Pattern Recognition, Vol. 26, No. 1, pp. 23–31, 1993
Printed in Great Britain

0031-3203/93 $6.00 + .00
Pergamon Press Ltd

TRACKING A PLANAR PATCH IN THREE-DIMENSIONAL SPACE BY AFFINE TRANSFORMATION IN MONOCULAR AND BINOCULAR VISION

SOO-CHANG PEI and LIN-GWO LIOU
Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, Republic of China

(*Received* 15 *July* 1991; *received for publication* 1 *April* 1992)

Abstract—The three-dimensional (3D) motion of a planar surface (or patch) can only be detected from the motion of its projected 2D contour or region image on the plane of vision. Acquirement of corresponding points is the most important procedure. In the past, many methods of how to find corresponding points were published frequently. Affine transformation is used for the matching work. By this transformation, the position, orientation and motion of a planar patch can be calculated. Numerical calculation is very simple in this method because only the mass moments need to be calculated not iterative or exhausting searching.

Motion analysis Affine transformation Computer vision

1. INTRODUCTION

Detecting the three-dimensional (3D) motion of a planar patch from the two-dimensional (2D) motion of its projected image is one of the most important tasks in computer vision and image processing. One of the prevailing approaches is first to detect the corresponding points and then to analyze their optical flow vectors.[1–3] However, it is well-known that the correspondence problem is one of the most difficult problems in image processing; not only because of the discouraging, boring, iterative searching but also because its highly noise-sensitive and ambiguous characteristics make us feel that a new method to detect 3D motion without searching corresponding points should be developed.

From reference (4) an idea was obtained about how mass moments were utilized in pattern recognition. Then, a new thought came to mind: "Why not use affine transformation to get motion parameters?"[5] So, the method described in reference (4) was modified. In later articles, the way of linking affine transformation with motion parameters will be shown. Before that, details on how the problem of tracking planar patch is formulated, will be described.

When the surface normal vector of a planar patch is not parallel to the viewing axis of the viewer (Z-axis), the perceived shape on the image plane is skewed. However, skewing is not the only case that can happen to the projected shape when the patch moves in 3D space. Translation, rotation and scaling should also be considered. After inspecting the projected shape R and R', which correspond to the planar patch before and after motion, that there *approximately* exists an affine transformation relationship between the two shapes, R and R' is assumed, as shown in Fig. 1. In fact, this relationship is just a kind of point correspondence.

Because the affine coefficients are easily derived by the mass moments of the shape R and R' (shown later), iterative searching is no longer required. In addition to easy calculation, mass moments have another excellent characteristic: noise-insensitivity.

Finally, some details should be especially noticed here.

(1) The reason why the moving rigid body in 3D space was restricted to being planar or approximately planar was so that the occlusion problem did not occur. Only those shapes projected from a planar patch can be utilized to do the matching work by affine transformation.

(2) Perspective projection is adopted to model the camera.

(3) In this paper, the use of one camera (monocular vision) or two cameras (binocular vision) to track a planar patch in 3D space is attempted.

(4) For simplicity, the projected image should be thresholded first to get a bilevel image. Level white for the projected shape and black for the background.

(5) Assume that it is known that the projected shapes come from the same planar patch in 3D space (i.e. recognition has been done previously).

2. USING MASS MOMENTS TO CALCULATE AFFINE TRANSFORMATION COEFFICIENTS

As shown in Fig. 1, there are two shapes R and R' on the image plane. Both of them can be considered as different projected shapes of the same object planar patch at different positions and orientations.

Now, a pixel point $\mathbf{p} = (x, y)$ in region R can map to another pixel point $\mathbf{p}' = (x', y')$ in region R' via the

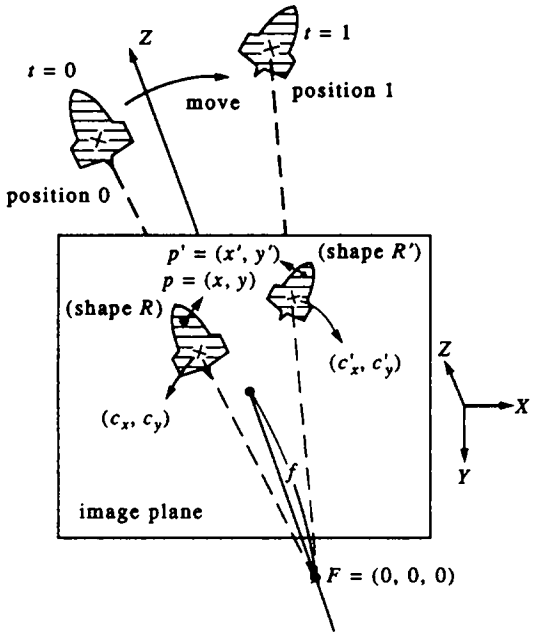


Fig. 1. Geometrical structure of the projected shapes and the object planar patch at different positions and orientation.

following affine transformation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + T. \quad (1)$$

Because the center point $\mathbf{c} = (c_x, c_y)$ of region R must map to the center point $\mathbf{c}' = (c'_x, c'_y)$ of region R', equation (1) can be stated in a more convenient form

$$\begin{bmatrix} x' - c'_x \\ y' - c'_y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x - c_x \\ y - c_y \end{bmatrix} = A \begin{bmatrix} x - c_x \\ y - c_y \end{bmatrix}. \quad (2)$$

Our purpose here is to acquire the four affine coefficients a, b, c, d.

