



PII: S0031-3203(96)00094-5

PERSPECTIVE-TRANSFORMATION-INVARIANT GENERALIZED HOUGH TRANSFORM FOR PERSPECTIVE PLANAR SHAPE DETECTION AND MATCHING¹

RONG-CHIN LO^{†‡} and WEN-HSIANG TSAI^{†*}

[†]Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C.

[‡]Department of Electronic Engineering, National Taipei Institute of Technology, Taipei, Taiwan 106, R.O.C.

(Received 12 July 1995; in revised form 20 June 1996; received for publication 10 July 1996)

Abstract—In the real world, most object shapes are perspectively transformed when imaged. How to recognize or locate such shapes in images is interesting and important. The conventional generalized Hough transform (GHT) is useful for detecting or locating translated two-dimensional (2D) planar shapes. However, it cannot be used for detecting perspectively transformed planar shapes. A new version of the GHT, called perspective-transformation-invariant generalized Hough transform (PTIGHT), is proposed to remove this weakness. The PTIGHT is based on the use of a new perspective reference table that is built up by applying both forward and inverse perspective transformations on a given template shape image from all viewing directions and positions. Due to the use of the point spread function to express the perspective reference table, the required dimensionality of the Hough counting space (HCS) for the PTIGHT is reduced to 2D. After performing the PTIGHT on an input image, the peaks in the HCS whose values are larger than a threshold is picked out as the candidate locations of the perspective shape to be detected in the input images. By performing an inverse PTIGHT on the candidates, one of the candidate locations whose corresponding shape matches best with the input shape is selected and the desired parameters of the perspective transformation can be obtained. Some experimental results are included to demonstrate the applicability of the proposed PTIGHT. © 1997 Pattern Recognition Society. Published by Elsevier Science Ltd.

Generalized Hough transform	Perspective transformation invariant
Perspective reference table	Point spread function Hough counting space
Cell value incrementation	Inverse generalized Hough transformation

1. INTRODUCTION

In practical applications, most object shapes are perspectively transformed when imaged. How to recognize or locate such object shapes in images is interesting and important. The Hough transform is useful for detecting or locating straight lines or analytic curves.^(1,2) A generalized version of the Hough transform, called the generalized Hough transform (GHT),⁽³⁾ was proposed for detecting arbitrary two-dimensional (2D) planar shapes. It has been used in many applications of computer vision such as shape detection and recognition, image registration, etc. However, a weakness of the GHT is that it can only be used to detect translated shapes and is unsuitable for detecting perspectively transformed planar shapes from unknown viewpoints. To improve the GHT, Silberberg *et al.*⁽⁴⁾ proposed an iterative Hough procedure for 3D object recognition, and Jeng and Tsai⁽⁵⁾ proposed a scale- and orientation-invariant GHT to detect rotated and scaled shapes.

To improve the GHT further, a new version of the GHT, called *perspective-transformation-invariant GHT* (PTIGHT), is proposed in this study. The PTIGHT can be

used for detecting a perspectively transformed planar shape in a perspective image that is taken from an unknown viewpoint. The new approach is used on the GHT using a new *perspective reference table* (PR-table). In order to build a PR-table from a given template shape image, we first apply an inverse perspective transformation on the given template shape image to derive the points of a certain 2D shape called the *original planar shape*. Next, all possible perspective shape images are derived by applying forward perspective transformations on the derived original planar shape from all viewing directions and positions. The position of each shape point in each perspective shape image is represented by a displacement vector relative to a reference point (usually selected to be the shape centroid). The displacement vectors are then rotated 180° with respect to the reference point. The PR-table, one kind of *point spread function*,^(5,7) is constituted by superimposing the displacement vectors of all the perspective shapes with respect to an identical reference point. The PR-table contains the information of all possible perspective transformations of the original planar shape. The process of constructing the PR-table is time-consuming but it can be completed in advance before the PTIGHT is performed.

To perform the PTIGHT, a 2D Hough counting space (HCS) consisting of accumulator cells is constructed.

* Author to whom correspondence should be addressed.
[†]Partially supported by National Science Council, Republic of China under contract number NSC82-0404-E009-157.

The cell value incrementation strategy of the PTIGHT is similar to that of the conventional GHT except that the PR-table instead of the conventional R-table⁽³⁾ is used here. Although more than two perspective transformation parameters are to be found, only two parameters, namely, the x - and y -translations, are required so that the HCS is reduced to 2D as that required by the conventional GHT. Furthermore, because the PR-table is built in the form of a point spread function which contains all the perspective-transformation informations, the processing time for cell value incrementation can be shortened. After performing the PTIGHT, the peaks whose values are larger than a threshold in the HCS are detected as the candidate locations of the perspective planar shape in the input perspective image. By performing an inverse process of the PTIGHT proposed in this study, called *inverse PTIGHT*, to the candidate locations and selecting one of them whose corresponding shape matches best with the input shape, the desired parameters of the perspective transformation of the input shape image can be obtained finally.

The remainder of this paper is organized as follows. In Section 2, we present the formulas for general forward and inverse perspective transformations between a perspective shape plane and its image plane. In Section 3, we review the conventional GHT and describe the pro-

posed inverse GHT. In Section 4, the proposed PTIGHT is described in detail. Several experimental results are given in Section 5. Some concluding remarks and discussions are given in Section 6.

2. FORWARD AND INVERSE PERSPECTIVE TRANSFORMATIONS

The formulas for general forward and inverse perspective transformations between two planes are described in many papers.⁽⁸⁻¹⁰⁾ Consider the perspective transformations relationship between the plane π_0 of a perspective planar shape and its image plane π as shown in Fig. 1. The transformation can be described with two coordinate systems. One is the camera coordinate system (CCS) denoted as X, Y, Z . The other is the natural coordinate system (NCS)⁽⁸⁾ denoted as X_0, Y_0 on perspective shape plane π_0 . The observer's viewpoint is located at the origin $O=(0, 0, 0)$ of the CCS and the visual axis coincides with the Z -axis. Point $F=(0, 0, C_2)$ in plane π_0 is called the fixation point, where the parameter C_2 represents the distance between the observer's viewpoint and the fixation point. Let the origin of the NCS coincide with the fixation point and its Y_0 -axis have the direction of the cross product of the vector normal to plane π_0 and that normal to plane π . It is convenient to use the NCS to

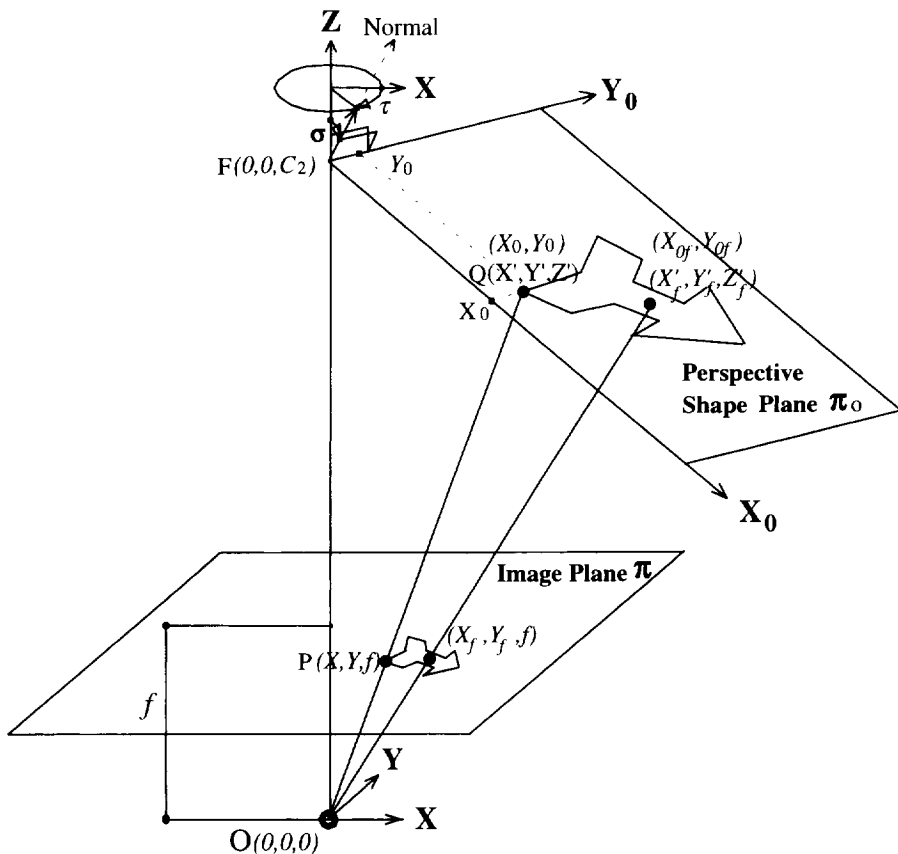


Fig. 1. Perspective transformation relationship between a perspective shape plane and its image plane. The reference point of the perspective planar shape is (X'_f, Y'_f, Z'_f) and the reference point of the perspective shape image is (X_f, Y_f, f) .

describe a planar shape in any perspective shape plane. Two other parameters of plane π_0 are the pan angle τ and the tilt angle σ , where τ is the angle between the X -axis and the projection of the normal to perspective shape plane π_0 on image plane π , and σ is the angle between the Z -axis and the normal. Assume that the direction of angle τ in Fig. 1 is positive and that of σ is negative. From the geometric relations shown in Fig. 1, the transformations between perspective shape plane π_0 and image plane π can be derived according to the following steps:

(1) The equation of image plane π in the CCS is

$$Z = f. \tag{1}$$

where f is the distance between the observer's viewpoint (or the lens center of the camera) and the image plane.

