quad \mathbf{r}_2 = c\hat{\mathbf{r}}_2 = c(\delta\mathbf{v}_1); \quad \frac{Z_r}{Z_{r0}} = \frac{1}{\sqrt{\lambda_2}} > 0 \quad (15)$$

where the vector $\mathbf{u}_1$ (or $\mathbf{v}_1$) is the first column vector of U (or V); δ is either 1 or -1. Readers should pay great attention to the signs of the matrices U and V obtained by equation (13). These signs should satisfy equation (12) simultaneously.

From the results derived in equations (14) and (15), two answers of the rotation matrix R are obtained. One of them is the correct answer, and the other one is the *reflection* answer.

2.4. Determine the translation

First, we determine T^* from equation (10)

$$T^* = Z_r \mathbf{b} - Z_{r0}\mathbf{r}_1. \quad (16)$$

Because the center point $C_0 = (0, 0, Z_{r0})$ of the reference patch P_{ref} corresponds to the center point $C = (C_X, C_Y, C_Z = Z_r)$ of the target patch P_{obs}, we have (from equation (5))

$$T_3 = Z_r - r_3 Z_{r0}. \quad (17)$$

The plane equation of the target patch P_{obs} can be obtained by using equation (8). Now, we finish the task of solving the problem (I).

3. FORMULATION AND ALGORITHM FOR PROBLEM (II)

3.1. Problem formulation

Consider a rigid object moving (including rotate and translate) in 3D space. This object is also positioned
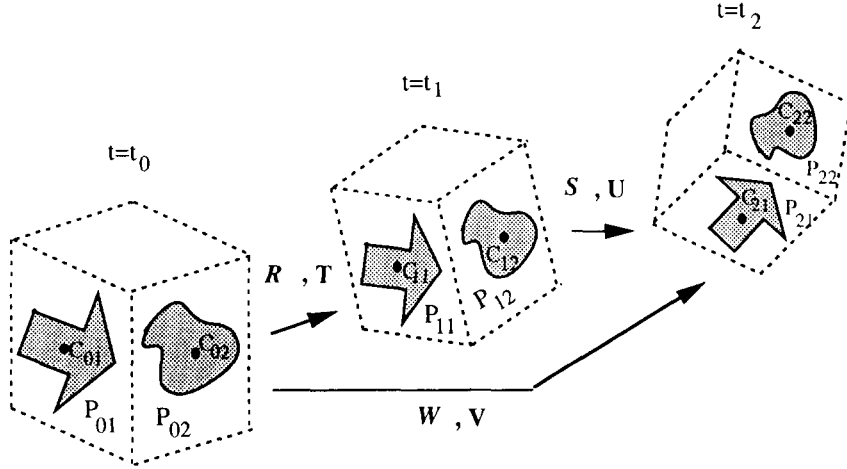
Fig. 4. Formulation of problem (II). There are N patches observed at three time instants by a camera. The jth planar patch P_{ij} at time t_i projects a 2D shape S_{ij} on the image plane. We try to determine the motion parameters and structure of these planar patches by using the relationships among the projected shapes S_{ij}s.

far enough away such that the scaled-orthographic projection can be used to approximate the perspective projection.

On this object, there are N planar patches which can be observed at three time instants $t_0, t_1,$ and t_2 by a steady camera (see Fig. 4). The jth planar patch at the time instant t_i is denoted by P_{ij}. This patch P_{ij} projects a 2D shape named S_{ij} on the image plane. The corresponding motion parameters including 3D rotation and translation of the object at different time instants are defined in Fig. 4 (rotation R, S, W; and translation T, U, V).

Now, the problem is: "How to estimate the motion and structure of the rigid object by using the approximate affine transform relationships among the corresponding image shapes $S_{0j}, S_{1j},$ and S_{2j} (for $j = 1$ to N)?"

3.2. Definition of variables

We first define the approximate affine relationships among the corresponding shapes $S_{0j}, S_{1j},$ and S_{2j} (index j is neglected in equations (18)–(20))

$$\begin{bmatrix} x_{i_2} \\ y_{i_2} \end{bmatrix} = A_{i_2 i_1} \begin{bmatrix} x_{i_1} \\ y_{i_1} \end{bmatrix} + b_{i_2 i_1} \qquad (18)$$

where $A_{i_2 i_1}$ and $b_{i_2 i_1}$ are the affine parameters from time t_{i_1} to t_{i_2}; $(x_i, y_i)^T$ is the corresponding image point at time t_i which is projected by scaled-orthographic projection

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{1}{Z_{ri}} \begin{bmatrix} X_i \\ Y_i \end{bmatrix}. \qquad (19)$$

Z_{ri} is an appropriate positive constant defined at time t_i which is similar to the projective depth Z_r described in equation (3). Here, the Z-component of the centroid $C_i = (C_{iX}, C_{iY}, C_{iZ})^T$ of the N planar patches P_{ij}s at a fixed time t_i is defined to be the projective depth Z_{ri}.

The plane equation for the jth planar patch P_{ij} at time t_i can be written as

$$n_{iX} X_i + n_{iY} Y_i + n_{iZ} Z_i = k_i. \qquad (20)$$

For the convenience in description, the rotation matrices $R, S,$ and W are similarly segmented as we have done in equation (6). (For examples, $S^*, W^*, s_1, s_2, w_1, w_2, s_3, w_3, \ldots,$ etc.) Similar to equation (15), we also define the unit vectors $\hat{s}_1, s_1^\perp, \hat{w}_1, \ldots,$ etc.

3.3. Solve the rotation matrices

From the derivation in equation (9), we have the following formula:

$$A_{10} = \left(\frac{Z_{r0}}{Z_{r1}} \right) (R^* - r_1 q_0^T) \qquad (21)$$

where $q_0 = [n_{0X}/n_{0Z}, n_{0Y}/n_{0Z}]^T$.

Pre-multiplying both sides of equation (21) by $(\hat{r}_1^\perp)^T$, we have

$$(\hat{r}_1^\perp)^T A_{10} = \left(\frac{Z_{r0}}{Z_{r1}} \right) [(\hat{r}_1^\perp)^T R^*] = -\left(\frac{Z_{r0}}{Z_{r1}} \right) (\hat{r}_2^\perp)^T \qquad (22)$$

where the relation $(\hat{r}_1^\perp)^T R^* = -(\hat{r}_2^\perp)^T$ is explained in Appendix A.

Equation (22) can be further rewritten as

$$(A_{10}^T J) \hat{r}_1 + J \left(\frac{Z_{r0}}{Z_{r1}} \hat{r}_2 \right) = 0. \qquad (23)$$

To the other affine matrices A_{21} and A_{20}, we can derive their relationships by using similar equations like equations (21) and (22).

$$(A_{21}^T J) \hat{s}_1 + J \left(\frac{Z_{r1}}{Z_{r2}} \hat{s}_2 \right) = 0 \qquad (24)$$

$$(A_{20}^T J) \hat{w}_1 + J \left(\frac{Z_{r0}}{Z_{r2}} \hat{w}_2 \right) = 0. \qquad (25)$$

For the jth planar patch at three time instants $(t_0, t_1,$ and $t_2)$, we have three sets of equations—equations (23)–(25). If the number N of the planar patches is larger than one $(N \geq 2)$, the vectors $\hat{r}_1, \hat{r}_2, \hat{s}_1, \hat{s}_2, \hat{w}_1,$

$\hat{\mathbf{w}}_2$, and the relative depths Z_{r1}/Z_{r0}, Z_{r2}/Z_{r0} can be easily solved. Notice that the solved normal vectors $(\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \hat{\mathbf{s}}_1, \ldots, \text{etc.})$ are still subject to a sign value $+1$ or -1.

From the derivations in Appendix A, we separate the rotation matrices $\mathbf{R}, \mathbf{S}$, and $\mathbf{W}$ by

$$\mathbf{R} = \begin{bmatrix} -r_3\hat{\mathbf{r}}_1\hat{\mathbf{r}}_2^T - \hat{\mathbf{r}}_1^{\perp}(\hat{\mathbf{r}}_2^{\perp})^T & c\hat{\mathbf{r}}_1 \\ c\hat{\mathbf{r}}_2^T & r_3 \end{bmatrix} \quad (26)$$

$$\mathbf{S} = \begin{bmatrix} -s_3\hat{\mathbf{s}}_1\hat{\mathbf{s}}_2^T - \hat{\mathbf{s}}_1^{\perp}(\hat{\mathbf{s}}_2^{\perp})^T & k\hat{\mathbf{s}}_1 \\ k\hat{\mathbf{s}}_2^T & s_3 \end{bmatrix} \quad (27)$$

$$\mathbf{W} = \begin{bmatrix} -w_3\hat{\mathbf{w}}_1\hat{\mathbf{w}}_2^T - \hat{\mathbf{w}}_1^{\perp}(\hat{\mathbf{w}}_2^{\perp})^T & p\hat{\mathbf{w}}_1 \\ p\hat{\mathbf{w}}_2^T & w_3 \end{bmatrix} \quad (28)$$

where $c = \delta_c|c|$, $k = \delta_k|k|$, and $p = \delta_p|p|$; δ_c, δ_k, and δ_p are either $+1$ or -1. Obviously, the undetermined signs of the solved unit vectors in equations (23)–(25) are included in the values of c, k, and p.