First, we define mass moments m_{ij} and m'_{ij}

$$m_{ij} = \frac{1}{S^{i+j}} \iint_R (x - c_x)^i (y - c_y)^j \, dx \, dy \quad (3)$$

$$m'_{ij} = \frac{1}{S^{i+j}} \iint_{R'} (x' - c'_x)^i (y' - c'_y)^j \, dx' \, dy' \quad (4)$$

where (order $= i + j$)

$$m_{00} = \iint_R dx \, dy = S, \quad m'_{00} = \iint_{R'} dx' \, dy' = S'$$

$$c_x = \iint_R x \, dx \, dy, \qquad c'_x = \iint_{R'} x' \, dx' \, dy'$$

$$c_y = \iint_R y \, dx \, dy, \qquad c'_y = \iint_{R'} y' \, dx' \, dy'$$

$$S = \det(A) S'. \quad (5)$$

According to equations (3) and (4), the mass moments of any order can be obtained as required. Although integration operation ($\int$) is used in equations (3) and (4), it should be replaced by summation operation ($\sum$) in actual processing.

Here, we define the dispersion matrix M and M' just as in reference (4)

$$M = \begin{bmatrix} m_{20} & m_{11} \\ m_{11} & m_{02} \end{bmatrix} \quad (6)$$

$$M' = \begin{bmatrix} m'_{20} & m'_{11} \\ m'_{11} & m'_{02} \end{bmatrix}. \quad (7)$$

Let $\lambda_1, \lambda_2, \mathbf{E}_1, \mathbf{E}_2$ be the eigenvalues and the corresponding eigenvectors of matrix M. In the same way, we also define $\lambda'_1, \lambda'_2, \mathbf{E}'_1, \mathbf{E}'_2$ of the matrix M'. Notice that $\|\mathbf{E}_1\| = \|\mathbf{E}_2\| = \|\mathbf{E}'_1\| = \|\mathbf{E}'_2\| = 1$.

Let us define matrix $E = [\mathbf{E}_1 | \mathbf{E}_2]$ and $E' = [\mathbf{E}'_1 | \mathbf{E}'_2]$. Because M and M' are symmetric, we must have $\mathbf{E}_1 \perp \mathbf{E}_2$ and $\mathbf{E}'_1 \perp \mathbf{E}'_2$ and $\lambda_1, \lambda_2, \lambda'_1, \lambda'_2 > 0$ (positive real solutions).

From similarity transformation, the matrix M can be decomposed into

$$M = E \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} E^{\mathrm{T}} = Z I Z^{\mathrm{T}}. \quad (8)$$

In the same way, the matrix M' can also be decomposed into

$$M' = E' \begin{bmatrix} \lambda'_1 & 0 \\ 0 & \lambda'_2 \end{bmatrix} E'^{\mathrm{T}} = Z' I Z'^{\mathrm{T}}. \quad (9)$$

Now, in shape R, if we change the coordinate system by $\mathbf{P}_1$ and $\mathbf{P}_2$ as the two basic vectors and (c_x, c_y) as

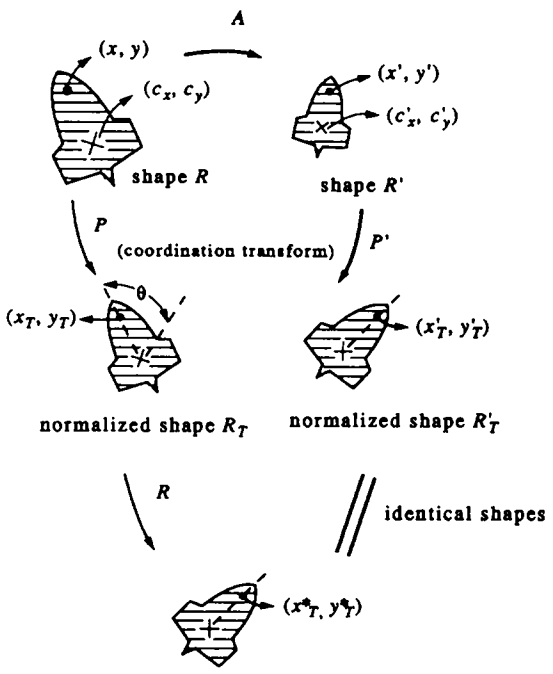


Fig. 2. The process of acquiring affine transformation between R and R' from normalizing shapes R_T, R^*_T and R'_T.

the center, the original point coordinate (x, y) can be represented by a new point coordinate (x_T, y_T), and we have a new transformed shape R_T (as shown in Fig. 2). Similarly, we have another new shape R'_T from R' via $\mathbf{P}'_1, \mathbf{P}'_2$ and (c'_x, c'_y).

$$\begin{bmatrix} x - c_x \\ y - c_y \end{bmatrix} = [\mathbf{P}_1 | \mathbf{P}_2] \begin{bmatrix} x_T \\ y_T \end{bmatrix} = P \begin{bmatrix} x_T \\ y_T \end{bmatrix}. \tag{10}$$

And we define the moment m_{Tij} and the dispersion matrix M_T as follows:

$$m_{Tij} = \frac{1}{S_T^{i+j}} \int\int_{R_T} (x_T)^i (y_T)^j \, dx_T \, dy_T \tag{11}$$

$$M_T = \begin{bmatrix} m_{T20} & m_{T11} \\ m_{T11} & m_{T22} \end{bmatrix}. \tag{12}$$

The dispersion matrix M and M_T will have the following relationship (similar to M' and M'_T):

$$M = P M_T P^{\mathrm{T}}. \tag{13}$$

Compared with equation (8), we can choose $P = Z$ such that the dispersion matrix M_T of the shape R_T is equal to the identity matrix I. Similarly, the shape R' is transformed to the shape R'_T by P', which is forced to be equal to Z' in order to make $M'_T = I$, too. These transformed shapes (like R_T and R'_T) in Fig. 2 whose dispersion matrices are equal to I are here called *normalized shapes*.