(2) The normal to perspective shape plane π_0 in the CCS is

$$n = (-\sin \sigma \cos \tau, -\sin \sigma \sin \tau, \cos \sigma). \tag{2}$$

(3) The equation of perspective shape plane π_0 in the CCS is $n \cdot (X', Y', Z')^T = C_2 \cos \sigma$ which can be transformed into

$$Z' = AX' + BY' + C_2. \tag{3}$$

where $A = \tan \sigma \cos \tau$, $B = \tan \sigma \sin \tau$.

(4) Let the viewpoint O , point P in π with CCS coordinates (X, Y, f) , and point Q in π_0 with CCS coordinates (X', Y', Z') be collinear. Then P is the image of Q . This implies the following equations:

$$\frac{X'}{X} = \frac{Y'}{Y} = \frac{Z'}{f}. \tag{4}$$

Then, we can solve X' , Y' , and Z' from equations (3) and (4) to be

$$\begin{aligned} X' &= \frac{C_2 X}{f - AX - BY}, & Y' &= \frac{C_2 Y}{f - AX - BY}, \\ Z' &= \frac{f C_2}{f - AX - BY}. \end{aligned} \tag{5}$$

(5) Now, we can derive the inverse perspective transformation from image plane π to perspective shape plane π_0 . First, we transform the CCS coordinates (X', Y', Z') of Q in π_0 into its NCS coordinates (X_0, Y_0) as follows:

$$\begin{bmatrix} X_0 \\ Y_0 \\ 0 \\ 1 \end{bmatrix} = \text{Tilt}(-\sigma) \text{Pan}(\tau) \text{Translation}(-C_2) \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}$$

where

$$\begin{aligned} \text{Tilt}(-\sigma) &= \begin{bmatrix} \cos \sigma & 0 & \sin \sigma & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \sigma & 0 & \cos \sigma & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \\ \text{Pan}(\tau) &= \begin{bmatrix} \cos \tau & \sin \tau & 0 & 0 \\ -\sin \tau & \cos \tau & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

$$\text{Translation}(-C_2) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -C_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

Then, from equations (5) and (6), we obtain the inverse perspective transformation, which maps the CCS coordinates (X, Y, f) of P to the NCS coordinates (X_0, Y_0) of Q , as follows:

$$\begin{aligned} X_0 &= C_2 \frac{(X \cos \tau + Y \sin \tau) / \cos \sigma}{f - X \tan \sigma \cos \tau - Y \tan \sigma \sin \tau}, \\ Y_0 &= C_2 \frac{(-X \sin \tau + Y \cos \tau)}{f - X \tan \sigma \cos \tau - Y \tan \sigma \sin \tau}. \end{aligned} \tag{7}$$

Recall that P on the image plane is the image of Q .

(6) Similarly, the forward perspective transformation from perspective shape plane π_0 to image plane π , which maps the NCS coordinates (X_0, Y_0) of Q to the CCS coordinates (X, Y, f) of P , can be obtained as follows:

$$\begin{aligned} X &= f \frac{X_0 \cos \sigma \cos \tau - Y_0 \sin \tau}{X_0 \sin \sigma - C_2}, \\ Y &= f \frac{X_0 \cos \sigma \sin \tau + Y_0 \cos \tau}{X_0 \sin \sigma + C_2}. \end{aligned} \tag{8}$$

3. CONVENTIONAL GENERALIZED HOUGH TRANSFORM AND PROPOSED INVERSE GENERALIZED HOUGH TRANSFORM

To use the conventional $\text{GHT}^{(3)}$ to detect an arbitrary template planar shape, it is necessary to set up a 2D HCS. Each cell in the HCS has a value specifying the possibility that the reference point of the template shape to be detected is located at the cell.

Before performing the GHT, an R-table for the template shape is built in advance by the following steps: (1) select a suitable pixel R in the given template shape as the reference point; (2) rotate the shape 180° with respect to R ; (3) trace all the template shape points and construct the R-table consisting of the displacement vectors between all the shape points and R .

In the GHT process, all the displacement vectors of the R-table are superimposed on each shape point in the input image. The value of each corresponding cell in the HCS pointed to by a displacement vector is incremented by one. If there exists any cell with its value exceeding a preselected threshold value and being the maximum in the HCS, then it is determined that the template shape is detected at the location of the cell [see Fig. 2(a)–(c) for all illustrative example].

For more complicated cases in which the HCS may receive a lot of noise or distortion^(6,7) as encountered in this study, the above peak detection method is found unstable. A peak so found sometimes does not necessarily correspond to a real template shape, but just a collection of shape pieces which come from different object shapes and happen to vote in the identical cell where the peak is located.

To overcome the above difficulty, one way is to detect those cells in the HCS that have values exceeding a certain threshold, regard the corresponding locations of

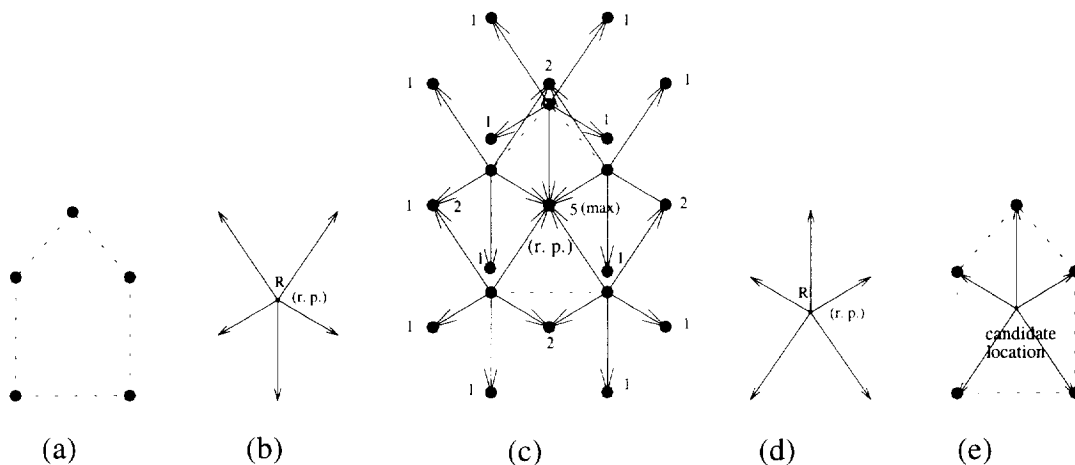


Fig. 2. Illustration of the conventional GHT and the inverse GHT: (a) the original template shape with five pixels; (b) the R-table; (c) the R-table is superimposed on the input image (containing one template shape), and a maximum of value 5 in the HCS is found at the reference point (r.p.); (d) the inverse R-table; (e) the inverse GHT is performed on the input image to verify the existence of the template shape at a candidate location, the counter of the candidate location here receives five votes.

these cells in the image space as *candidate shape locations*, and verify the existence of the template shape at each of these locations.

For the purpose of such template shape verification, we propose in this study a new technique called *inverse GHT*. To apply the inverse GHT, first an *inverse R-table* is constructed in all the same way as that for the R-table except that the template shape is not rotated 180° before the displacement vectors are collected. Then, to verify if the template shape exists at a candidate location in the input image, the inverse R-table is superimposed on the location and a counter is created for this location. The counter is incremented by one for each shape point in the input image whose location is pointed to by a displacement vector of the inverse R-table. If the final counter value is close to the number of the total pixels in the template shape, then it is decided that the template shape exists at the candidate location. This process is illustrated in Figs 2(d) and (e). Sometimes, several candidate locations may be verified, and the one with the largest counter value is selected, which will be called the *optimal candidate*.

4. PROPOSED PERSPECTIVE-TRANSFORMATION-INVARIANT GENERALIZED HOUGH TRANSFORM

In actual applications, given a template image of a planar shape and an input perspective image containing a perspectively transformed version of the planar shape, how do we verify that the shape in the input image is a perspectively transformed version of the planar shape in the template image? It is mentioned previously that the conventional GHT is unsuitable for this problem. So the PTIGHT is proposed. The basic steps are similar to those of the conventional GHT, including construction of a PR-table, creation of an HCS, incrementation of cell values, and HCS peak detection. However, an additional process, called inverse PTIGHT, is proposed for verification of

candidate shape locations. The details are described in the following.

4.1. PR-table construction

The purpose of constructing the PR-table is to include the information of all possible perspective shapes. Before describing the steps for constructing the PR-table, we have to derive appropriate formulas for creating all possible perspective shapes from the given template shape image. For this, basically we do the following two major works: (1) perform an inverse perspective transformation to derive the original planar shape from the given template image; (2) perform forward perspective transformations to derive all possible perspectively transformed versions of the original planar shape.