Because these rotation matrices must satisfy $\mathbf{W} = \mathbf{SR}$, we have

$$\mathbf{W}^* = \mathbf{S}^*\mathbf{R}^* + \mathbf{s}_1\mathbf{r}_2^T \quad (29)$$

$$\mathbf{w}_1 = \mathbf{S}^*\mathbf{r}_1 + r_3\mathbf{s}_1 \quad (30)$$

$$\mathbf{w}_2^T = \mathbf{s}_2^T\mathbf{R}^* + s_3\mathbf{r}_2^T \quad (31)$$

$$w_3 = \mathbf{s}_2^T\mathbf{r}_1 + s_3 r_3 = (kc)\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1 + s_3 r_3. \quad (32)$$

After substituting the $\mathbf{R}^*$ and $\mathbf{S}^*$ in equations (30) and (31) by the forms described in equations (26) and (27), we have

$$[-kr_3(\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1) + cs_3]\hat{\mathbf{s}}_1 + [-k(\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1^{\perp})]\hat{\mathbf{s}}_1^{\perp} = p\hat{\mathbf{w}}_1 \quad (33)$$

$$[-cs_3(\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1) + kr_3]\hat{\mathbf{r}}_2^T + [-c((\hat{\mathbf{s}}_2^{\perp})^T\hat{\mathbf{r}}_1)](\hat{\mathbf{r}}_2^{\perp})^T = p\hat{\mathbf{w}}_2^T. \quad (34)$$

If we separately dot (inner product) equations (33) and (34) by $\hat{\mathbf{s}}_1^{\perp}$ and $\hat{\mathbf{r}}_2^{\perp}$, we can solve c and k in terms of p.

$$c = \left(\frac{-\hat{\mathbf{w}}_1^T\hat{\mathbf{s}}_1^{\perp}}{(\hat{\mathbf{s}}_2^{\perp})^T\hat{\mathbf{r}}_1}\right)p = \alpha p; \quad k = \left(\frac{-\hat{\mathbf{w}}_2^T\hat{\mathbf{r}}_2^{\perp}}{\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1^{\perp}}\right)p = \beta p. \quad (35)$$

Next, we separately dot equations (33) and (34) by $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{r}}_2$; then we will have two linear equations of the two unknowns r_3 and s_3 as follows:

$$(\beta)r_3 + (-\alpha\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1)s_3 = (\hat{\mathbf{w}}_1^T\hat{\mathbf{s}}_1);$$

$$(-\beta\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1)r_3 + (\alpha)s_3 = (\hat{\mathbf{w}}_2^T\hat{\mathbf{r}}_2). \quad (36)$$

It is easy to uniquely solve r_3 and s_3 by equation (36). And from the orthogonality of $\mathbf{R}$ and $\mathbf{S}$, we have $|c| = \sqrt{(1 - r_3^2)}$ and $|k| = \sqrt{(1 - s_3^2)}$. Besides, $|p|$ is equal to $|k|/|\beta|$ or $|c|/|\alpha|$.

However, once δ_p is determined, the exact values of c and k can also be uniquely determined by the relations $c = \alpha p$ and $k = \beta p$. Then from equation (32), we can uniquely decide the value of w_3 by

$$w_3 = \mathbf{s}_2^T\mathbf{r}_1 + s_3 r_3 = (\alpha\beta p^2)\hat{\mathbf{s}}_2^T\hat{\mathbf{r}}_1 + s_3 r_3. \quad (37)$$

Now, with the known values c, k, p, r_3, s_3, and w_3, we can acquire the rotation matrices $\mathbf{R}, \mathbf{S}$, and $\mathbf{W}$, which completes the task of finding the rotation matrices.

Before leaving this section, two things must be noticed: (1) δ_p is still undetermined. Thus, two ambiguous solutions must be discussed. (2) Although we successfully solve the rotation matrices, it does not mean that the solved rotation matrices may always satisfy the characteristics of an orthogonal matrix exactly. It is because the affine matrices $\mathbf{A}_{i_2 i_1}$ are approximate results. Hence the final estimation should be slightly modified for satisfying the orthogonal constraints.

3.4. Solve the translation and the plane equation for P_{0j}

First, from equation (21), $\mathbf{q}_0$ can be determined (index j is neglected in equations (38) and (39))

$$\mathbf{q}_0^T = \frac{1}{c}\left[-\left(\frac{Z_{r1}}{Z_{r0}}\right)(\hat{\mathbf{r}}_1^T\mathbf{A}_{10}) + (\hat{\mathbf{r}}_1^T\mathbf{R}^*) \right] \quad (38)$$

and the orientation vector $\mathbf{n}_0$ of the planar patch P_{0j} is

$$\mathbf{n}_0 = (n_{0X}, n_{0Y}, n_{0Z})^T = \frac{(q_{0x}, q_{0y}, 1)^T}{\sqrt{(q_{0x}^2 + q_{0y}^2 + 1)}}. \quad (39)$$

Of course, the orientation vectors ($\mathbf{n}_1$ and $\mathbf{n}_2$) of the planar patches P_{1j} and P_{2j} are equal to $\mathbf{Rn}_0$ and $\mathbf{Wn}_0$.

Now, we want to determine the translation parameters $\mathbf{T}$ and $\mathbf{V}$. If the center of the planar patch P_{ij} is denoted by $\mathbf{C}_{ij}$, the average center $\mathbf{C}_i$ of the N planar patches P_{ij} at a fixed time t_i is defined as

$$\mathbf{C}_i = \sum_{j=1}^{N}\left(\frac{a_{ij}}{\sum_{j=1}^{N} a_{ij}}\right)\mathbf{C}_{ij} \quad (40)$$

where a_{ij} is the area of the planar patch P_{ij}. On the other hands, if the center and the area of the projected image shape S_{ij} is separately denoted by $\mathbf{c}_{ij}$ and a'_{ij}, equation (40) can be rewritten in a new form (under the scaled-orthographic projection)

$$\mathbf{C}_i = \sum_{j=1}^{N}\left(\frac{(a'_{ij}/|n_{ij,z}|)}{\sum_{j=1}^{N}(a'_{ij}/|n_{ij,z}|)}\right)\begin{bmatrix} Z_{ri}\mathbf{c}_{ij} \\ C_{ij,z} \end{bmatrix} \quad (41)$$

where the Z-component of $\mathbf{C}_i$ has been defined to be the projective depth Z_{ri} at time t_i, $n_{ij,z}$ is the Z-component of the orientation vector $\mathbf{n}_{ij}$.

We may define an average image center $\mathbf{c}_i$ at a fixed time t_i by

$$\mathbf{c}_i = \sum_{j=1}^{N}\frac{(a'_{ij}/|n_{ij,z}|)}{\sum_{j=1}^{N}(a'_{ij}/|n_{ij,z}|)}\mathbf{c}_{ij}. \quad (42)$$

So we have

$$\mathbf{C}_i = Z_{ri}\begin{bmatrix} \mathbf{c}_i \\ 1 \end{bmatrix}. \quad (43)$$

Because $\mathbf{C}_1 = \mathbf{RC}_0 + \mathbf{T}$ and $\mathbf{C}_2 = \mathbf{WC}_0 + \mathbf{V}$, we can easily determine the translation vectors

$$\frac{\mathbf{T}}{Z_{r0}} = \frac{Z_{r1}}{Z_{r0}}\begin{bmatrix} \mathbf{c}_1 \\ 1 \end{bmatrix} - \mathbf{R}\begin{bmatrix} \mathbf{c}_0 \\ 1 \end{bmatrix}; \quad \frac{\mathbf{V}}{Z_{r0}} = \frac{Z_{r2}}{Z_{r0}}\begin{bmatrix} \mathbf{c}_2 \\ 1 \end{bmatrix} - \mathbf{W}\begin{bmatrix} \mathbf{c}_0 \\ 1 \end{bmatrix}.$$

$$(44)$$

Until now, we have solved the rotation matrices $\mathbf{R}$ and $\mathbf{W}$, translation vectors $\mathbf{T}/Z_{r0}$ and $\mathbf{V}/Z_{r0}$, normal vector $\mathbf{n}_{0j}$ of the planar patch P_{0j}. The final unknown is the position vector of the center $\mathbf{C}_{0j}$ of the planar patch P_{0j}. Because the relationships $\mathbf{C}_{1j} = \mathbf{R}\mathbf{C}_{0j} + \mathbf{T}$ and $\mathbf{C}_{2j} = \mathbf{W}\mathbf{C}_{0j} + \mathbf{V}$ should be satisfied, we have

$$\begin{bmatrix} Z_{r1}\mathbf{c}_{1j} \\ C_{1j,z} \end{bmatrix} = \mathbf{R} \begin{bmatrix} Z_{r0}\mathbf{c}_{0j} \\ C_{0j,z} \end{bmatrix} + \mathbf{T};$$

$$\begin{bmatrix} Z_{r2}\mathbf{c}_{2j} \\ C_{2j,z} \end{bmatrix} = \mathbf{W} \begin{bmatrix} Z_{r0}\mathbf{c}_{0j} \\ C_{0j,z} \end{bmatrix} + \mathbf{V}. \qquad (45)$$

Then the value $C_{0j,z}/Z_{r0}$ can be easily solved by using

$$\left(\frac{Z_{r1}}{Z_{r0}}\right)\mathbf{c}_{1j} = \mathbf{R}^*\mathbf{c}_{0j} + \left(\frac{C_{0j,z}}{Z_{r0}}\right)\mathbf{r}_1 + \left(\frac{\mathbf{T}^*}{Z_{r0}}\right) \qquad (46)$$

and $\mathbf{C}_{0j}/Z_{r0} = (\mathbf{c}_{0j}, C_{0j,z}/Z_{r0})^{\mathrm{T}}$.

Now, the whole work is complete.

4. DISCUSSION

4.1. Indeterminacy

The so-called *indeterminacy* here is defined to be the quantity that is impossible to be determined only from the input image data. For example, the Z-com-

ponent of the 3D translation vector is an irrecoverable quantity if orthographic projection is used. However, there is *no* indeterminacy in scaled-orthographic projection *if* the projective depth Z_r has been appropriately defined. For examples, let us consider a rigid object which is translating in the direction of the Z-axis (see Fig. 5). If the projective depth Z_r is a constant which will not be affected by the depth of the rigid object, all of the projected images will be the same regardless of the object's depth. So the depth of the object is now an indeterminacy. On the contrary, if Z_r is defined to be the Z-component of the object's center, the change of object's depth will induce different images. So there is no indeterminacy any more.