In spite of the same dispersion matrix I, it does not mean that the shapes R_T and R'_T are really identical. In fact, R_T differs from R'_T just by a rotation angle θ. If we rotate the shape R_T by this angle θ and transform the shape R_T to R_T^* (as shown in Fig. 2), then R'_T and R_T^* will be identical. Thus we have

$$\begin{bmatrix} x_T \\ y_T \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_T^* \\ y_T^* \end{bmatrix} = R(\theta) \begin{bmatrix} x_T^* \\ y_T^* \end{bmatrix} \tag{14}$$

$$\begin{bmatrix} x_T^* \\ y_T^* \end{bmatrix} = \begin{bmatrix} x'_T \\ y'_T \end{bmatrix}. \tag{15}$$

From equations (10), (14) and (15), we have

$$\begin{bmatrix} x' - c'_x \\ y' - c'_y \end{bmatrix} = P' R^{\mathrm{T}} P^{-1} \begin{bmatrix} x - c_x \\ y - c_y \end{bmatrix}. \tag{16}$$

Comparing with equation (2) gives

$$A = P' R^{\mathrm{T}} P^{-1}. \tag{17}$$

If θ is known, we can easily calculate all the affine coefficients. So here comes a question: "How can we know the value of θ?"

Obviously, the Fourier transformation (FT) and the complex log mapping (CLM) are the two appropriate choices to obtain θ. In the following, we will provide two other methods to obtain θ without a complex value operation and matching process as FT or CLM did before.

(1) The first method is very simple. Let us look at the two normalized shapes R_T and R'_T in Fig. 3. There *often exists* a point D_T (or D'_T) on the outer edge of

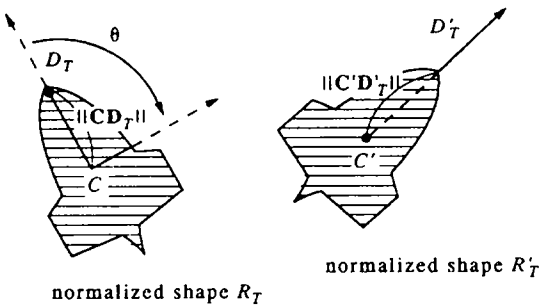


Fig. 3. Finding the difference angle θ between $\mathbf{CD}_T$ and $\mathbf{C'D'}_T$.

R_T (or R'_T) such that the distance $\|\mathbf{CD}_T\|$ (or $\|\mathbf{C'D'}_T\|$) is maximized. Then we just need to calculate the difference angle θ between $\mathbf{CD}_T$ and $\mathbf{C'D'}_T$.

Careful readers may have found that D_T (or D'_T) does not always exist uniquely. So we especially point out the two words: *often exist*.

(2) The second method is a little more complex than the first. Higher order moments (order = 3) are used to solve the rotation angle θ. Because all the mass moments m_{ij} and m'_{ij} whose orders are equal to or less than 2 will not suffice to fully determine the four affine coefficients (in equations (2) and (14)) between the two shapes, R and R', there is still one degree of freedom (that is θ) to be solved. Here we utilize higher order moments to solve the last unknown θ. Naturally, third-order moments are the best candidates. From equations (2)–(4) we have

$$m'_{30} = \frac{1}{(\det(A))^2} [a^3 m_{30} + 3a^2 b m_{21} + 3ab^2 m_{12} + b^3 m_{03}]. \tag{18}$$

From equation (17), the coefficients a, b, c, d are all functions of θ and then we have to solve equation (18) to get θ. Of course, we may have more solutions than only one. However, if the object planar patch is not so symmetric (e.g. rectangle, square, circle, etc.), most of the time we can get a unique solution. Even multiple solutions do exist, it is natural to choose the solution whose matching ratio is highest to solve the ambiguity.

Because digitized summation $\sum$ is used instead of continuous integration $\int$, values of mass moments cannot avoid quantization error, which makes us feel that equation (18) is not as precise as we derive. But later experiments will show that the error of the desired affine coefficients from solved θ in equation (18) is quite small.

Just as we have discussed previously there are many methods to solve θ. Any one of them will suit our purpose.

3. TRACKING A PLANAR PATCH BY TWO CAMERAS

Information about orientation and position of a planar patch in 3D space is the key to track this object patch. The goal here is to get this information im-

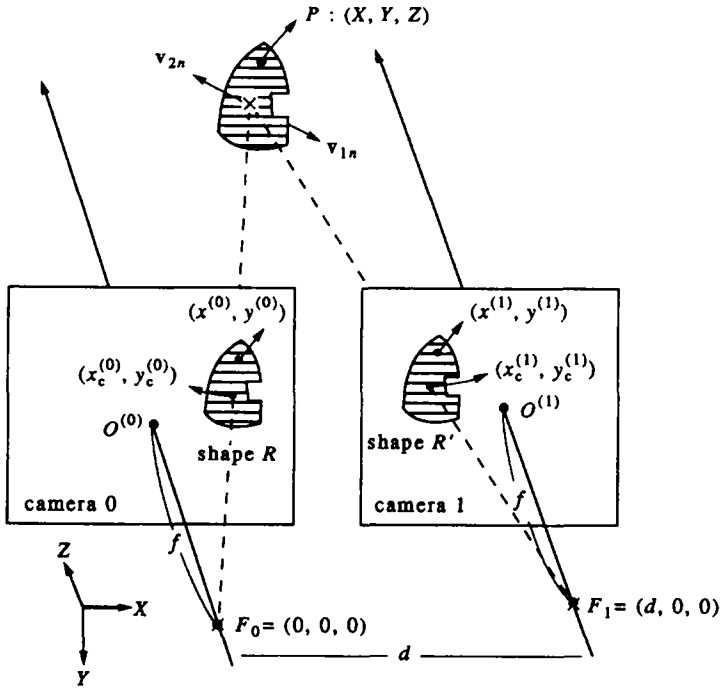


Fig. 4. Geometrical structure of binocular vision.