4.1.1. Formula derivation. The detailed steps we employ to derive formulas for creating possible perspective shapes are as follows.

(1) Assume that the shape T on the given template image is the image of a certain shape S , called the *original planar shape*, taken without any pan and tilt at a distance C_1 from the origin of the CCS under the condition that the template image plane and the plane on which S appears (called *original shape plane* henceforth) are parallel (see Fig. 3 for an illustration).

(2) Let P_1 be a point with CCS coordinate (x, y, f) of shape T on the given template image, as shown in Fig. 3.

(3) Derive the CCS coordinates (x', y', C_1) of point Q_1 on original planar shape S corresponding to P_1 on T using equations (5). Note that the original shape plane is just a special case of the perspective shape plane mentioned previously (see Fig. 1), so equations (5) are applicable here with tilt $\sigma=0$, pan $\tau=0$, and $C_2=C_1$. The derivation results are as follows:

$$x' = C_1 \frac{x}{f}, \quad y' = C_1 \frac{y}{f}. \quad (9)$$

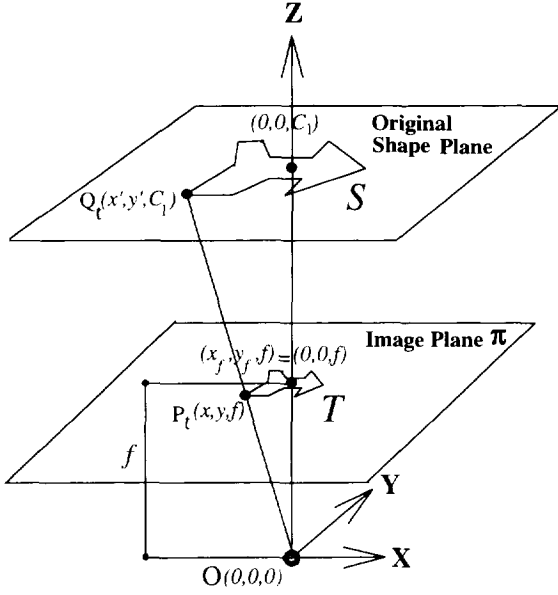


Fig. 3. Transformation relationship between the original planar shape S and its template shape image T . The reference point of the original planar shape is $(0, 0, C_1)$ and the reference point of its template shape image is $(0, 0, f)$.

or

$$\begin{aligned} \frac{X_{0f}}{C_2} &= \frac{(X_f \cos \tau + Y_f \sin \tau) / \cos \sigma}{f - X_f \tan \sigma \cos \tau - Y_f \tan \sigma \sin \tau} \\ \frac{Y_{0f}}{C_2} &= \frac{-X_f \sin \tau + Y_f \cos \tau}{f - X_f \tan \sigma \cos \tau - Y_f \tan \sigma \sin \tau} \end{aligned} \quad (10)$$

(5) Since the shape appearing in the original shape plane (see Fig. 3) is actually identical to that appearing in perspective shape plane π_0 (see Fig. 1), each pair of corresponding points in the two shapes have the same relative displacements with respect to their respective reference points. Moreover, the NCS coordinates (X_0, Y_0) of each point Q of the shape on perspective shape plane π_0 (see Fig. 1) can be expressed as $X_0 = X_{0f} + \Delta X_0$, $Y_0 = Y_{0f} + \Delta Y_0$, where $(\Delta X_0, \Delta Y_0)$ are the relative displacements of Q with respect to the reference point (X_{0f}, Y_{0f}) . Also, the relative displacement of point Q_t on original planar shape S (see Fig. 3), which corresponds to point Q on perspective shape plane π_0 , with respect to the reference point $(0, 0, C_1)$ is just (x', y', C_1) . Therefore, we have $\Delta X_0 = x'$, $\Delta Y_0 = y'$. Further from equations (9), we get

$$X_0 = X_{0f} + C_1 \frac{x'}{f}, \quad Y_0 = Y_{0f} + C_1 \frac{y'}{f}. \quad (11)$$

(6) Thus, from equations (10) and (11), equation (8) can be rewritten as follows:

$$\begin{aligned} X &= f \frac{X_0 \cos \sigma \cos \tau - Y_0 \sin \tau}{X_0 \sin \sigma + C_2} \\ &= f \frac{[(X_{0f}/C_2) + (C_1/C_2)(x'/f)] \cos \sigma \cos \tau - [(Y_{0f}/C_2) + (C_1/C_2)(y'/f)] \sin \tau}{[(X_{0f}/C_2) + (C_1/C_2)(x'/f)] \sin \sigma + 1} \\ &= f \frac{[(X_{0f}/C_2) - C(x'/f)] \cos \sigma \cos \tau - [(Y_{0f}/C_2) + C(y'/f)] \sin \tau}{[(X_{0f}/C_2) + C(x'/f)] \sin \sigma + 1} \\ Y &= f \frac{X_0 \cos \sigma \sin \tau - Y_0 \cos \tau}{X_0 \sin \sigma + C_2} \\ &= f \frac{[(X_{0f}/C_2) + (C_1/C_2)(x'/f)] \cos \sigma \sin \tau + [(Y_{0f}/C_2) + (C_1/C_2)(y'/f)] \cos \tau}{[(X_{0f}/C_2) + (C_1/C_2)(x'/f)] \sin \sigma + 1} \\ &= f \frac{[(X_{0f}/C_2) + C(x'/f)] \cos \sigma \sin \tau + [(Y_{0f}/C_2) - C(y'/f)] \cos \tau}{[(X_{0f}/C_2) - C(x'/f)] \sin \sigma + 1} \end{aligned} \quad (12),$$

Note that in the derivation we have $X = x$, $X' = x'$, $Y = y$, $Y' = y'$ in equations (5).

(4) When using the PTIGHT to check if a perspective transformed shape exists in the input image, it will result in the form of locating a reference point of the perspective transformed shape in the input image. Let the CCS coordinates of the reference point of the perspective transformed shape be (X_f, Y_f, f) . Then the NCS coordinates (X_{0f}, Y_{0f}) of the corresponding reference point of the shape in perspective shape plane π_0 (see Fig. 1) can be derived from equations (7) as follows:

$$\begin{aligned} X_{0f} &= C_2 \frac{(X_f \cos \tau + Y_f \sin \tau) / \cos \sigma}{f - X_f \tan \sigma \cos \tau - Y_f \tan \sigma \sin \tau} \\ Y_{0f} &= C_2 \frac{-X_f \sin \tau + Y_f \cos \tau}{f - X_f \tan \sigma \cos \tau - Y_f \tan \sigma \sin \tau} \end{aligned}$$

where X_{0f}/C_2 and Y_{0f}/C_2 are computed by equations (10) and $C = C_1/C_2$. Note that in equations (12), there are totally five controllable parameters C , σ , τ , and (X_f, Y_f) where (X_f, Y_f) appears in the computation of $(X_{0f}/C_2, Y_{0f}/C_2)$ by equations (10).

4.1.2. Algorithm for construction of PR-table. Now, given a template image T of a planar shape, we can construct the PR-table from it using the relevant formulas derived previously. As an illustrative example, let the given template shape image T be shown in Fig. 4(a) which includes a planar arrow shape. The PR-table are constructed from image T by the following major steps:

(1) Regard each image pixel in the image plane with CCS coordinates (X_f, Y_f, f) as the reference point