Notice that the inherent indeterminacy of true values for monocular vision is not considered here. For example, all we have are *relative* answers such as $\mathbf{T}/Z_{r0}$ and Z_{r1}/Z_{r0}, not true values like $\mathbf{T}$ and Z_{r1}. In problem (I), we can obtain the true values because the value of Z_{r0} is known as prior information; however, in problem (II), only relative values are obtained because we do not know the exact value of Z_{r0}.

4.2. Ambiguity

The so-called *ambiguity* is the case when the image data cannot distinguish several obtained answers and
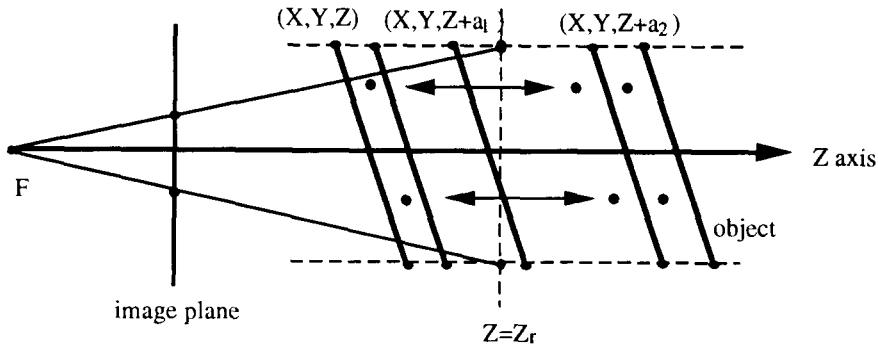

Fig. 5. Indeterminacy of depth in scaled-orthographic projection. Once the projective depth Z_r is fixed, any point on the line which passes through a fixed point (X, Y, Z) and parallel to the Z-axis will project to the same image point $(X/Z_r, Y/Z_r)$. Therefore, if all the points on the rigid object translate a random distance to the direction of the Z-axis, their projection image will not change.
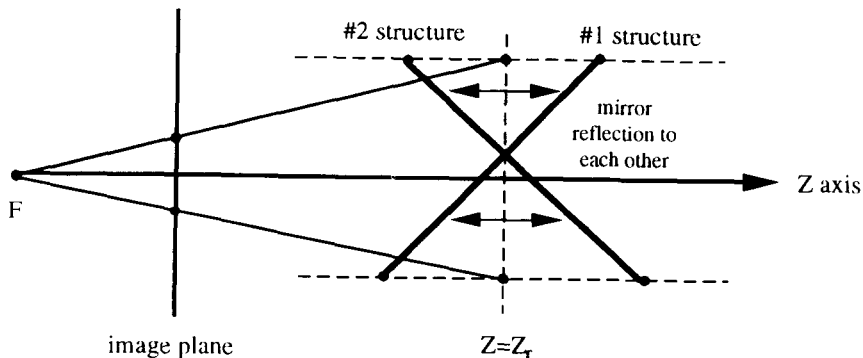

Fig. 6. Reflection structures (to the $Z = Z_r$ projective plane) when using scaled-orthographic projection. We also find that these two structures project the same image on the image plane.

decide which one is correct. Just as the orthographic projection does, the scaled-orthographic projection suffers from the ambiguities, that is *reflection*. The original object and its mirror object (to some plane $Z = Z_r$) in 3D space projects exactly the same image on the image plane if the scaled-orthographic projection model is used (see Fig. 6). If the motion is constrained to be rigid, each one of the original and mirror objects can *rigidly* move to project the same image data in an image sequence, which causes ambiguities. That is the reason why we can obtain two sets of solutions in problems (I) and (II).

Another possible ambiguity occurs when the affine transformation between two corresponding shapes is not unique. For example, if the reference patch P_{ref} used in problem (I) is a square, or rectangle, or circle, there is no unique affine transformation between S_{ref} and S_{obs}. In such a case, more ambiguities should be considered.

4.3. Degeneracy

The so-called *degeneracy* is the case when the input image data do not contain enough information. It usually results from a degenerate motion or degenerate structure.

In problem (I), data degeneracy happens *only* when the normal vector n of the observed patch P_{obs} is parallel to the image plane, which produces a line segment on the image plane. Of course, no meaningful affine transformation can be obtained now.

In problem (II), except for the similar ill-projection stated above, there are several degenerate cases (except the fourth one). We will briefly discuss them now.

1. If all the N planar patches on the target object have the same orientation vector (not necessarily all of them are on a plane), their affine matrices $A_{i_2i_1}$s are also the same (see equation (9)). (It means there are not enough independent equations for us to solve the unknown variables in equations (23)–(25).)

2. If any one of the rotation matrices R, S, and W is a rotation around the Z-axis (including the identity matrix), data degeneracy happens. Take R for example, the affine matrices A_{10}s for all the N planar patches on the object are equal because r_1 and r_2 are zero vectors (see equation (9)). However, when it really happens, we have no need to solve equation (23) any more. In fact, the affine matrix A_{10} is now exactly equal to $(Z_{r0}/Z_{r1})R^*$, which implies the R matrix can be uniquely determined.

How about the remaining two unknown rotation matrices S and W? After acquiring R and the unit vectors $\hat{s}_i$ and $\hat{w}_i$ (obtained by equations (24) and (25)), we still cannot determine these two matrices from equations (30)–(32). That is to say, data degeneracy does exist in this case.

Besides, if all of the matrices R, S, and W are rotations around the Z-axis, these rotation matrices will be easily determined. However, we still cannot determine the

normal vectors n_{ij}s of the planar patches P_{ij}s. Data degeneracy still exist!

3. If all the matrices R, S, and W are all identity matrices (means 3D translation only), the normal vectors n_{ij} still cannot be determined. In this case, only the relative projective depth values Z_{r1}/Z_{r0} and Z_{r2}/Z_{r0} can be determined.

4. A hazard case of problem (II) happens when $\hat{s}_2^T \hat{r}_1 = \pm 1$. In this situation, we cannot determine the parameters such as c, k, and p by using equations (35) and (36). However, we can determine those parameters by directly using the following two equations (similar to equation (21))

$$A_{10} = \left(\frac{Z_{r0}}{Z_{r1}}\right)(\hat{r}_1(-r_3\hat{r}_2^T - c q_0^T) - \hat{r}_1^\perp (\hat{r}_2^\perp)^T) \qquad (47)$$

$$A_{20} = \left(\frac{Z_{r0}}{Z_{r2}}\right)(\hat{w}_1(-w_3\hat{w}_2^T - p q_0^T) - \hat{w}_1^\perp (\hat{w}_2^\perp)^T). \qquad (48)$$

We first solve the two vectors: $a_{10} = -r_3\hat{r}_2^T - c q_0^T$ and $a_{20} = -w_3\hat{w}_2^T - p q_0^T$ from equations (47) and (48); de-compose q_0 into $k_1\hat{r}_2 + k_2\hat{r}_2^\perp$ and substitute it back to a_{10} and a_{20}. Finally, we can obtain R and W (still to a reflection). So it is not a degenerate case.

Of course, it seems too rough to say that the above-listed cases of data degeneracy in problem (II) contain all possible degenerate situations. At least, we provide some insights to data degeneracy. Remember that degenerate data always cause indeterminacy of some parameters.

5. SIMULATED EXPERIMENTS

5.1. Rotation correction

Before using our algorithms derived by scaled-orthographic projection, a pre-processing job called *rotation correction* must be executed first. In Section 1, we have said that the ortho-perspective projection is a rotated form of scaled-orthographic projection. On the other hand, we know scaled-orthographic projection is as good as ortho-perspective projection when the object's viewing direction is pointing toward the Z-axis. Therefore, we can suitably rotate the camera around its own focal point F such that the new projected image is near the center of the new image plane. Notice the new projected image can be directly generated by the old image if a suitable rotation $\bar{R}$ has been given. After the motion and structure parameters are obtained by using the new projected image, we can determine the true solutions by transforming the new coordinate system back to the old coordinate system. Details are not shown here.