mediately when the projected shapes of the object planar patch are seen on each image plane of the two cameras (see Fig. 4). There were many methods which attempted to solve this problem.[2,3] However, these methods look complex.

Affine transformation still plays a main role in our method. Surprisingly, when two cameras are used for tracking, acquiring the four affine coefficients becomes much easier than using just one camera. When using two cameras, there is no necessity to calculate higher order moments.

The two cameras, named camera 0 and 1, are placed as shown in Fig. 4. Both of them have the same focal length f, and their focal points F_0 and F_1 are put at $(0,0,0)$ and $(d,0,0)$.

Under the rules of perspective projection, a point $P=(X,Y,Z)$ in 3D space can project two image points $(x^{(0)},y^{(0)})$ and $(x^{(1)},y^{(1)})$ (as shown in Fig. 4). We can write the following equations:

$$(x^{(0)},y^{(0)})=\left(\frac{f}{Z}\right)[X,Y] \tag{19}$$

$$(x^{(1)},y^{(1)})=\left(\frac{f}{Z}\right)[X-d,Y]. \tag{20}$$

It is easy to find that $y^{(0)}=y^{(1)}$ for the object point P in 3D space, but $x^{(0)}\neq x^{(1)}$ if $d\neq 0$. It is easy to find that

$$Z=\frac{fd}{x^{(0)}-x^{(1)}} \tag{21}$$

$$(X,Y)=\frac{Z}{f}[x^{(0)},y^{(0)}]. \tag{22}$$

From equations (21) and (22), we can completely recover the 3D coordinate (X,Y,Z) of the point P.

Now, if this point P is any one point on the object planar patch which projects two shapes R and R' on two image planes, this point P must project two image points $(x^{(0)},y^{(0)}),(x^{(1)},y^{(1)})$. Suppose there approximately exist an affine transformation between the two shapes, R and R', then we will have

$$\begin{bmatrix}x^{(1)}-x_c^{(1)}\\ y^{(1)}-y_c^{(1)}\end{bmatrix}=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}x^{(0)}-x_c^{(0)}\\ y^{(0)}-y_c^{(0)}\end{bmatrix} \tag{23}$$

or

$$\begin{bmatrix}x^{(1)}\\ y^{(1)}\end{bmatrix}=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}x^{(0)}\\ y^{(0)}\end{bmatrix}+\begin{bmatrix}-ax_c^{(0)}-by_c^{(0)}+x_c^{(1)}\\ -cx_c^{(0)}-dy_c^{(0)}+y_c^{(1)}\end{bmatrix} \tag{24}$$

where $(x_c^{(0)},y_c^{(0)})=\mathbf{c}^{(0)},(x_c^{(1)},y_c^{(1)})=\mathbf{c}^{(1)}$ are the centers of the shape R and R'.

Because of the special structure of these two cameras, we have $y^{(0)}=y^{(1)}$ for all points in shapes R and R'. Then equation (24) should be modified to

$$\begin{bmatrix}x^{(1)}\\ y^{(1)}\end{bmatrix}=\begin{bmatrix}a & b\\ 0 & 1\end{bmatrix}\begin{bmatrix}x^{(0)}\\ y^{(0)}\end{bmatrix}+\begin{bmatrix}b_x\\ 0\end{bmatrix}. \tag{25}$$

From equations (21) and (25), we have

$$Z=\frac{fd}{x^{(0)}-x^{(1)}}=\frac{fd}{(1-a)x^{(0)}-by^{(0)}-b_x}. \tag{26}$$

Once we know $a,b,b_x,x^{(0)},y^{(0)}$, any one point $P=(X,Y,Z)$ on the planar patch can be obtained from equations (22) and (26). Sometimes, we hope that the normal vector of the object planar patch can also be

obtained. Substituting equation (19) into equation (26)

$$(1-a)X-(b)Y-\left(\frac{b_x}{f}\right)Z=d \tag{27}$$

we found that equation (27) is just a representation of a plane in 3D space and its normal vectors are

$$\mathbf{v}_n=\pm\frac{\left(1-a,-b,-\frac{b_x}{f}\right)}{\sqrt{\left((1-a)^2+b^2+\left(\frac{-b_x}{f}\right)^2\right)}}. \tag{28}$$

Here we choose the sign that can make $v_{nZ}<0$.