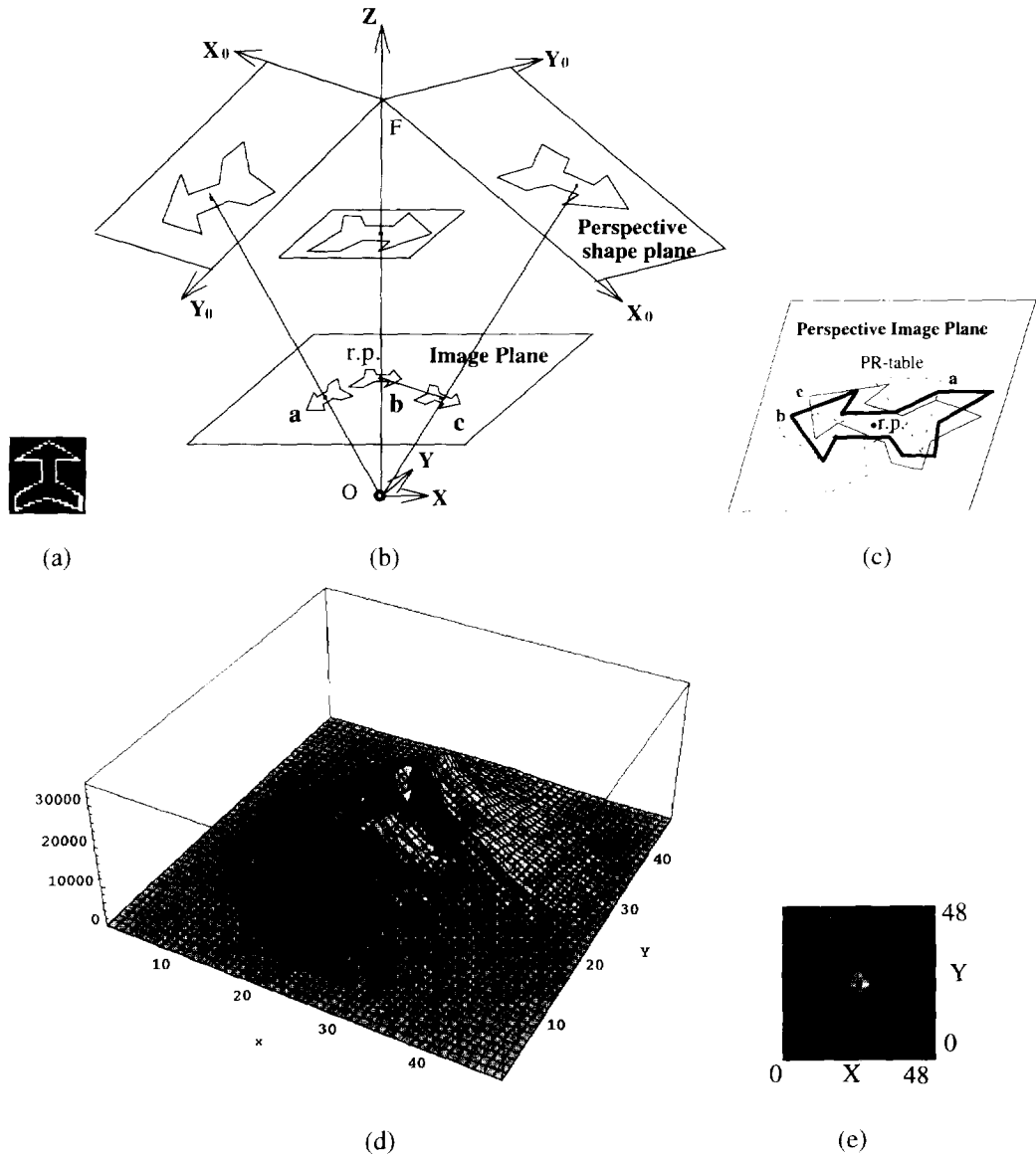


Fig. 4. Building a PR-table: (a) the known template image with an arrow shape; (b) three perspective images (denoted as a, b, and c) derived by applying perspective transformations with different parameters; (c) the perspective images a, b, and c are rotated through 180° with respect to their reference points, translated, and superimposed together to constitute the PR-table, note that the sizes of a, b, and c are magnified; (d) the complete PR-table of the known template image; (e) the PR-table of (d) shown as an image.

of the certain perspective shapes, and compute the corresponding reference point with NCS coordinates (X_{0f}, Y_{0f}) on the perspective shape plane by equations (10).

(2) Compute the locations of the image points of all possible perspective shapes in the image plane by equations (12) with different perspective parameters. For example, three possible perspective shape images denoted as a, b and c are shown in Fig. 4(b).

(3) Rotate each perspective shape in the image plane through 180° with respect to its reference point and then translate it in such a way that its reference point coincides

with a preselected common reference point selected to be the centroid of T . Finally, the PR-table is constituted by superimposing all the perspective shape images, as shown in Fig. 4(c).

The detailed steps to build the PR-table are described in the following as an algorithm.

Algorithm 1. Building a PR-table from a given template shape image.

Input:

1. A given template image T of a planar shape.
2. The focus length f of the camera.

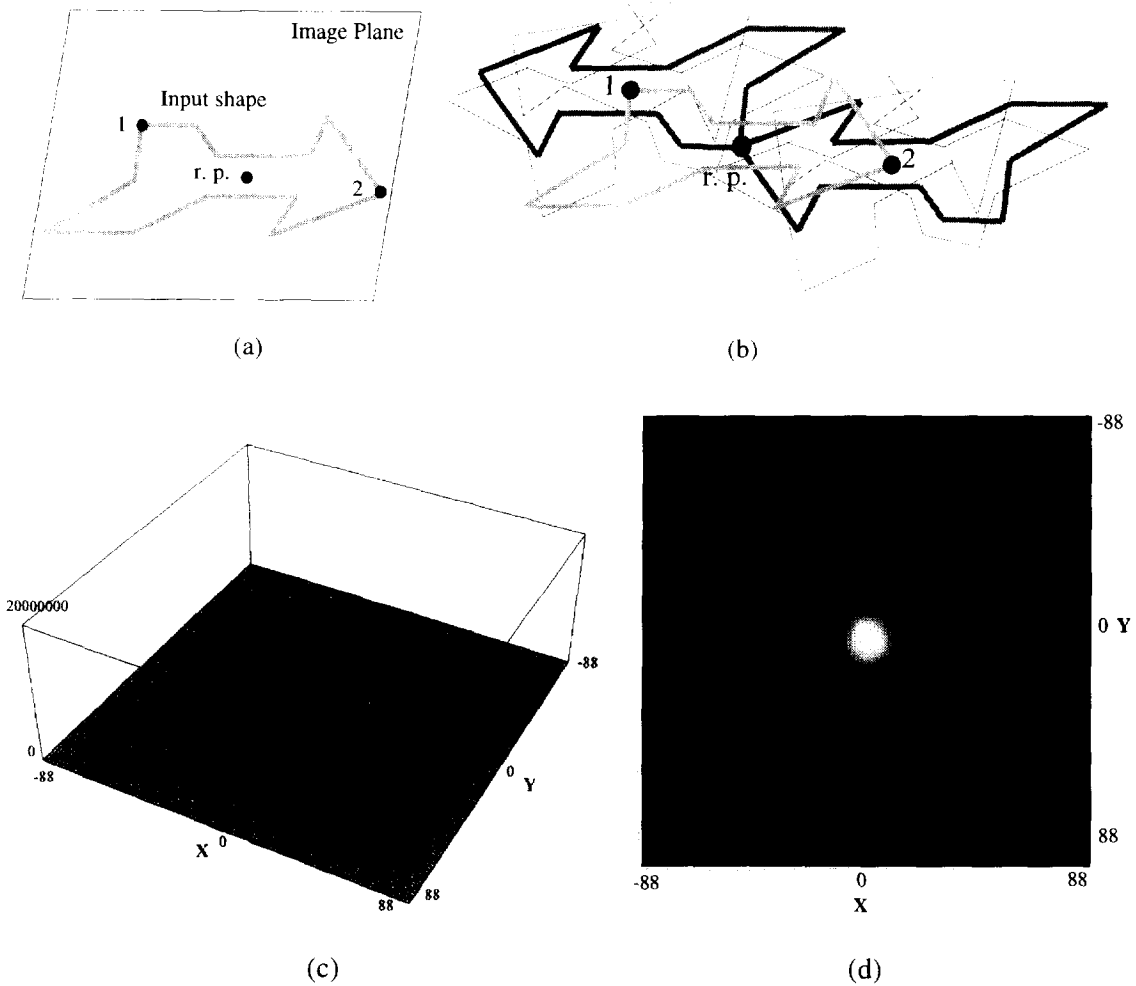


Fig. 5. Illustration of cell value incrementation: (a) an input image containing one perspectively transformed shape; (b) superimposition of part of the PR-table on two shape points (denoted as 1 and 2) of the input image; (c) 3D shape of the HCS after completing cell value incrementation; (d) the HCS of (c) shown as an image.

Output:

A PR-table.

Steps:

1. **Initialization:** Form a 2D array B as the PR-table, and set the values of all the array elements to zero.
2. **Building PR-table:** Select the centroid Q of the template image T as a reference point and let its CCS coordinates be $(0, 0, f)$. Also, regard Q as a common reference point for all perspective shapes on the image plane to be derived. Let the corresponding reference point of each possible perspective planar shape be denoted as E and the corresponding reference point of the perspective shape on the image plane be denoted as R .
 - 2.1. For each possible location with CCS coordinates (X_f, Y_f) in the image plane (regarded as a reference point R of possible perspective shapes).
 - 2.2. for $C = 0.5$ to 2.0 with increment of 0.1 ,
 - 2.3. for $\sigma = -45^\circ$ to 45° with increment of 5° ,
 - 2.4. for $\tau = -45^\circ$ to 45° with increment of 5° ,
 - 2.5. compute the NCS coordinates (X_{of}, Y_{of}) of the reference point E in the perspective shape plane corresponding to R by equations (10) [note that (X_{of}, Y_{of}) are used in computation of equations (12) of Step 2.7];
 - 2.6. for each point P_i of the template shape in the given image T ,
 - 2.7. compute the location of the corresponding point P'_i of the perspective shape in the image plane by equations (12), rotate P'_i through 180° with respect to R , and translate the displacement vector V_i between P'_i and R in such a way that R coincides with the common reference point Q ;
 - 2.8. increment by 1 the value of the cell in B which is pointed to by the displacement vector V_i ;
 - 2.9. end;
3. End.