5.2. Simulation experiments

Our simulation experiments are divided into two parts for proving our algorithms derived in Sections 3 and 4. The first goal of our experiments is to prove that our derivations are exactly correct if the image is projected by using the scaled-orthographic projection
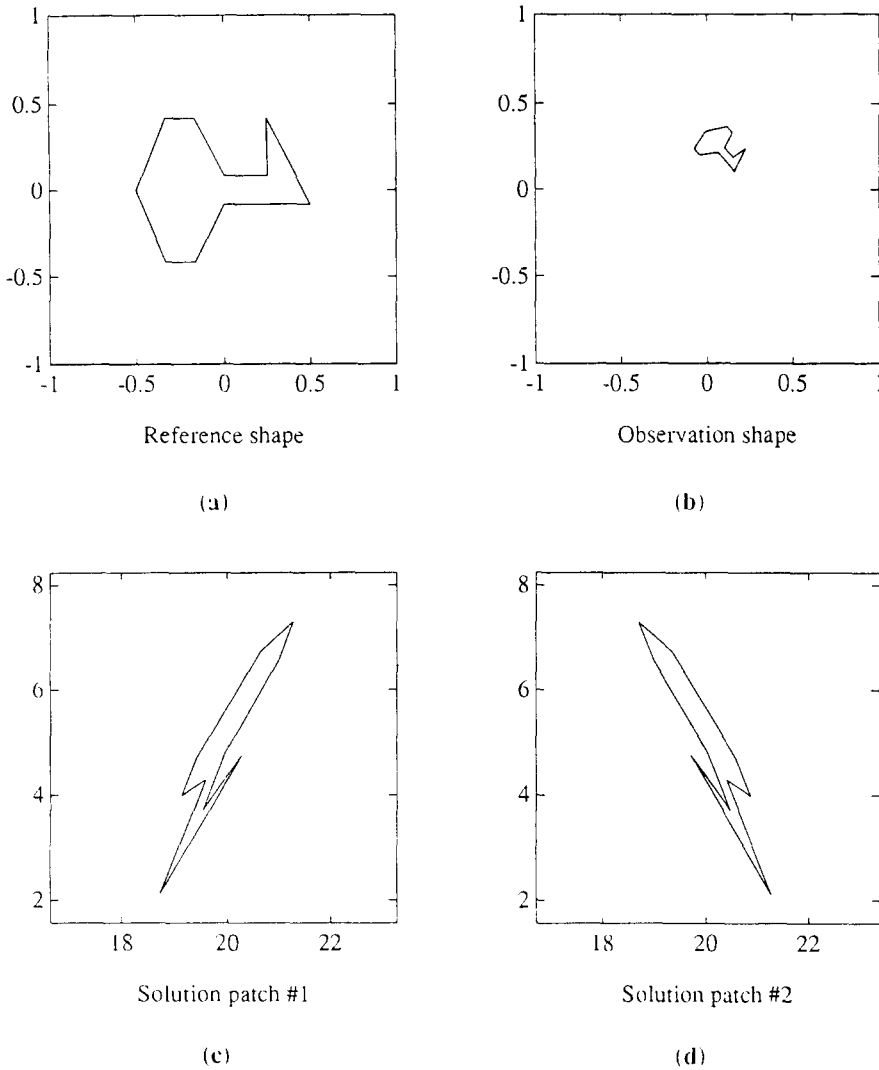
Reference shape

(a)

Observation shape

(b)

Solution patch #1

(c)

Solution patch #2

(d)

Fig. 7. Reference shape which is the image of the reference patch placed at center position $= (0,0,6)$ and orientation vector $= (0,0,1)$. (b) Observed shape with an unknown position and orientation. (c), (d) Two solutions (#1 and #2) are watched from their side lines to show their differences in structure. Here, #1 is the previously designed true solution. Both of them can project the same image on the image plane if scaled-orthographic projection is used as the image model.

model (no modeling errors). The second goal is to prove that our estimation is very close to the true solution if the image is projected by using the perspective projection model (considering modeling errors). The third goal is to test the error sensitivity. Experiments E1–E3 are designed for problem (I). Experiments E4–E6 are designed for problem (II).

E1. Figure 7(a) shows the reference shape S_{ref}. It is projected by the reference patch P_{ref} placed on the plane $Z_{r0} = 6$. The image center of S_{ref} has been moved to the origin of the image plane. Then we randomly assign a rigid motion $\mathbf{R}$ and $\mathbf{T}$ to P_{ref} to create a new patch P_{obs} which projects a new image shape S_{obs} by scaled-orthographic projection (the modeling errors are not considered, see Fig. 7(b)). Two estimated solution patches, #1 and #2, are displayed in Figs 7(c) and (d) (side-viewing). Estimated parameters are listed in Table 1. The #1 solution is almost the same as the true solution. These errors are introduced by quantization

Table 1. Comparison between the real solution and our estimation results of problem (I). S_{obs} is generated by scaled-orthographic projection for not considering the modeling errors

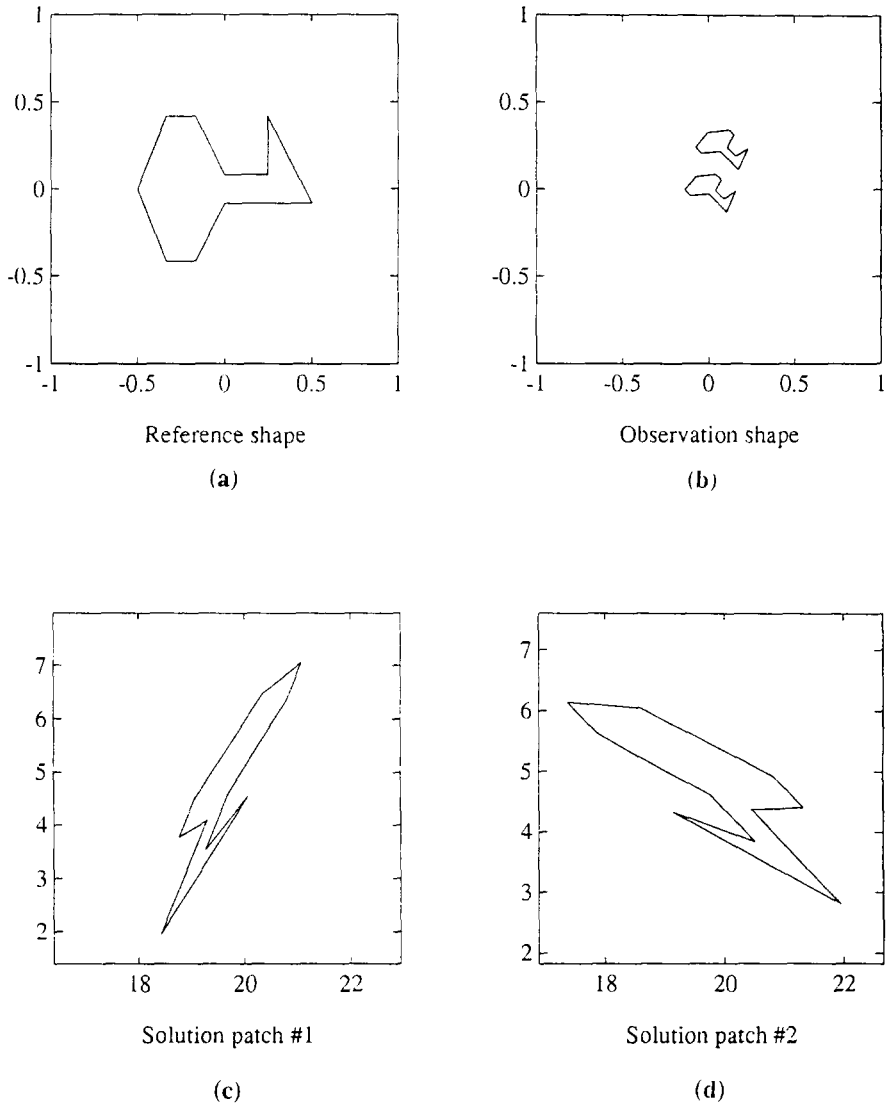| Parameters	Real	Est. #1	Est. #2	Err. of #1
Rotation	$(0.5000, 0.1000, -0.9000)$	$(0.5000, 0.1000, -0.9000)$	$(-0.5000, -0.1000, -0.9000)$	$5.6 \times 10^{-13}\%$
Center	$(1.3304, 5.0789, 20.0000)$	$(1.3304, 5.0789, 20.0000)$	$(1.3304, 5.0789, 20.0000)$	$2.5 \times 10^{-13}\%$
Orientation	$(-0.1225, -0.4566, 0.8812)$	$(-0.1225, -0.4566, 0.8812)$	$(-0.1225, -0.4566, -0.8812)$	$3.5 \times 10^{-15}\,\text{deg}$

Fig. 8. Similarly defined as that in Fig. 7. However, perspective projection is used as the projection model. The shape near the image center in (b) is the center-corrected image shape. All of the motion parameters are the same as that used in Fig. 7. Notice the #2 solution in (d), which is obviously different from that in Fig. 7(d).
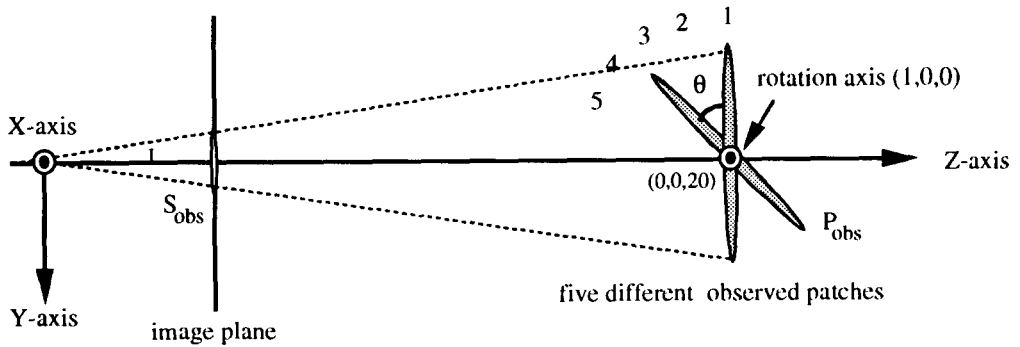

Fig. 9. Five cases designed in experiment E3. All of them use the same patch in previous experiments E1 and E2. Their centers are all placed at (0,0,20). The only difference among them is the orientation. Cases 1–5 are separately corresponding to the rotated patch in case 1 by rotating around axis (1,0,0) and angle 0, 15, 30, 45, and 60 deg.

The standard derivation δ of added error.