Next, how can we obtain the coefficients a, b in equation (25)? Of course, the method described in Section 2 can be used. But we also notice that the four affine coefficients in equation (23) are not completely unknown ($c=0, d=1$). We cannot help suspecting that there should be a shortcut in solving the last two unknowns a and b. From equations (4) and (23) with $c=0$, $d=1$, we have the following results:

$$m'_{11}=\frac{b}{a}m_{02}+m_{11} \tag{29}$$

$$m'_{20}=am_{20}+2bm_{11}+\frac{b^2}{a}m_{02} \tag{30}$$

$$m'_{02}=\frac{1}{a}m_{02}. \tag{31}$$

We can solve a and b from equations (29) and (31) and get

$$a=\frac{m_{02}}{m'_{02}} \tag{32}$$

$$b=\frac{m'_{11}-m_{11}}{m'_{02}}. \tag{33}$$

Once we get a and b from measured moments m_{ij} and m'_{ij}, the normal vector of object planar patch can be recovered from equation (28).

4. TRACKING A PLANAR PATCH BY ONE CAMERA

Many similar researchers[1-3] have presented such a problem in the past, monocular vision was adopted by them. Basically, monocular tracking is quite different from binocular tracking because there is not enough information from just one camera. For this reason, some *prior information* or *reference* should be provided first. In the following paragraph, a new method based on affine transformation will be presented.

First, for the object planar patch, we should know its prior information. The so-called prior information is only one projected shape R_{ref} of the objected planar patch with *known* position and orientation (as shown in Fig. 6). Now, if the patch moves to another place with new position and orientation and then projects a new shape R_{obs} on the same image plane, we can acquire the desired information of the moved planar patch (such as normal vector $\mathbf{v}_n$) by comparing the two shapes R_{ref} and R_{obs}.

Just as mentioned before, there approximately exists an affine transformation relationship between the two shapes R_{ref} and R_{obs}. However, the shortcut described in Section 3 cannot be used here because no special relationship should exist between these two shapes. It means that the affine coefficients here can only be determined by the general method described in Section 2.

As known before, a planar patch is completely determined if the coordinates of at least three points on it are known (not on the same line). Referring to

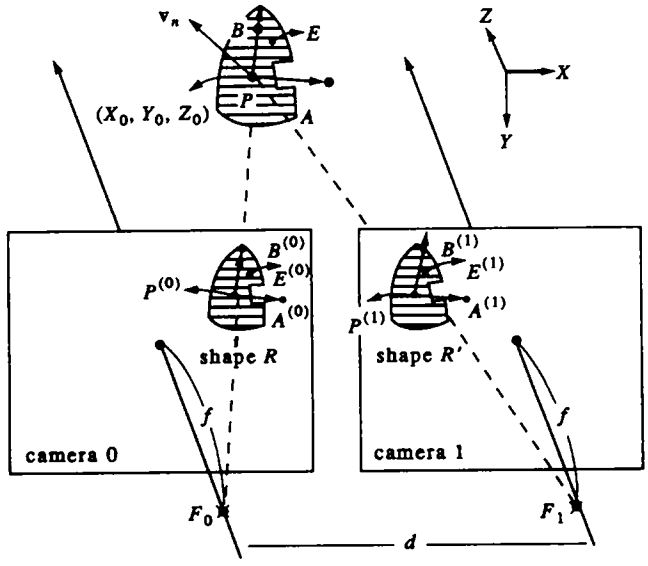


Fig. 5. Illustration of the structure in analyzing the affine transformation between the shapes R and R'.

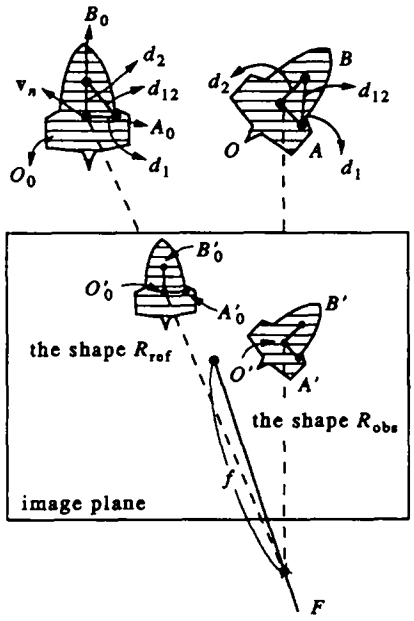


Fig. 6. Geometrical configuration in monocular vision.

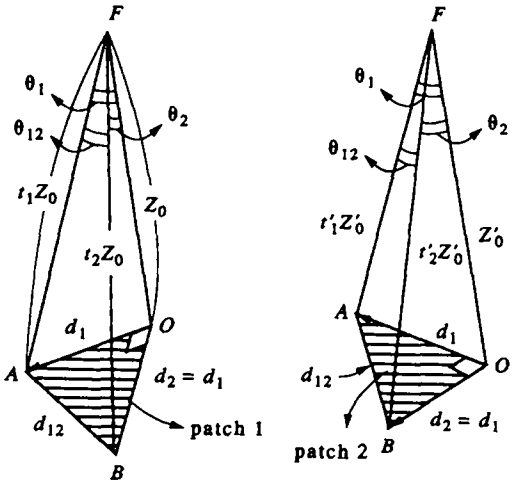


Fig. 7. Ambiguous solution patches (1 and 2) from different sets of (t_1, t_2) with the same set of $(\theta_1, \theta_2, \theta_{12})$ (see Fig. 6).

Fig. 6 (or Fig. 7), we define the following parameters:

$$\|\mathbf{FO}\| = Z_0, \|\mathbf{FA}\| = Z_1 = t_1 Z_0, \|\mathbf{FB}\| = Z_2 = t_2 Z_0$$
$$\angle AFO = \theta_1, \angle BFO = \theta_2, \angle AFB = \theta_{12}$$
$$\|\mathbf{OA}\| = d_1, \|\mathbf{OB}\| = d_2, \|\mathbf{AB}\| = d_{12}$$

where O is the center of the patch.