Note that in the above algorithm, the possible ranges for the five parameters are limited to those found ade-

quate for our experiments and most applications. Continuing the illustrative example, after performing Algorithm 1 in Fig. 4(a), we get the complete PR-table as shown in Figs 4(d) and (e). The PR-table not only records the displacement vector between each pixel and the reference point of each perspective shape image but preserves the perspective information. Furthermore, the PR-table is built up intrinsically as a point spread function.

4.2. Cell value incrementation

The cell value incrementation strategy is similar to that of the conventional GHT except that the PR-table instead of the R-table is used and that the cell increment value is not always 1. To perform the PTIGHT, a 2D HCS consisting of accumulator cells is constructed first. Then, superimpose the common reference point of the PR-table on each shape point in the input image and add the value of each array element B_i of the PR-table to the corresponding cell of the HCS which is pointed to by the displacement vector of B_i . Note that the value of B_i in the PR-table is used as the increment value here, which is not necessarily the value of 1 as used in the cell value

incrementation stage of the conventional GHT. As an illustrative example, let an input image be shown in Fig. 5(a). An illustration of the result of superimposing part of the PR-table (including only three perspective shapes) on two shape points is shown in Fig. 5(b). The result of complete cell value incrementation is shown in Figs 5(c) and (d). Figure 5(c) shows the 3D shape of the resulting HCS. Figure 5(d) shows the HCS as an image.

4.3. Candidate shape locations and inverse PTIGHT

After the cell value incrementation process is completed, we locate candidate locations of the perspective shape in the input perspective image by finding the peaks whose values in the 2D HCS are larger than a threshold value. The reason why we do not detect the perspective shape simply by finding the peak in the HCS with the maximum value has been explained in Section 3 where we proposed the inverse GHT. In order to find the optimal candidate and retrieve the corresponding perspective parameters, we perform the *inverse PTIGHT* on all the candidate locations. The inverse PTIGHT is similar to the inverse GHT and the detailed steps are described in the following algorithm.

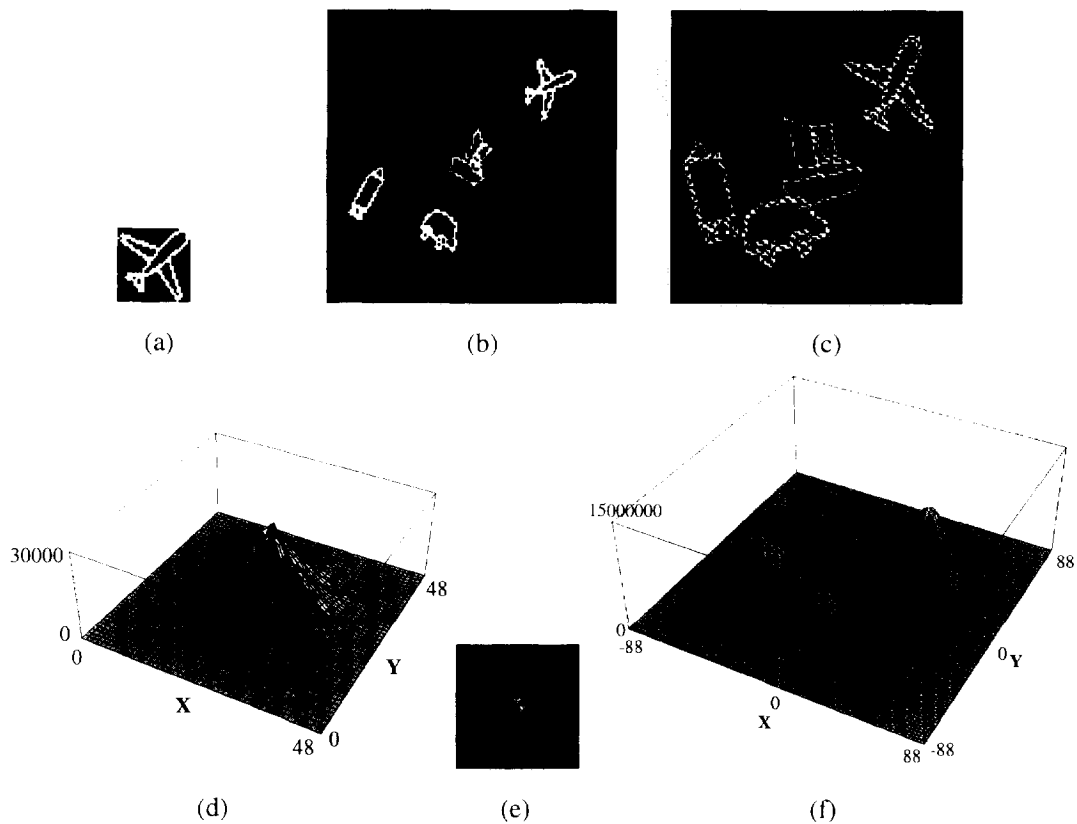


Fig. 6. Illustration of perspective airplane detection: (a) the template airplane image; (b) input image simulated by using $\sigma=-45$, $\tau=30$, $C=0.8$ and reference point (r.p.) of the perspective airplane image at $(32, -32)$; (c) another input image like (b) simulated by using $\sigma=30$, $\tau=-15$, $C=1.4$ and the r.p. at $(32, -32)$; (d) the PR-table shown as 3D shape; (e) the same as (d) but shown as an image; (f) and (g) the resulting HCSs after the PTIGHT are performed on (b) and (c) respectively; (h) and (i) all detected candidates in (f) and (g), respectively; (j) and (k) recomputed perspective shapes using retrieved sets of parameters.

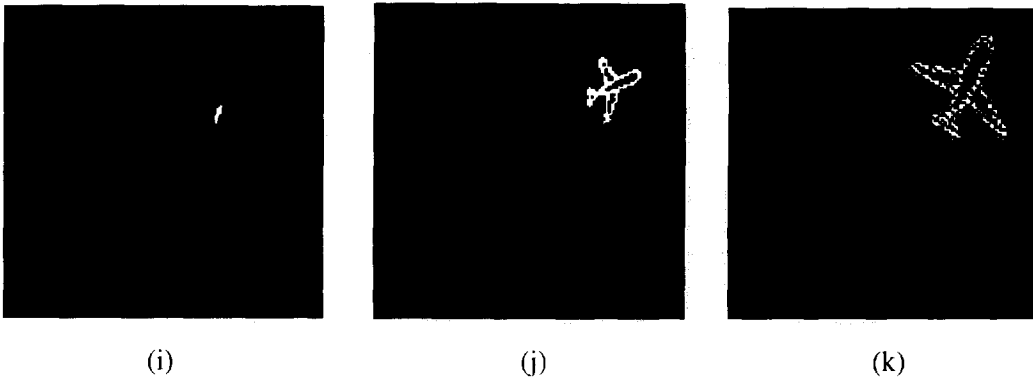
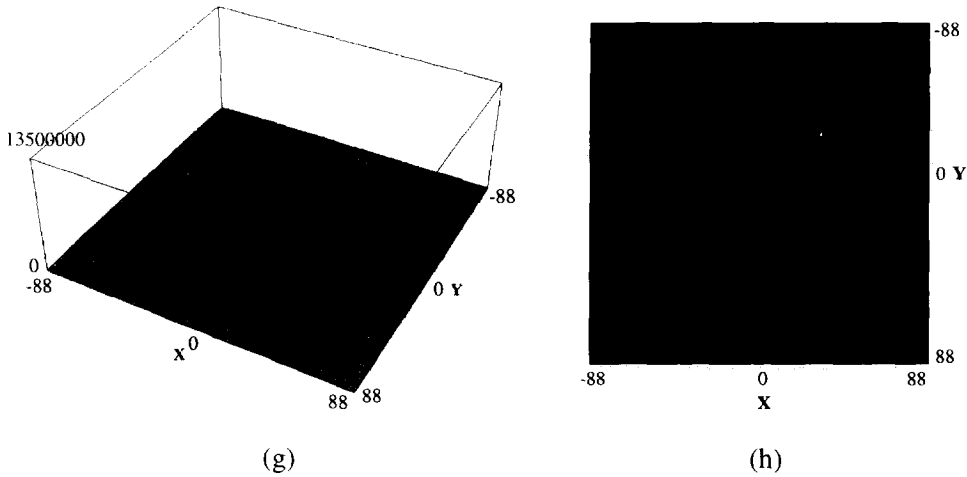


Fig. 6. (Continued).

Algorithm 2. Inverse PTIGHT.

Input: A set of candidate shape locations in the HCS and the input template shape image T .

Output: A set of desired perspective transformation parameters of the input perspective shape in the input image.

Steps:

1. For each candidate shape location (X_f, Y_f) ,
 - 1.1. for each possible set of perspective parameters C, σ and τ .
 - 1.2. *Construct perspective shape:* Construct the image of a perspective shape U from original template shape image T by equations (12) using $(X_f, Y_f), C, \sigma$ and τ , and count the total number C_1 of pixels of U ;
 - 1.3. *Build inverse PR-table:* Considering the perspective shape U as a template shape, select the centroid R of U as a reference point, form a 2D cell array S_i as the inverse PR-table, set all values of S_i to zero, and for each point P_i of U , increment by 1 the value of the cell in S_i which is pointed to by the displacement vector V_i between P_i and R ;
 - 1.4. *Inverse counting:* Create a counter C_2 , set its content to zero, superimpose all the displacement

vectors of the inverse PR-table on the candidate shape location, and increment C_2 by 1 each time a point of U is pointed to by a displacement vector of the inverse PR-table.