(a)

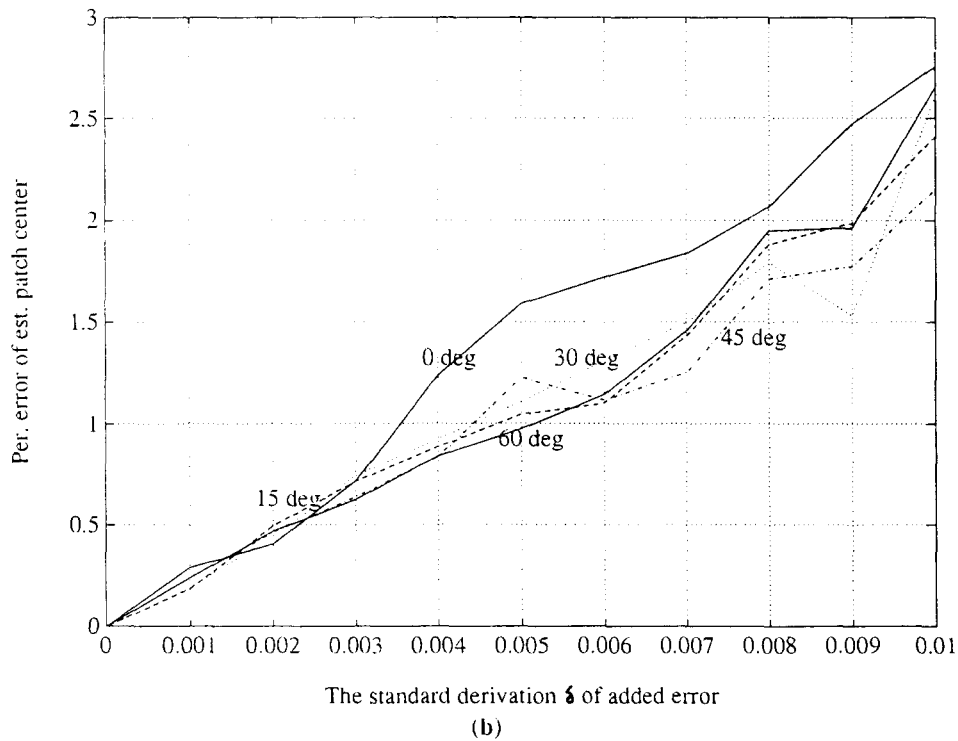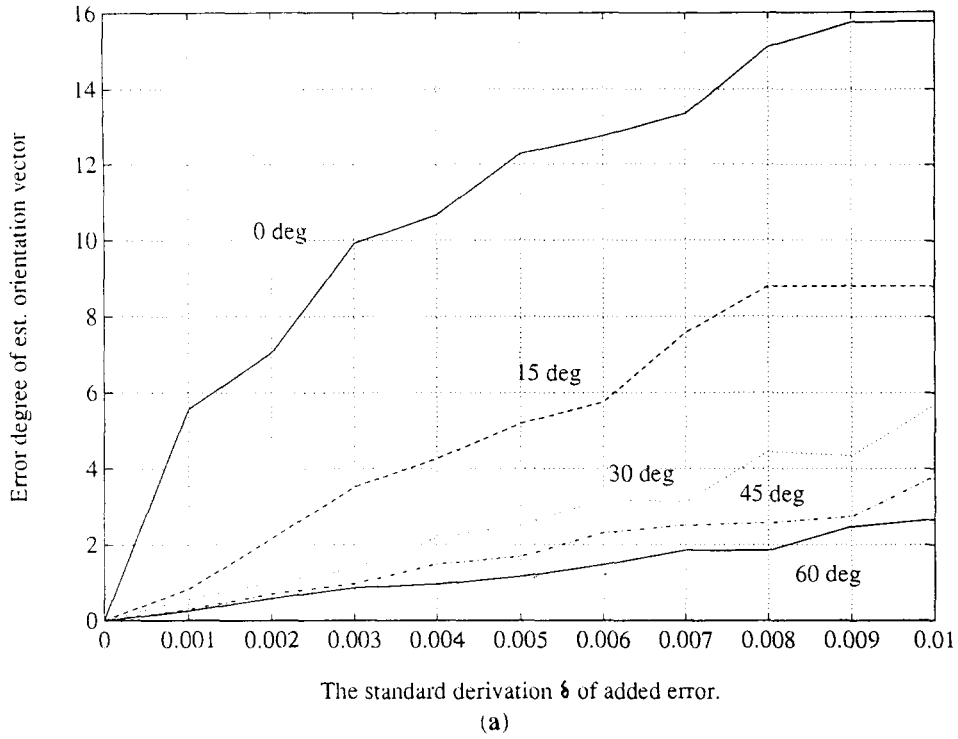The standard derivation δ of added error

(b)

Fig. 10. Error analysis of the five cases described in Fig. 9. (a) Degree errors of the estimated orientation vector. (b) Percentage errors of the estimated center position.

S.-C. Pei and L.-G. Liou

Table 2. Comparison between the real solution and our estimation results of problem (I). S_{obs} is generated by perspective projection for considering the modeling errors

Parameters	Real	Est. #1	Est. #2	Err. of #1
Rotation	(0.5000, 0.1000, −0.9000)	(0.5312, 0.0882, −0.8848)	(−1.0383, −0.1836, −0.8265)	3.55%
Center	(1.3304, 5.0789, 20.0000)	(1.3292, 4.8595, 19.7076)	(1.3292, 4.8595, 19.7076)	1.7682%
Orientation	(−0.1225, −0.4566, 0.8812)	(−0.1414, −0.4769, 0.8675)	(0.2352, 0.8196, 0.5225)	1.7699 deg

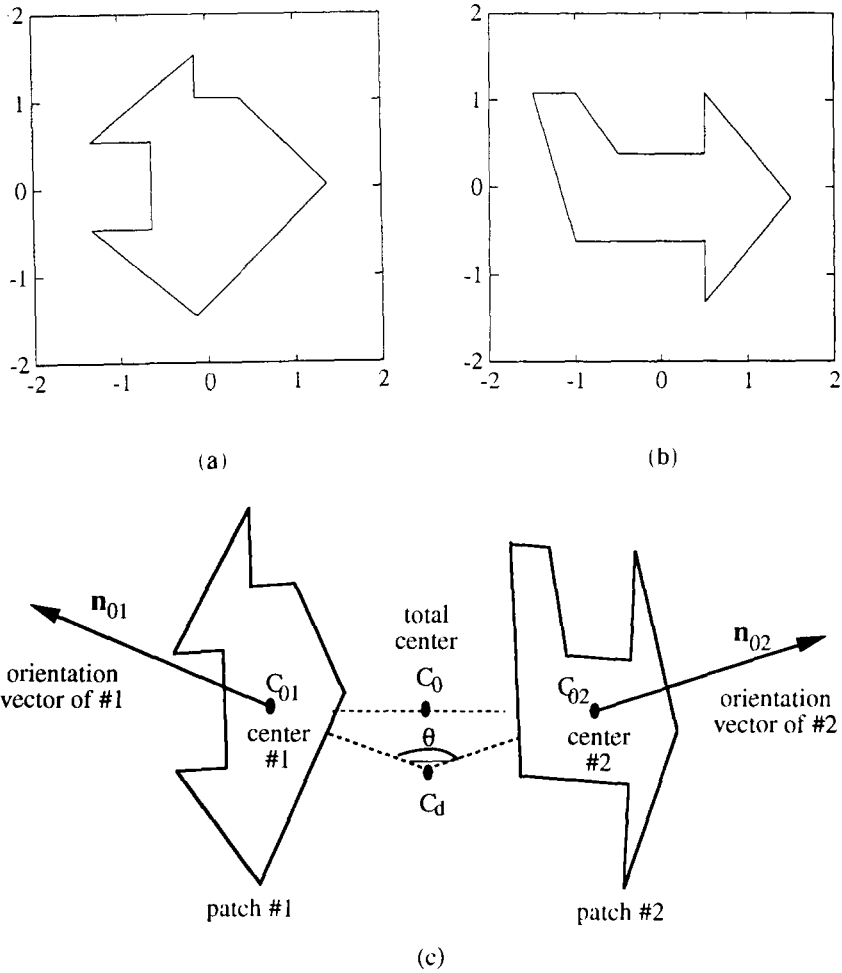(a)   (b)

(c)

Fig. 11. Construction of the observed two planar patches on a rigid object. (a) Actual size of #1 patch. (b) Actual size of #2 patch. (c) The structure of the test object. For simplicity, we assume that $n_{01}, n_{02}, C_{01}, C_{02}$, and C_0 are on the same plane. The angle θ is defined as the intersecting angle of these two patches.

Table 3. Comparison between the real solution and estimation results of problem (II). Two sets ($N = 2$) of corresponding planar patches are used. Scaled-orthographic projection is used as image model, and no observation error is added

Parameters	Real	Est. #1	Est. #2	Err. of #2
R	(0.4000, 0.2000, 0.2000)	(−0.4000, −0.2000, 0.2000)	(0.4000, 0.2000, 0.2000)	$2.7 \times 10^{-7}\%$
W	(0.8000, 0.6000, 0.6000)	(−0.8000, −0.6000, 0.6000)	(0.8000, 0.6000, 0.6000)	$2.5 \times 10^{-7}\%$
T	(−3.5868, 10.0605, 5.1132)	(5.6652, −4.5231, 4.8074)	(−3.5868, 10.0605, 5.1132)	$3.25 \times 10^{-7}\%$
V	(−11.4181, 12.7151, 14.1758)	(16.0602, −6.0890, 13.6579)	(−11.4181, 12.7151, 14.1758)	$3.0 \times 10^{-7}\%$
n_{01}	(0.7068, 0.0000, 0.7074)	(−0.7068, 0.0000, 0.7074)	(0.7068, 0.0000, 0.7074)	$3.5 \times 10^{-15}\%$
n_{02}	(−0.7068, 0.0000, 0.7074)	(0.7068, 0.0000, 0.7074)	(−0.7068, 0.0000, 0.7074)	$7.4 \times 10^{-15}\%$
C_0	(1.0000, 0.0000, 20.0000)	(1.0000, 0.0000, 20.0000)	(1.0000, 0.0000, 20.0000)	$1.2 \times 10^{-7}\%$
C_1	(2.0000, 3.0000, 23.0000)	(2.0000, 3.0000, 23.0000)	(2.0000, 3.0000, 23.0000)	$1.4 \times 10^{-7}\%$
C_2	(3.0000, 4.0000, 25.0000)	(3.0000, 4.0000, 25.0000)	(3.0000, 4.0000, 25.0000)	$2.0 \times 10^{-8}\%$

effect in computation. Besides, notice the change of signs in the two estimated rotation vectors and orientation vectors. From the listed estimation results, we can guarantee that our derivations are correct.