The points A, B, O are three points on the object planar patch and each of them projects a point (A', B', O') on the image plane. It is easy to find that the object planar patch can be completely determined once the parameters t_1, t_2, Z_0 are acquired. The question is: "How can we get them?"

Assume that the projected points A', B', O' can be acquired by the known affine relationship between R_{ref} and R_{obs} (A, B, O corresponds to A_0, B_0, C_0 which were previously chosen in the reference patch, see Fig. 6). We also assume that the distance (d_1, d_2, d_{12}) is also known from prior information. From the three triangles $\triangle FOA$, $\triangle FOB$, $\triangle FAB$ (see Fig. 7), and applying the cosine theorem, we have

$$(k)^2 = t_1^2 - 2t_1 \cos\theta_1 + 1 \tag{34}$$

$$(ak)^2 = t_2^2 - 2t_2 \cos\theta_2 + 1 \tag{35}$$

$$(bk)^2 = t_2^2 - 2t_2 t_1 \cos\theta_{12} + t_1^2 \tag{36}$$

where we define $k = d_1/Z_0, a = d_2/d_1, b = d_{12}/d_1$.

Although there are three unknowns k^2, t_1, t_2 in three equations, it is very hard to solve them by brute force. For simplicity, we further assume $a = 1$, $\angle AOB = 90°$. This assumption is easily fitted if we properly choose the three points A_0, B_0, O_0 in the reference patch. From the above assumption, we have $\mathbf{OA} \perp \mathbf{OB}$ and then

$$t_1 t_2 \cos\theta_{12} - t_2 \cos\theta_2 - t_1 \cos\theta_1 + 1 = 0 \tag{37}$$

when $\cos\theta_{12} \neq \cos\theta_1 \cos\theta_2$, we have

$$t_2 = -\frac{1 - t_1 \cos\theta_1}{\cos\theta_2 - t_1 \cos\theta_{12}} \tag{38}$$

when $\cos\theta_{12} = \cos\theta_1 \cos\theta_2$, we have

$$(t_1 \cos\theta_1 - 1)(t_2 \cos\theta_2 - 1) = 0. \tag{39}$$

Substituting equation (38) into equations (34) and (35), then we have

$$a_4 t_1^4 + a_3 t_1^3 + a_2 t_1^2 + a_1 t_1 + a_0 = 0 \tag{40}$$

where

$$a_4 = \cos^2\theta_{12}$$
$$a_3 = -2\cos\theta_2 \cos\theta_{12} - 2\cos\theta_1 \cos^2\theta_{12}$$
$$a_2 = 6\cos\theta_1 \cos\theta_2 \cos\theta_{12} + \cos^2\theta_2 - \cos^2\theta_1$$
$$a_1 = -4\cos\theta_1 \cos^2\theta_2 - 2\cos\theta_2 \cos\theta_{12} + 2\cos\theta_1$$
$$a_0 = 2\cos^2\theta_2 - 1.$$

Solving equation (40) to get t_1; and from equations (38) and (34), we can acquire t_2 and Z_0 (from k). Then the position and orientation of this object planar patch are now known!

Because equation (40) is a fourth-order equation of t_1, it seems that t_1 has four solutions. But in a practical situation, only two or three solutions are reasonably adopted (it must have $t_1 > 0$, $t_2 > 0$).

Careful readers may have found that there can be many solutions to t_1, which means that we may get two or more ambiguous tracking results as shown in Fig. 7.[7] Can the correct solution be distinguished from the wrong ones? Because affine transformation is used to approximate the matching result between the two shapes, R_{ref} and R_{obs}, and it needs a more accurate matching to solve the ambiguity, we do not think that affine transformation is accurate enough to disambiguate them.

5. FROM POSITIONS AND ORIENTATIONS OF AN OBJECT PLANAR PATCH TO OBTAIN MOTION PARAMETERS BETWEEN TWO DIFFERENT TIMES

If we can separately obtain the positions and orientations of the moving object planar patch at two different times (t_0 and t_1), it is easy to acquire the motion parameters of this moving patch, binocular vision is used here because we do not want to discuss ambiguity from monocular vision.

The object planar patch at $t = t_0$ projects two shapes R_0 and R_1 on cameras 0 and 1, and the same planar patch at $t = t_1$ also projects two shapes R'_0 and R'_1 on the two cameras. We can get the normal vector $\mathbf{u}_3$ and center (X_0, Y_0, Z_0) at $t = t_0$ from the method described in Section 2. Similarly, we can also get $\mathbf{u}'_3$ and center (X'_0, Y'_0, Z'_0) at $t = t_1$. Now, we randomly choose a point A on the object planar patch at $t = 0$. This point A corresponds to another point A' on the object at $t = t_1$. Then we define

$$\mathbf{u}_1 = \frac{\mathbf{OA}}{\|\mathbf{OA}\|}; \quad \mathbf{u}'_1 = \frac{\mathbf{O'A'}}{\|\mathbf{O'A'}\|}$$

$$\mathbf{u}_2 = \mathbf{u}_3 \times \mathbf{u}_1; \quad \mathbf{u}'_2 = \mathbf{u}'_3 \times \mathbf{u}'_1$$

and

$$U = [\mathbf{u}_1 | \mathbf{u}_2 | \mathbf{u}_3] \tag{41}$$

$$U' = [\mathbf{u}'_1 | \mathbf{u}'_2 | \mathbf{u}'_3]. \tag{42}$$

Then a point $P = (X, Y, Z)$ and its corresponding point $P' = (X', Y', Z')$ must have the following relationships:

$$\begin{bmatrix} X' - X'_0 \\ Y' - Y'_0 \\ Z' - Z'_0 \end{bmatrix} = U'U^{-1} \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix} = R \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix}. \tag{43}$$

The matrix R is the desired rotation matrix (R is a Hermitian matrix).