1.5. end.

2. *Optimal candidate detection:* Find the optimal candidate location with the maximum ratio of C_2 to C_1 among all possible sets of perspective parameters and all candidate locations. The perspective transformation parameters of the corresponding perspectively transformed template shape are output as the desired parameters.
3. End.

4.4. PTIGHT algorithm

Now we are ready to describe the overall PTIGHT algorithm as follows:

Algorithm 3. PTIGHT for detecting a template shape from an input perspective image.

Input:

1. A given template image T of a known planar shape.
2. An input image containing a partial or full perspective transformed version of the known planar shape whose viewpoint is unknown.

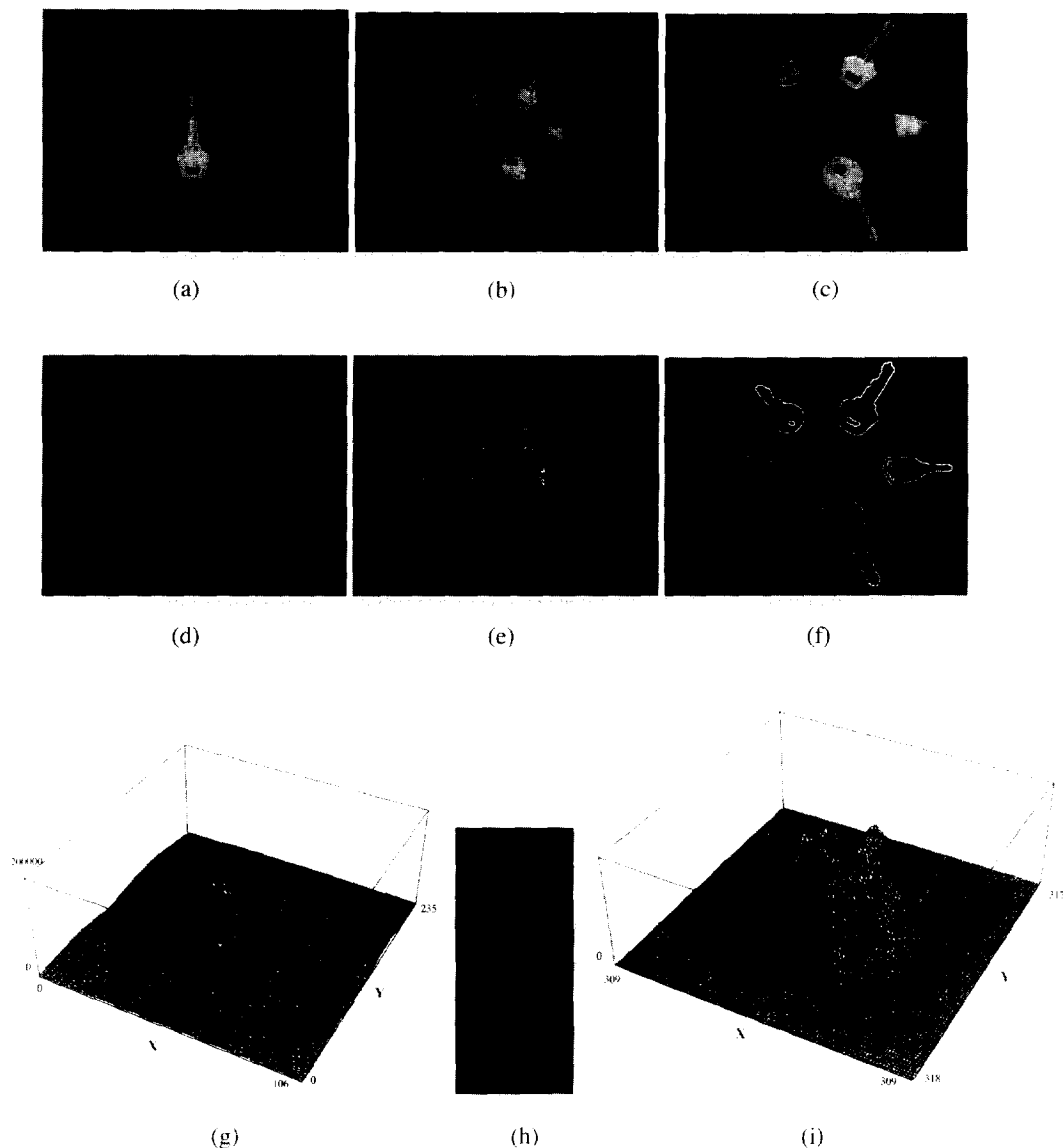


Fig. 7. Illustration of practical perspective key detection: (a) a real template key image; (b) and (c) two real input images; (d), (e) and (f) the contours of (a), (b) and (c); (g) the PR-table of (d) shown as 3D shape; (h) the same as (g) but shown as an image; (i) and (j) the resulting HCSs after the PTIGHT are performed on (e) and (f); (k) and (l) all detected candidates in (e) and (f), respectively; (m) and (n) recomputed perspective template using retrieved sets of parameters.

3. The focus length f of the camera and a threshold value t_v .

Output:

1. A location in the input image where the perspective transformed shape appears (or more specifically, where the reference point of the perspectively transformed shape is located).
2. A set of parameters for describing the perspective transformation from the perspective shape plane to the image plane.

Steps:

1. *Initialization:* Form a 2D HCS and set all values of the cells in the HCS to zero.

2. *Building PR-table:* Build the PR-table from the given template image T by Algorithm 1.

3. *Cell value incrementation:* Increment the values of the cells in the HCS using the PR-table by the process described in Section 4.2.

4. *Detecting candidate shape locations:* Find the candidate shape locations in the HCS with all values exceeding t_v .

5. *Performing inverse PTIGHT:* Detect the optimal candidate and retrieve the corresponding perspective parameters by Algorithm 2.

6. End.

In the above algorithm, Step 2 is time-consuming but can be completed beforehand. Due to using the point

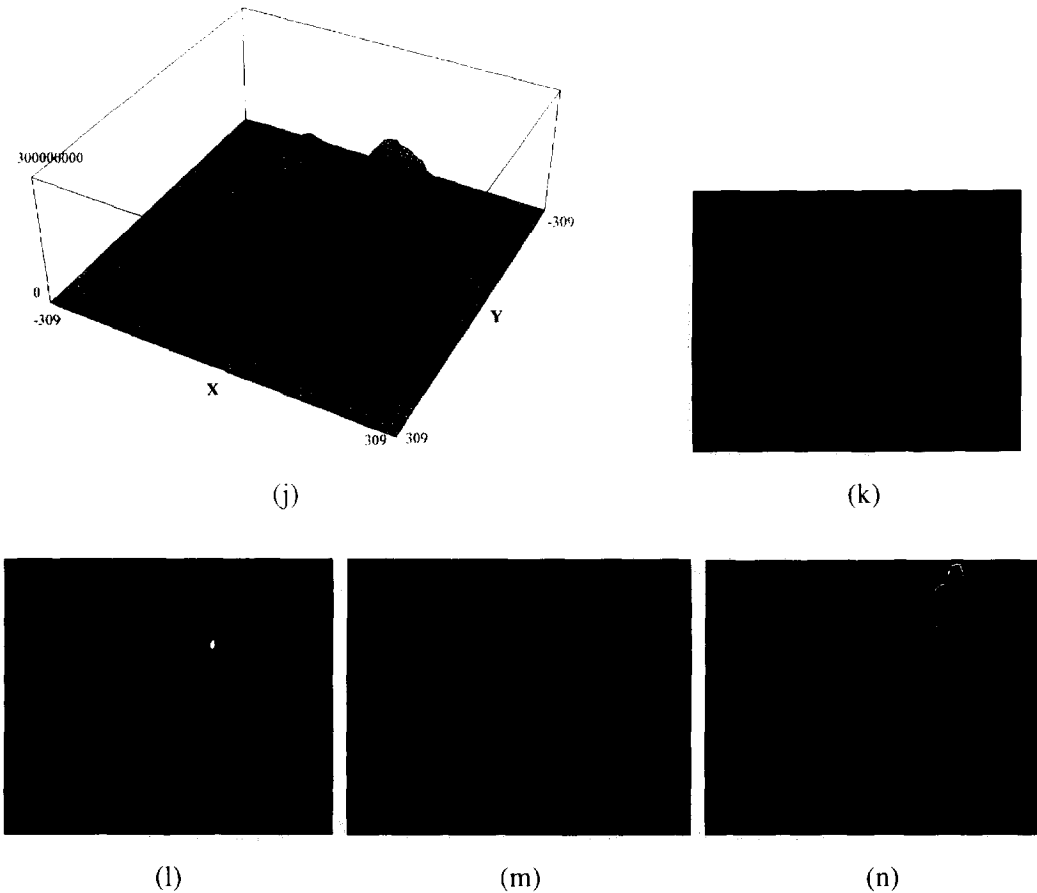


Fig. 7. (Continued).

spread function to build the PR-table in Step 2, the cell value incrementation stage in Step 3 is similar to that of the conventional GHT except that the PR-table instead of the conventional R-table is used here. Only two parameters in the HCS, namely, x - and y -translations, are required.