E2. All the planar patches and motion parameters are the same as that in E1. However, the image shape S_{obs} is now projected by perspective projection (the modeling errors are considered, see Fig. 8). After a rotation correction, a center-corrected shape S'_{obs} can be generated by the old shape S_{obs}. Notice the S'_{obs} is very near to the origin of the image plane. The best-approximated affine transformation between S_{ref} and S'_{obs} is utilized to determine the motion parameters. Numerical results listed in Table 2 are very close to the true answer (#1 solution).

E3. In this experiment, we try to *qualitatively* discuss the error sensitivity of the estimations (of problem (I)). There are many kinds of errors in the real situation: such
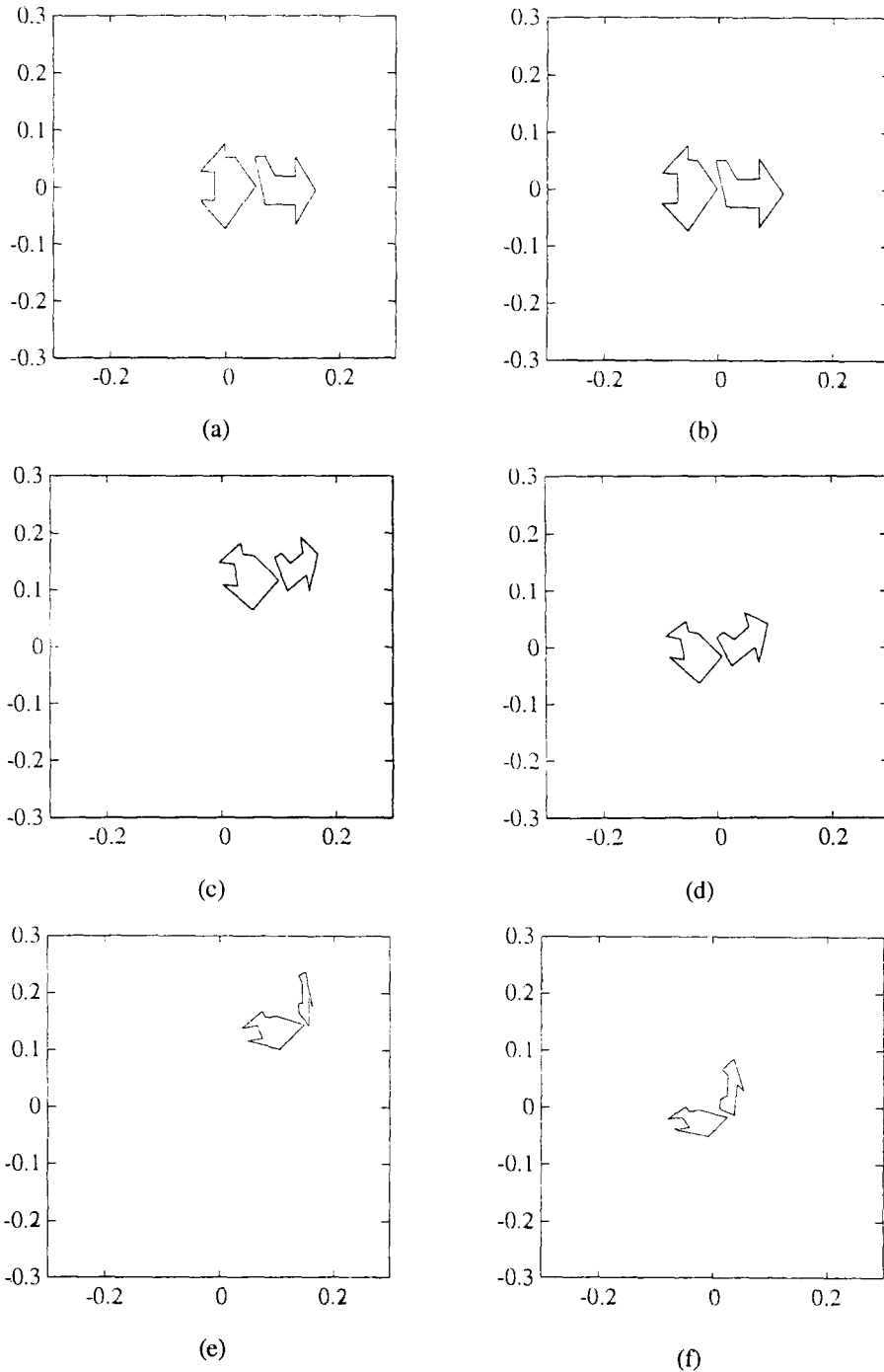

Fig. 12. The corresponding images of the case used in experiment E5. (a), (c), (e) are the projected image shapes at $t = t_0, t_1$, and t_2. (b), (d), (f) are the corresponding center-corrected image shapes of (a), (c), (e).

Table 4. Comparison between the real solution and estimation results of problem (II). Perspective projection is used as image model, and no observation error is added. Notice that the #2 solution is very close to the real solution

Parameters	Real	Est. #1	Est. #2	Err. of #2
R	$(0.4000, 0.2000, 0.2000)$	$(-0.6137, -0.0930, 0.2881)$	$(0.3911, 0.1863, 0.1980)$	3.36%
W	$(0.8000, 0.6000, 0.6000)$	$(-0.9570, -0.3682, 0.8346)$	$(0.7913, 0.5810, 0.5993)$	1.79%
T	$(-3.5868, 10.0605, 5.1132)$	$(4.4925, -8.3189, 7.0587)$	$(-3.2724, 9.9852, 5.3564)$	3.42%
V	$(-11.4181, 12.7151, 14.1758)$	$(14.7289, -7.9957, 14.9665)$	$(-11.0474, 12.9024, 14.7281)$	3.11%
n_{01}	$(0.7068, 0.0000, 0.7074)$	$(0.6306, 0.0279, -0.7756)$	$(0.7088, 0.0273, 0.7049)$	$1.57 \deg$
n_{02}	$(-0.7068, 0.0000, 0.7074)$	$(-0.7725, 0.0053, -0.6350)$	$(-0.7014, 0.0047, 0.7128)$	$0.52 \deg$
C_0	$(1.0000, 0.0000, 20.0000)$	$(1.0052, -0.0074, 19.9689)$	$(1.0068, -0.0075, 19.9996)$	0.05%
C_1	$(2.0000, 3.0000, 23.0000)$	$(2.0428, 3.0345, 23.3335)$	$(2.0459, 3.0391, 23.3694)$	1.61%
C_2	$(3.0000, 4.0000, 25.0000)$	$(3.1073, 4.1328, 25.8111)$	$(3.1121, 4.1391, 25.8508)$	3.41%

as modeling errors, quantization errors, and position errors in extracting the image shapes,..., etc. The final errors of the affine transformation are in fact a total-collect of these factors. Hence, we directly add some error quantities on the affine transformation to test the sensitivity of our algorithm. Notice that we use scaled-orthographic projection here for neglecting the modeling errors.

In this experiment, five cases with different orientation vectors of the same planar patches used in experiment E1 are analyzed (see Fig. 9). These test cases are generated by rotating the P_{ref} around the axis $(1, 0, 0)^T$ with five different θ angles $(0°, 15°, 30°, 45°,$ and $60°)$ and, without loss of generality, placing all their centers at point $(0, 0, 20)$. Orientation errors (deg) and percentage center position errors $(\%)$ are our main discussion topics.

The errors added on the affine matrix A and translation vector b are created as follows:

$$\tilde{A} = A + \begin{bmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{bmatrix}; \quad \tilde{b} = b + \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} \quad (49)$$

where the δ_{ij} and σ_i are random variables of normal distribution with zero means and standard deviations δ and σ.

Because the image shape S_{obs} has been placed at the image center previously, we do not need to discuss the influence from σ. Hence, σ is set to zero in all five test cases. The results are shown in Fig. 10. Every point on them is iterated 100 times, then averaged.

We can clearly observe that the same error level will induce remarkably different performance in orientation vector estimation, but only a very slight difference in center position estimation. It implies that the estimation result of a less slanted patch is more sensitive than that of a more slanted one in solving their orientation vectors.

E4. This experiment is very similar to experiment E1. It is designed for solving problem (II). Scaled-orthographic projection is used as the projection model for not considering the modeling errors. The object shown in Fig. 11 contains two planar patches, P_{01} and P_{02} (at time t_0). The center points and orientation vectors of the two planar patches are separately denoted by C_{01}, C_{02}, n_{01}, and n_{02}. These center points and orien-

tation vectors are lying on the same plane. So we can define an intersecting angle $\theta = \cos^{-1}(|n_{01} \cdot n_{02}|)$ as shown in Fig. 11(c).

Numerical results shown in Table 3 prove our derivations. Notice the estimated translation T and center positions are properly scaled for an easier comparison.

E5. Of course, this experiment, designed for problem (II), is also similar to experiment E2. The structure of the target patches and motion parameters are the same as that in experiment E4. However, the projected shapes on the image plane are obtained by perspective projection (the modeling errors are considered). The process of rotation correction is also executed as that in experiment E2 for a better approximation (see Fig. 12).

The estimation results shown in Table 4 are quite close to the true solutions.

E6. Analyzing the error sensitivity of problem (II) is a very troublesome problem because there are too many parameters that should be specified and tested. For example, different structures, motion parameters, positions will affect the error sensitivity. There seems to be not enough room to list all these results in our paper. Therefore, only a very simple case is tested. Even so, it still provides some important properties in error sensitivity. Notice here that we use scaled-orthographic projection for not considering the modeling errors.