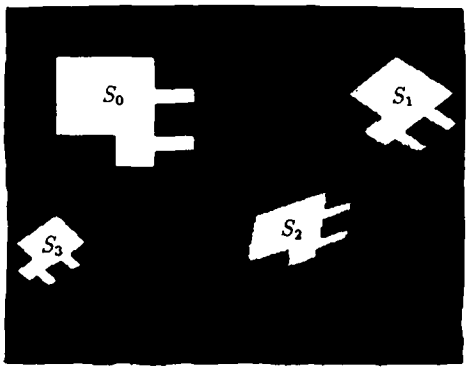


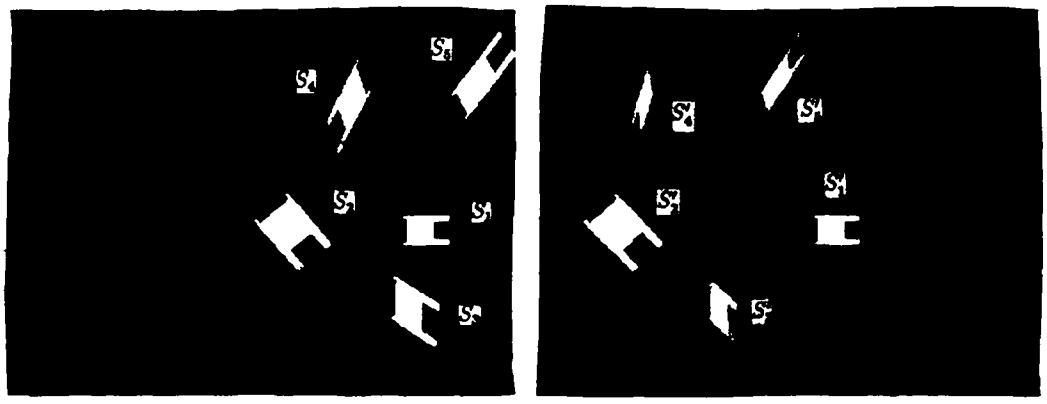
Fig. 8. Four example shapes with the relation of affine transformation. All cameras with the same focal length $f = 0.25$ m.

6. EXPERIMENT

In the following paragraph, we will present three experiments. The first experiment tests the method described in Section 2. The second experiment tests

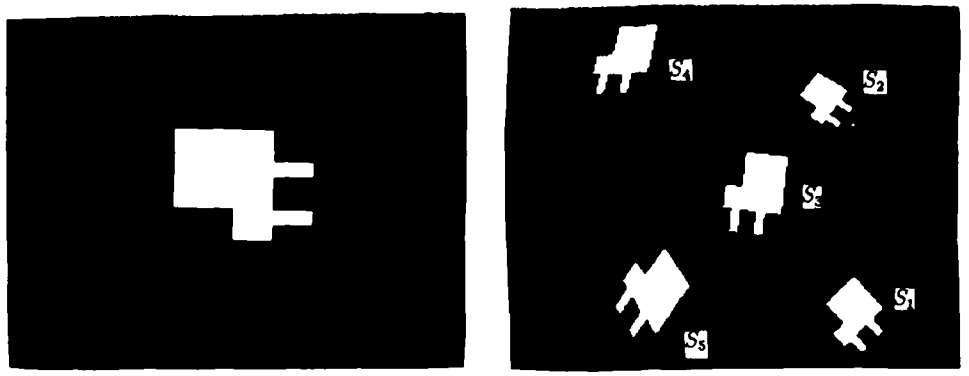
Table 1. Results of experiment A

Affine trans.	Real affine matrix	Est. affine matrix	Match ratio
$S_0 \to S_1$	$\begin{bmatrix} 0.5829 & -0.4879 \\ 0.4720 & 0.4879 \end{bmatrix}$	$\begin{bmatrix} 0.5830 & -0.4990 \\ 0.4797 & 0.4861 \end{bmatrix}$	0.9876
$S_0 \to S_2$	$\begin{bmatrix} 0.7272 & -0.0849 \\ -0.3087 & 0.6041 \end{bmatrix}$	$\begin{bmatrix} 0.7376 & -0.0806 \\ -0.3183 & 0.6106 \end{bmatrix}$	0.9815
$S_0 \to S_3$	$\begin{bmatrix} 0.2893 & -0.4060 \\ 0.3447 & 0.3407 \end{bmatrix}$	$\begin{bmatrix} 0.2945 & -0.4168 \\ 0.3647 & 0.3380 \end{bmatrix}$	0.9709



(a) (b)

Fig. 9. (a) Binocular tracking test: projected shapes of the same object planar patch with different positions and orientations on camera 0. (b) Corresponding projected shapes on camera 1.



(a) (b)

Fig. 10. (a) Monocular tracking test: R_{ref}: placed at (0, 0, 1), orientation at (0, 0, −1). (b) Different R_{obs}.