5. EXPERIMENTAL RESULTS

The proposed PTIGHT algorithm has been implemented on a SUN SPARC 10 workstation and several sets of images have been tested. Some experimental results are shown in Figs 6–8 using the focus length $f=1628$ pixels.

In Fig. 6, an airplane template image with 32×32 pixels is shown in Fig. 6(a) and two input perspective images with 128×128 pixels are shown in Figs 6(b) and (c). The input perspective image of Fig. 6(b) is obtained by simulation with parameters $\sigma=-45$, $\tau=30$ and $C=0.8$. The perspective template image is translated to $(X_f, Y_f) = (32, -32)$. The input perspective image of Fig. 6(c) is obtained by simulation with parameters $\sigma=30$, $\tau=-15$ and $C=1.4$. The perspective template image is translated to $(X_f, Y_f) = (32, -32)$. Figure 6(d) is the PR-table of Fig. 6(a). Figure 6(e) shows the PR-table of Fig. 6(a) as an image. Figures 6(f) and (g) show the resulting HCSs for Figs 6(b) and (c), respectively, after the PTIGHT is performed. We detect all

peaks that have cell values larger than the threshold value t_v and find the optimal candidate from them. Then, the perspective transformation parameters of the optimal candidate are retrieved. Figures 6(h) and (i) show all detected candidates. For Fig. 6(b), the retrieved parameters are $\sigma=-45$, $\tau=30$, $C=0.8$ and $(X_f, Y_f) = (32, -32)$; and for Fig. 6(c), the parameters are $\sigma=30$, $\tau=-15$, $C=1.4$ and $(X_f, Y_f) = (32, -32)$. By using these two sets of parameters to recompute the locations of the perspective template images, the results are shown in Figs 6(j) and (k). Comparing Fig. 6(b) with Fig. 6(j), and Fig. 6(c) with Fig. 6(k), we see that they have identical locations and shape points.

In Fig. 7, a key template image with 71×157 pixels is shown in Fig. 7(a) and two input perspective images with 512×400 pixels are shown in Figs 7(b) and (c). Figures 7(a)–(c) are obtained through a real TV camera. Figures 7(d)–(f) are the contours of Figs 7(a)–(c), respectively. Figures 7(g) and (h) are the PR-table of Fig. 7(d). Figure 7(g) shows the PR-table as a 3D shape and Fig. 7(h) shows the PR-table as an image. After the PTIGHT is performed in Figs 7(e) and (f), the resulting HCSs are shown in Figs 7(i) and (j). We detect all peaks that have cell values larger than the threshold value t_v and find the optimal candidate from them. Then, the perspective transformation parameters of the optimal candidate are retrieved. Figures 7(k) and (l) show all detected candi-

dates. For Fig. 7(b), the retrieved perspective parameters are $\sigma=10$, $\tau=35$, $C=0.6$ and $(X_f, Y_f) = (39, -73)$ and for Fig. 7(c), the parameters are $\sigma=-5$, $\tau=35$, $C=1.0$ and $(X_f, Y_f) = (80, -125)$. By using these two sets of parameters to recompute the template shapes, the results are shown in Figs 7(m) and (n). Comparing Figs 7(b) and (e) with (m), and Figs 7(c) and (f) with (n), we see that the

detected locations and the recomputed shapes of the keys are quite close to their original locations and the shapes.

In Fig. 8, a real image containing a film box template is shown in Fig. 8(a) and an input real perspective-transformed image containing three kinds of boxes as well as one chewing gum is shown in Fig. 8(d). Both of Figs 8(a)

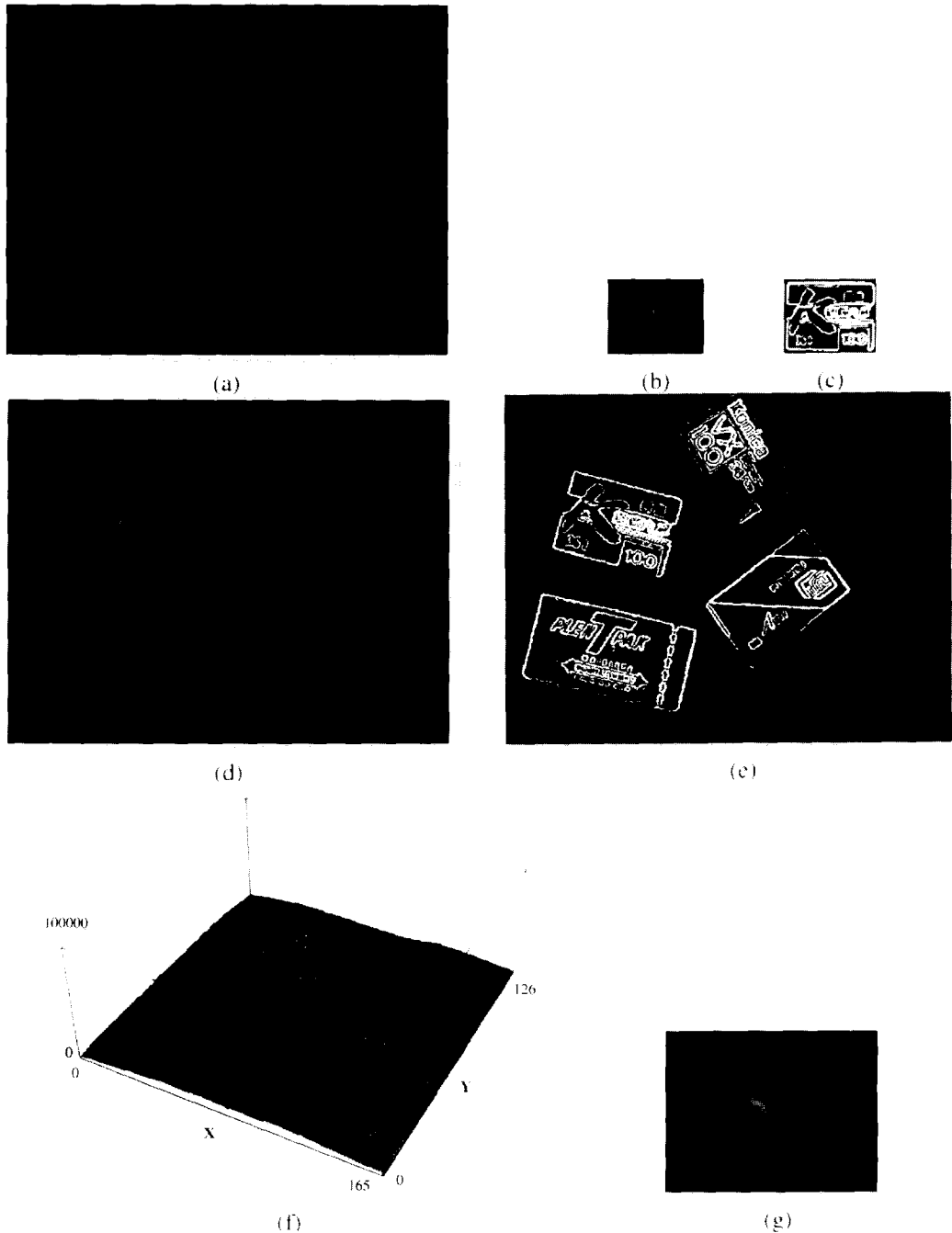


Fig. 8. Illustration of practical perspective film box detection: (a) a real image containing film box template; (b) the segmented template image from (a); (c) the resulting feature detection from (b); (d) the input image; (e) the resulting feature detected from (d); (f) and (g) the constructed PR-table from (c) shown as a 3D shape and an image, respectively; (h) and (i) the resulting HCSs shown as a 3D shape and an image, respectively, after performing the PTIGHT on (e); (j) all detected candidates; (k) recomputed perspective template using retrieved parameters.