The object defined in experiment E4 is used here. The distance $\overline{C_{01} C_{02}}$ is set to zero. Then we set six different intersecting angles θ to create six test structures in this experiment $(15°, 30°, 45°, 60°, 75°,$ and $90°)$. Errors are added to the affine matrices in a similar method defined in experiment E3. σ is still set to zero because all the centers of these two patches are placed on the Z-axis.

Without loss of generality, all of the average centers C_i of the two patches P_{i1} and P_{i2} at time t_i, $(i = 0-2)$ are placed on the Z-axis. These centers are $(0, 0, 18)$, $(0, 0, 21)$, and $(0, 0, 24)$. The corresponding rotation vectors to R and W are $0.3 \times (1/\sqrt{2}, 0, 1/\sqrt{2})$ and $0.7 \times (1/\sqrt{2}, 0, 1/\sqrt{2})$.

Rotation vector errors $(\%)$ and center position errors $(\%)$ are our main discussion topics. Results are shown in Fig. 13. It is easy to find that the case of smaller intersecting angle θ is more error sensitive. However, there is no remarkable difference among the errors of the estimated center positions for all six cases.
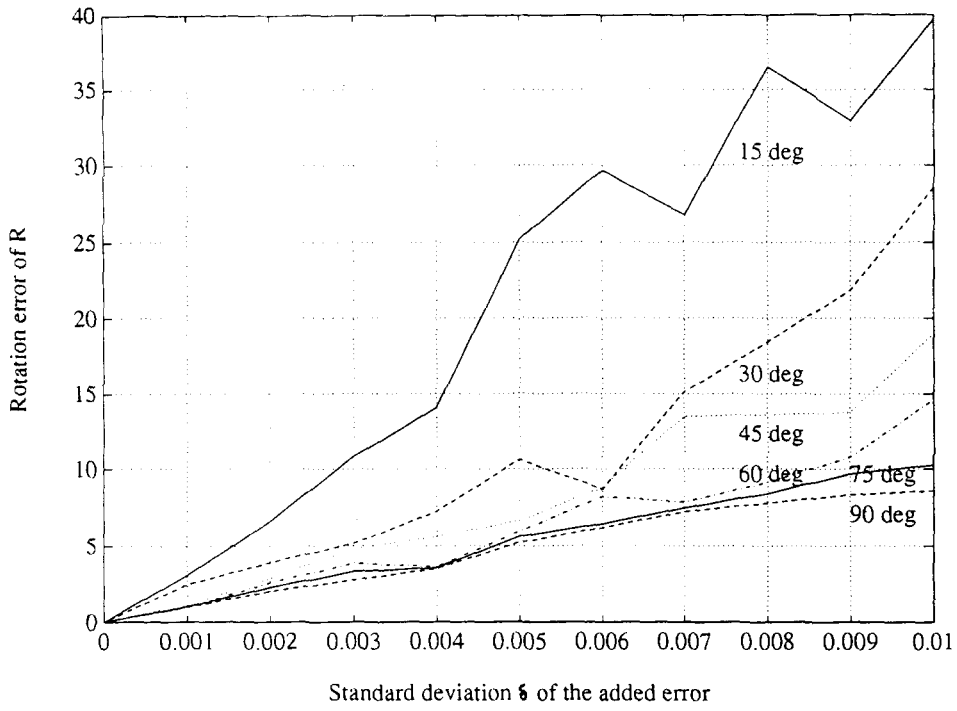
(a)
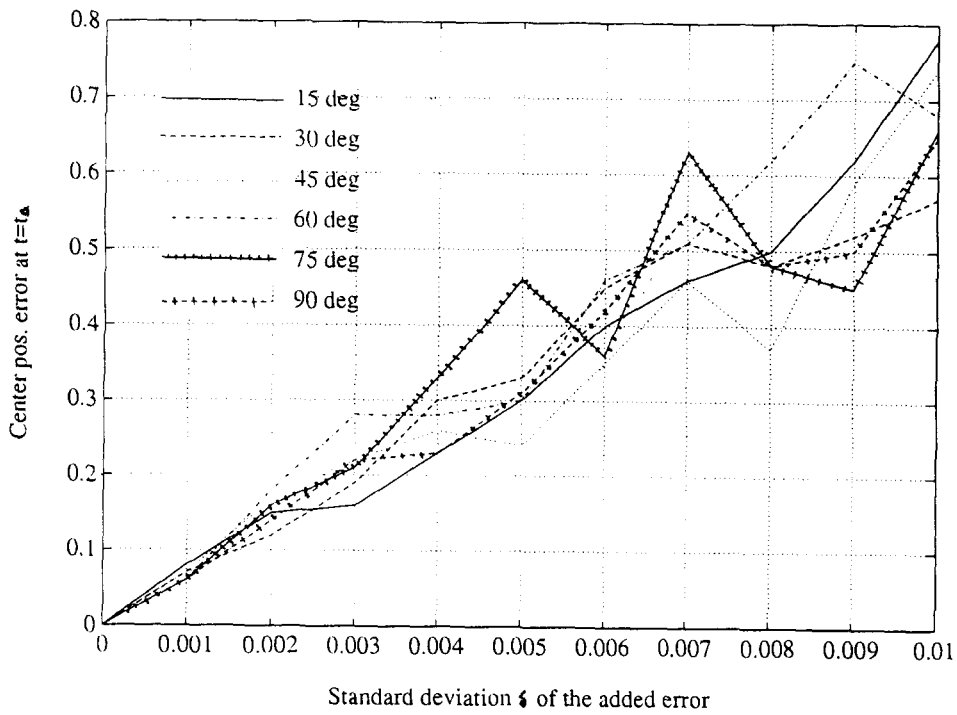
(b)

Fig. 13. Error analysis of the six cases described in experiment E6. (a) Percentage error of the estimated rotation vector corresponding to **R**. (b) Percentage error of the estimated center positon at time t_0.

6. CONCLUSIONS

In this paper, we derive our algorithm by using scaled-orthographic projection, instead of perspective projection. Once again, readers should realize that scaled-orthographic projection is close to perspective projection when the size of the target is relatively small (compared to the depth of this object) and placed near to the Z-axis. If the target does not satisfy these constraints, the estimations will be poor.

At least four advantages can be obtained by using scaled-orthographic projection:

(1) After making a rotation correction to the original image data, the scaled-orthographic projection model will be as good as a better model, ortho-perspective projection.

(2) The complexity and difficulty of mathematical derivations of perspective projection is so greatly reduced in our algorithm that we can obtain a closed-form solution of 3D motion and plane equations.

(3) From the derivations in our paper, we know the limits when we try to recover a unique 3D motion of a far object by using perspective projection. There are two distinct solutions (the true solution and the reflection solution) which can generate almost the same image on the image plane.

(4) Solutions obtained from scaled-orthographic projection can be used as a good initial guess of non-linear searching or iteration in many algorithms which derive their equations from perspective projection.

Indeterminacy and ambiguities are carefully discussed in our paper. Some readers may think that these uncertain properties of scaled-orthographic projection cause some serious problems in motion estimation. However, we would rather believe that these uncertain properties introduced by using scaled-orthographic projection faithfully reflect the inherent weakness of motion recovery when the object is small and positioned far enough from the camera. Besides, some degenerate cases resulting from degenerate structure or motion are also discussed for providing more insights to the problem.

In future work, we try to derive another similar closed-form solution from other kinds of image features (such as image lines) under the scaled-orthographic projection.

REFERENCES

1. J. Aaloimonos, Perspective approximations, *Image Vision Comput.* **8**(3), 179–192 (1990).
2. K. Kanatani, Structure and motion from optical flow under orthographic projection, *Comput. Vision Graphics Image Process.* **35**, 181–199 (1986).
3. X. Hu and N. Ahuja, Motion estimation under orthographic projection, *IEEE Trans. Robotics Automn* **7**(6), 848–853 (1991).
4. T. S. Huang and C. H. Lee, Motion and structure from orthographic projections, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-11**(5), 536–540 (1989).
5. K. Sugihara and N. Sugie, Recovery of rigid structure from orthographically projected optical flow, *Comput.*