Table 2. Results of experiment B

Patch at time	Real center [Est. center]	Real orientation [Est. orientation]	Error angle
t_1	(0.200, 0.000, 2.600) [(0.201, 0.000, 2.591)]	(0.000, 0.000, −1.00) [(0.000, 0.000, −1.00)]	0°
t_2	(−0.100, 0.000, 1.700) [(−0.100, 0.000, 1.698)]	(0.087, 0.291, −0.953) [(0.069, 0.295, −0.953)]	1.05°
t_3	(0.100, 0.120, 1.550) [(0.101, 0.120, 1.547)]	(−0.839, 0.177, −0.515) [(−0.830, 0.167, −0.532)]	1.22°
t_4	(0.000, −0.180, 1.470) [(0.001, −0.180, 1.468)]	(−0.891, −0.110, −0.441) [(−0.881, −0.094, −0.464)]	1.73°
t_5	(0.200, −0.220, 1.500) [(0.202, −0.221, 1.499)]	(−0.799, −0.362, −0.481) [(−0.798, −0.352, −0.489)]	0.77°

Table 3. Results of experiment C

Patch at time	Real center [Est. center]	Real orientation [Est. orientation]	Error angle
t_1	(0.200, 0.240, 2.000)	(0.000, 0.000, −1.00)	0.0°
#1	[(0.195, 0.234, 1.952)	[(−0.000, 0.000, −1.00)	68.8°
#2	(0.195, 0.235, 1.955)]	(0.000, 0.000, −1.00)]	▷2.24°
t_2	(0.170, −0.220, 2.500)	(−0.309, −0.432, −0.847)	0.0°
#1	[(0.164, −0.209, 2.390)	[(−0.263, −0.458, −0.849)	▷3.02°
#2	(0.164, −0.209, 2.395)]	(0.114, 0.601, −0.792)]	67.9°
t_3	(0.000, 0.000, 1.800)	(0.000, 0.000, −1.000)	0°
#1	[(0.000, 0.000, 1.752)	[(−0.000, 0.000, −0.999)	▷0.85°
#2	(0.000, 0.000, 1.752)	(−0.045, −0.006, −0.999)	2.74°
#3	(0.000, 0.000, 1.752)]	(0.006, −0.045, −0.999)]	2.75°
t_4	(−0.240, −0.230, 1.780)	(−0.259, −0.483, −0.837)	0°
#1	[(−0.236, −0.224, 1.747)	[(−0.255, −0.487, −0.836)	▷0.32°
#2	(−0.231, −0.219, 1.711)]	(0.453, 0.665, −0.594)]	86.6°
t_5	(−0.160, 0.150, 1.500)	(−0.656, −0.040, −0.754)	0°
#1	[(−0.155, 0.146, 1.463)	[(0.806, −0.106, −0.582)	85.1°
#2	(−0.155, 0.147, 1.466)]	(−0.662, −0.080, −0.745)]	▷2.4°

the method mentioned in Section 3. The third experiment tests the method described in Section 4. All the simulation results consider quantization effect. Real answers are compared with estimated results.

(A) There are four shapes in Fig. 8 (size: 256 × 240). They are S_0, S_1, S_2, S_3.

(B) In this experiment, we use two cameras to track a moving object patch (at $t = t_0$). Here, Fig. 9(a) shows the image plane of camera 0, Fig. 9(b) the image plane of camera 1. The object planar patch at five randomly chosen positions is considered. At each time $t = t_i$, the object patch projects two shapes S_i and S_i' on cameras 0 and 1 (both Figs 9(a) and (b) are 256 × 240 pixels).

Nonlinearity from perspective projection affects our tracking results the most. When this factor outstands, our tracking error angle grows rapidly to about 6°–10°. However, when this factor is not apparent, the error angle is reduced to about 0°–3°.

(C) In this experiment, we use only one camera to track an object patch. Figure 10 (256 × 240) shows five projected shape S_i from different positions of the same planar patch at time $t = t_i$.

Notice that we have multiple solutions for each patch T_i, and only one of them corresponds to the correct answer.

7. CONCLUSION

In this paper, we present two methods to track an object planar patch—monocular tracking and binocular tracking. Both of them are based on affine transformation which makes the matching process become much easier than it used to be. For this reason, our methods seem to be easier than others.

Finally, there is still one thing to be emphasized. Because affine transformation is a linear transformation, any nonlinear factors will affect our results. Nonlinearity from perspective projection has the most threatening effect on the accuracy of matching from affine transformation. To reduce this crisis, the size of the projected shape should be confined to a suitably small range. However, it should not be too large or too small. On one hand, a too large shape makes the nonlinearity from perspective projection outstand; on the other hand, a too small shape forces the inference from quantization error and noise to grow up. Neither of them should happen!

REFERENCES

1. K. Kanatani, Detection of surface orientation and motion from texture by a stereological technique, *Artif. Intell.* **23**, 213–217 (1984).
2. K. Kanatani, Tracking planar surface motion from projection without knowing correspondence, *Comput. Vision Graphics Image Process.* **29**, 1–12 (1985).
3. K. Kanatani, Detecting the motion of a planar surface by line and surface integral, *Comput. Vision Graphics Image Process.* **29**, 13–22 (1985).
4. J. G. Leu, Shape normalization through compacting, *Pattern Recognition Lett.* **10**, 243–250 (1989).
5. K. Wohn and J. Wu, Estimating the finite displacement using moments, *Pattern Recognition Lett.* **11**, 371 378 (1990).
6. R. Jain, S. L. Bartlett and N. O'Brien, Motion stereo using ego-motion complex logarithmic mapping, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-9**, 356–369 (1987).
7. S. Linnainmaa, D. Harwood and L. S. Davis, Pose determination of a three-dimensional object using triangle pairs, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-10**(5), 634–647 (1988).

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.