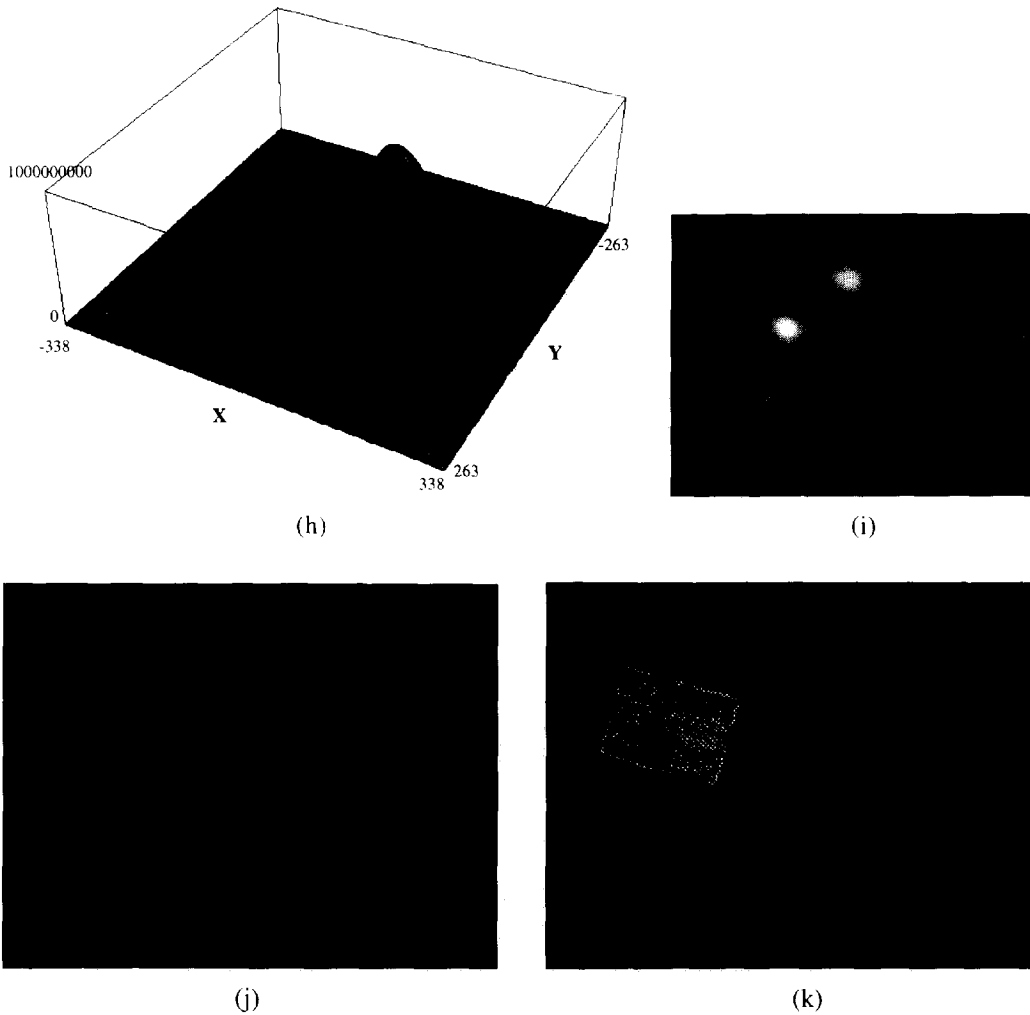


Fig. 8. (Continued).

and (d), whose image sizes are 512×400 pixels, are obtained through a real TV camera. Figure 8(b) is the film box template, whose size is 110×84 pixels, segmented from Fig. 8(a). After performing the preprocessing steps of edge detection, inverse and thresholding in Figs 8(b) and (d), Figs 8(c) and (e) are obtained, respectively. Figure 8(f) is the PR-table with size 165×126 pixels constructed from Fig. 8(c), and Fig. 8(g) is the 2D image of Fig. 8(f). After the PTIGHT is performed in Fig. 8(e), the resulting HCS is shown as a 3D shape and an image in Figs 8(h) and (i), respectively. We detect all peaks that have cell values larger than the threshold value t_v , and find the optimal candidate from them. Then the perspective transformation parameters of the optimal candidate are retrieved. Figure 8(j) shows all the detected candidates. For Fig. 8(d), the retrieved perspective parameters are $\sigma=5$, $\tau=15$, $C=1.2$ and $(X_f, Y_f) = (-122, -51)$.

The result of recomputing the template shape using the parameters is shown in Fig. 8(k). We see that the detected location and the recomputed shape of the film box in Fig. 8(k) are quite close to the original location and shape in Fig. 8(e).

It takes about 10 min on an average to process an input image. The speed can be reduced if the resolution of the HCS space is reduced.

6. CONCLUSIONS

A new version of the GHT, called PTIGHT, has been proposed, which can be used to detect and locate a template planar shape in a perspective input image with an unknown viewpoint. The new transform is perspective transformation invariant. In the proposed PTIGHT process, a PR-table instead of the R-table of the conventional GHT is constructed first. The PR-table contains all perspective transformation information of the template shape. Constructing the PR-table is time-consuming but it can be done in advance before the PTIGHT is performed. The cell value incrementation strategy of the PTIGHT is similar to that of the conventional GHT. It only needs a 2D HCS so that the storage and computation requirements can be effectively reduced. After detecting the peaks in the HCS whose values are larger than a threshold value as the candidate shape locations, the perspective transformation parameters of the optimal

candidate are retrieved by using the proposed inverse PTIGHT. From the experimental results it is seen that the PTIGHT has good potential for practical applications. How to extend the PTIGHT to gray-scale and color images is worth further research.

REFERENCES

1. P.V.C. Hough, Method and means for recognizing complex patterns, *U.S. Patent*, 3069654 (1962).
2. R. O. Duda and P. E. Hart, Use the Hough transform to detect lines and curves in pictures, *Commun. Assoc. Comput.* **15**, 11–15 (1972).
3. D. H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, *Pattern Recognition* **13**, 111–122 (1981).
4. T. M. Silberberg, L. Davis and D. Harwood, An iterative Hough procedure for three-dimensional object recognition, *Pattern Recognition* **17**, 621–629 (1984).
5. S.C. Jeng and W.H. Tsai, Scale- and orientation-invariant generalized Hough transform—A new approach, *Pattern Recognition* **24**, 1037–1051 (1991).
6. T. M. Vanveen and F. C. Groen, Discretization errors in the Hough transform, *Pattern Recognition* **14**, 137–145 (1981).
7. C. M. Brown, Inherent bias and noisy in the Hough transform, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-5**, 493–505 (1983).
8. Z. Pizza and A. Rosenfeld, Recognition of planar shapes from perspective images using contour-based invariants, *CVGIP: Image Understanding* **56**, 330–350 (1992).
9. K. C. Wong and J. Kittler, Recognizing polyhedral objects from a single perspective view, *Image and Vision Computing* **11**, 211–220 (1993).
10. M. Dhome, M. Richetin, J. T. Laprest and G. Rives, Determination of the attitude of 3-D objects from a single perspective view, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-11**, 1265–1278 (1989).

About the Author — WEN-HSIANG TSAI (S'78-M'80-SM'91) received his B.S. degree from National Taiwan University in 1973, M.S. degree from Brown University in 1977, and Ph.D. degree from Purdue University in 1979, all in electrical engineering. Since November 1979, he has been associated with faculty of the Institute of Computer Science and Information Engineering at National Chiao Tung University, Hsinchu, Taiwan. From 1984 to 1984, he was an Assistant Director and later an Associate Director of the Microelectronics and Information Science and Technology Research Center at National Chiao Tung University. He joined the Department of Computer and Information Science at National Chiao Tung University in August 1984, acted as the Head of the Department from 1984 to 1988, and is currently a Professor. He now acts as the Dean of General Affairs of the University. He serves as a consultant to several research institutes and industrial companies. His current research interests include computer vision, image processing, pattern recognition, and neural networks. Dr Tsai is an Associate Editor of *Pattern Recognition*, *Journal of the Chinese Institute of Engineers*, and *Journal of Information Science and Engineering*, and was an Associate Editor of *International Journal of Pattern Recognition and Artificial Intelligence*, *Computer Quarterly*, and *Proceedings of National Science Council of the Republic of China (Part A)*. He was elected as an Outstanding Talent of Information Science of the Republic of China in 1986. He was the winner of the 13th Annual Best Paper Award of the Pattern Recognition Society of the U.S.A. He obtained the 1987 Outstanding Research Award and the 1988–1989, 1990–1991 Distinguished Research Awards of the National Science Council of the Republic of China. He also obtained the 1989 Distinguished Teaching Award of the Ministry of Education of the Republic of China. He was the winner of the 1989 AceR Long Term Award for Outstanding Ph.D. dissertation supervision, the 1991 Xerox Foundation Award for Ph.D. dissertation study supervision, and the 31st S.K. Chuang Foundation Award for Science Research. He was also the winner of 1990 Outstanding Paper Award of the Computer Society of the Republic of China. Dr Tsai has published more than 80 papers in well-known international journals. He is a senior member of the IEEE, and a member of the Chinese Image Processing and Pattern Recognition Society, the Computing Linguistics Society of the Republic of China, and the Medical Engineering Society of the Republic of China.

About the Author — RONG-CHIN LO received B.S. degree in physics from National Cheng-Kung University in 1974 and M.S. degree in electrical engineering from Tatung Institute of Technology in 1982. He worked as a Supervisor at Great Ltd. from 1982 to 1984, and as a Deputy Manager at China Data Processing Center from 1984 to 1985. His major work was to develop new products. He joined the Department of Electronic Engineering at S. John's & S. Mary's Institute of Technology from 1985 to 1989. In August 1989, he joined the Department of Electronic Engineering at National Taipei Institute of Technology, and is currently a lecturer there. He is now working toward his Ph.D. degree in computer and information science at National Chiao Tung University. His research interests include pattern recognition, image processing, and autonomous vehicle navigation.