Vision Graphics Image Process. **27**, 309–320 (1984).
6. H. H. Chen and T. S. Huang, Using motion from orthographic projections to verify 3-D point matches, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-13**(9), 872–878 (1991).
7. H. H. Chen, Determining motion and depth from binocular orthographic views, *Comput. Vision Graphics Image Process: Image Understanding* **54**, 47–55 (1991).
8. C. H. Lee and T. S. Huang, Finding point correspondences and determining motion of a rigid object from two weak perspective views, *Comput. Vision Graphics Image Process.* **52**, 309–327 (1990).
9. T. S. Huang, Motion and structure from two orthographic projections, Technical Report, Coordinated Science Lab, University of Illinois (1986).
10. H. C. Longuet-Higgins, A method of obtaining the relative positions of four points from three perspective projections, *Image Vision Comput.* **10**(5), 266–270 (1992).
11. B. M. Bennett *et al.*, Structure from two orthographic views of rigid motion, *J. Opt. Soc. Am. A* **6**(7), 1052–1069 (1989).
12. J. Aloimonos and C. K. Brown, On the kinetic depth effect, *Biol. Cybernetics* **60**, 445–455 (1989).
13. J. J. Koenderink and A. J. van Doorn, Affine structure from motion, *J. Opt. Soc. Am. A* **8**(2), 377–385 (1991).
14. R. Y. Tsai, T. S. Huang and W.-L. Zhu, Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition, *IEEE Trans. Acoustics Speech Signal Process.* **ASSP-30**(4), 525–534 (1982).
15. R. Y. Tsai and T. S. Huang, Estimating three-dimensional motion parameters of a rigid planar patch, III: finite point correspondences and three-view problem, *IEEE Trans. Acoustics Speech Signal Process.* **ASSP-32**(2), 213–220 (1984).
16. K. Kanatani, Tracing planar surface motion from a projection without knowing the correspondence, *Comput. Vision Graphics Image Process.* **35**, 1–12 (1985).
17. J. Weng, N. Ahuja and T. S. Huang, Motion and structure from point correspondences with error estimation: planar surface, *IEEE Trans. Signal Process.* **39**(12), 2691–2717 (1991).
18. R. Mukundan, Estimation of quaternion parameters from two dimensional image moments, *CVGIP: Graphical Models Image Process.* **54**(4), 345–350 (1992).
19. S. Negahdaripour, Closed-form relationship between the two interpretations of a moving plane, *J. Opt. Soc. Am. A* **7**(2), 279–285 (1990).
20. S. C. Pei and Lin-Gwo Liou, Tracking a planar patch in 3-D space by affine transformation in monocular and binocular vision, *Pattern Recognition* **26**, 23–31 (1993).
21. S. Chaudhuri and S. Chatterjee, Motion analysis of a homogeneously deformable object using subset correspondences, *Pattern Recognition* **24**, 739–745 (1991).
22. T. L. Faber and E. M. Stokely, Orientation of 3-D structures in medical image, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-10**(5), 626–633 (1988).
23. T. Y. Young and Y. L. Wang, Analysis of 3D rotation and linear changes, *Pattern Recognition Lett.* **2**, 239–242 (1984).
24. K. Wohn and J. Wu, Estimating the finite displacement using moments, *Pattern Recognition Lett.* **11**, 371–378 (1990).
25. R. Safaee-Rad, K. C. Smith, B. Benhabib and I. Tchoukanov, Application of moment and Fourier descriptors to accurate estimation of elliptical-shape parameters, *Pattern Recognition Lett.* **13**, 497–508 (1992).
26. Z. C. Li, Y. Y. Tang, T. D. Bui and C. Y. Suen, Shape transformation models and their applications in pattern recognition, *Int. J. Pattern Recognition Artif. Intell.* **4**(1), 65–94 (1990).
27. J. G. Leu, Shape normalization through compacting, *Pattern Recognition Lett.* **10**, 243–250 (1989).
28. Y. S. Abu-Mostafa and D. Psaltis, Image normalization by complex moments, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-7**(1), 46–55 (1985).

APPENDIX A. RELATIONSHIPS OF AN ORTHOGONAL MATRIX

Assume matrix $\mathbf{R}$ is an orthogonal matrix (especially used in the rigid object motion) can be divided by the form defined in equation (6). We try to find the relationships among $\mathbf{R}^*, \mathbf{r}_1$, $\mathbf{r}_2$, and r_3. Because $\mathbf{RR}^T = \mathbf{R}^T\mathbf{R} = \mathbf{I}_3$ (identity matrix), we have the following equations:

$$\mathbf{R}^*\mathbf{R}^{*T} + \mathbf{r}_1\mathbf{r}_1^T = \mathbf{I}_2; \quad \mathbf{R}^{*T}\mathbf{R}^* + \mathbf{r}_2\mathbf{r}_2^T = \mathbf{I}_2 \quad (A1)$$

$$\mathbf{R}^*\mathbf{r}_2 + r_3\mathbf{r}_1 = 0; \quad \mathbf{R}^{*T}\mathbf{r}_1 + r_3\mathbf{r}_2 = 0 \quad (A2)$$

$$\mathbf{r}_2^T\mathbf{r}_2 + r_3^2 = 1; \quad \mathbf{r}_1^T\mathbf{r}_1 + r_3^2 = 1 \quad (A3)$$

$$\det(\mathbf{R}) = 1. \quad (A4)$$

From equation (A3), we know $|\mathbf{r}_1| = |\mathbf{r}_2|, \mathbf{r}_1 = c\hat{\mathbf{r}}_1$ and $\mathbf{r}_2 = c\hat{\mathbf{r}}_2$. $\hat{\mathbf{r}}_1^\perp$ and $\hat{\mathbf{r}}_2^\perp$ are defined as that in equation (11). By premultiplying both sides of equation A2(a) by $\hat{\mathbf{r}}_1^\perp$ and rearranging equation A2(b), we have

$$(\hat{\mathbf{r}}_1^\perp)^T\mathbf{R}^* = a(\hat{\mathbf{r}}_2^\perp)^T \;;\; \hat{\mathbf{r}}_1^T\mathbf{R}^* = -r_3\hat{\mathbf{r}}_2^T \quad (A5)$$

where a is a constant. Therefore, the matrix $\mathbf{R}^*$ can be decomposed into

$$\mathbf{R}^* = [\hat{\mathbf{r}}_1 \hat{\mathbf{r}}_1^\perp] \begin{bmatrix} -r_3 & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} \hat{\mathbf{r}}_2^T \\ (\hat{\mathbf{r}}_2^\perp)^T \end{bmatrix} \quad (A6)$$

and $\det(\mathbf{R}^*) = -r_3 a$.

Pre- and post-multiply both sides of equation A1(a) by $(\hat{\mathbf{r}}_1^\perp)^T$ and $\hat{\mathbf{r}}_1^\perp$, we have $a^2 = 1$. Because the cross product of the first column vector and the second column vector of an orthogonal matrix $\mathbf{R}$ is equal to the third column vector, we know $\det(R^*) = r_3$. Thus $a = -1$.

By using a similar method, readers can derive another important property of an orthogonal matrix.

APPENDIX B. METHODS OF ACQUIRING THE AFFINE TRANSFORMATION

In order to keep our paper short and neat, we just show the main ideas here.

If there exists an unknown affine transformation between two corresponding shapes S_1 and S_2 on the image plane, how do we solve this unknown transformation from the two shapes S_1 and S_2? (see Fig. B1).

If the corresponding points on these two shapes are available, all we have to do is to solve a set of linear equations, that seems very easy. However, if these corresponding points
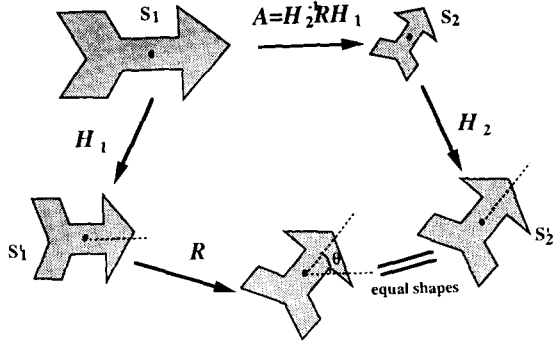

Fig. B1. The process of acquiring the affine matrix $\mathbf{A}$ by using shape normalization method.

are not available, it is not so easy as before. A lot of methods without were proposed using corresponding points.[20-26]

First, we separately normalize the two image shapes S_1 and S_2 into another two "normalized" shapes S_1' and S_2' (see Fig. B1).[27,28] The so-called "normalized" shape is a shape whose dispersion matrix (defined later) is an identity matrix. Dispersion matrix $\mathbf{M}$ is defined as

$$\mathbf{M} = \frac{1}{|S|} \int\int_S \mathbf{pp}^T \, dx \, dy \quad (B1)$$

where $|S|$ is the area of the image shape S; $\mathbf{p}$ is the coordinate vector relative to the shape center $\mathbf{c}$ of the shape S.

This symmetric matrix $\mathbf{M}$ can be decomposed into $\mathbf{Q}\Lambda\mathbf{Q}^T$ by using similarity transformation. If we make a coordinate transform to $\mathbf{p}$ such that $\mathbf{p}' = (\mathbf{Q}\Lambda^{-1/2})^T\mathbf{p} = \mathbf{Hp}$, then the new dispersion matrix $\mathbf{M}'$ of the normalized shape S' is just an identity one.

Therefore, from the above procedure, we have $\mathbf{H}_1, \mathbf{H}_2$, and the normalized shapes S_1' and S_2'. Now, because the normalized shapes S_1' and S_2' are only different by a rotation angle θ, it is easy to show that the affine matrix $\mathbf{A}$ can be written by $\mathbf{A} = \mathbf{H}_2^{-1}\mathbf{R}(\theta)\mathbf{H}_1$, and the translation vector $\mathbf{b}$ by $\mathbf{b} = \mathbf{c}_2 - \mathbf{Ac}_1$.

The rotation $\mathbf{R}(\theta)$ can be obtained by many methods. Because the two shapes S_1' and S_2' are only different by an angle θ, this problem must be easier than the general affine case. We do not wish to discuss these methods here.

Once the matrix $\mathbf{R}$ is acquired, the whole problem is completely solved. A detailed algorithm is described in *reference* (20).

About the Author—SOO-CHANG PEI was born in Soo-Auo, Taiwan, China on 20 February 1949. He received the B.S. degree from the National Taiwan University in 1970 and the M.S. and Ph.D. degrees from the University of California, Santa Barbara, in 1972 and 1975, respectively, all in electrical engineering. He was an engineering officer in the Chinese Navy Shipyard at Peng Fu Island from 1970 to 1971 and a Research Assistant at the University of California, Santa Barbara, from 1971 to 1975. He was Professor and Chairman in the Department of Electrical Engineering at Tatung Institute of Technology from 1981 to 1983. Presently, he is the Professor of the Department of Electrical Engineering at the National Taiwan University. His research interests include digital signal processing, digital picture processing, optical information processing, laser and holography. Dr Pei is a member of the IEEE, Eta Kappa Nu and the Optical Society of America.

About the Author—LIN-GWO LIOU was born in Taiwan. He received the B.S. degree from the National Chiao Tung University (N.C.T.U.) in Taiwan in 1989. Now, he is at the National Taiwan University (N.T.U.) studying for a Ph.D. degree. His research interests include motion image analysis, methods for 3D object reconstruction, pattern recognition in image application.