New approach for scale, rotation, and translation invariant pattern recognition

Wen-Hao Wang
Yung-Chang Chen
National Tsing Hua University
Institute of Electrical Engineering
Hsinchu, Taiwan 30043
E-mail: ycchen@ee.nthu.edu.tw

Abstract. A new approach for scaling, rotation, and translation invariant object recognition is proposed. The geometric discrepancy between the test pattern and the related reference pattern can be reduced by applying affine transform to the test pattern with the parameters estimated from the affine relation between the control points of both parties. The automatically paired control points are the centroids of patterns' desired regions with specific gray level distributions. By means of the analysis of intrinsic invariance properties, a small number of reference templates are required to deal with the residual geometric variation of the transformed test pattern. Consequently, a simple similarity measure can be employed directly without any scale and rotation invariant feature extraction.
[S0091-3286(97)01804-7]

Subject terms: invariant pattern recognition; affine transform; intrinsic invariance property.

Paper 31066 received June 22, 1996; revised manuscript received Oct. 23, 1996; accepted for publication Oct. 24, 1996.

1 Introduction

This paper is primarily concerned with the object recognition invariant to object's scaling, rotation, and translation (SRT). This type of recognition plays an important role in machine vision, pattern recognition [e.g., optical character recognition (OCR)], or automatic inspection. It is well known that the conventional correlation method is susceptible to object's geometric variation (rotation and scaling), i.e., only a small range of variation can be tolerated. Hence, it entails extracting invariant features before recognition. Many methods for invariant recognition have been proposed by using[1] (1) transform coefficient features, (2) algebraic features, (3) visual features, and (4) moment-based methods.

The moment invariants method[2,3] is often utilized for invariant recognition or character recognition. Flusser and Suk[4] used the affine moment invariants for object matching to estimate the affine parameters in the application of image registration. Sheng and Shen[5] proposed the orthogonal Fourier-Mellin moments, which are thought of as generalized Zernike moments and orthogonalized complex moments. However, moment-based methods tend to be time-consuming and fall short of expectation in case of noise degradation. Teh and Chin[6] analyzed the effect of noise of image on various types of moments, and examined the properties and the interrelations among them.

Circular harmonic (CH) method[7,8] has been well utilized to deal with the issue of rotation invariant recognition. The technique equivalent to the CH is the Mellin radial harmonics,[9] which was employed for scale invariant recognition; alternatively, that can also be achieved by means of the higher order autocorrelation features.[10] Except the CH and Mellin radial harmonics, most of the invariant recognition methods are capable of identifying objects invariant to SRT simultaneously, e.g., Fourier-Mellin descriptors.[11] Fang's self-transform[12] employed the pattern itself in the kernel of a transform that is invariant to objects SRT variations. By means of this transform, a number of coefficients extracted from the transformed domain can be utilized as the invariant features.

On the other hand, some approaches were proposed for shape recognition. Wang *et al.*'s moment Fourier descriptor[13] was exploited to recognize object by shape (application of OCR). Flusser[1] proposed an improved shape matrix method to depict the shape of object, and then matching likelihood coefficients were used for invariant matching. Log-polar mapping[14] was another popular method by which object scale and orientation can be transformed into translations. In addition to the invariant recognition, the method also offers the advantage of estimating the geometric features of the test pattern — scale and rotation. This function is not incorporated in other methods.

The solutions to recognition or classification problems can be considered as the combination of the following four components[15]: (1) a feature space, (2) a search space, (3) a search strategy, and (4) a similarity metric. On account of the variations of SRT of the test pattern, the search space is the whole 4-D space. The search will be very time-consuming. Caelli and Liu[16] proposed an adaptive approach which is based on the analysis of pattern's ''*intrinsic invariance property*.'' By virtue of the analysis of the property, a set of exquisite templates can be found and designed with different orientations and scales for different classes of patterns, i.e., the search space can be reduced. It is claimed that the considerable number of correlations can be processed in a parallel manner. It is deplorable, however, that enormous numbers of templates are expected to render the parallel implementation impractical; still, the search space is too large.

To reduce the large search space, a registration-based search strategy is proposed in this paper. The main principle is to apply the affine transform to the test pattern to acquire an image that is, in geometry, close to the related reference pattern. Hence, the geometric discrepancy and the search space can be reduced. The affine parameters are estimated by paired control points. These points are the centroids of segmented regions obtained from the test and the reference patterns, respectively. These points are used for matching (point pairing) and estimating the required affine parameters. Due to the utilization of the affine transform, all kinds of affine deformations of the test pattern can be transformed to be similar to the reference pattern. This is much different from Sheng and Arsenault's method,[17] which is a normalization approach and only invariant to SRT.

The proposed system, which needs no invariant feature extraction, can be divided into two stages. The first stage is responsible for registering the test pattern to be close to the related reference pattern using the centroids of thresholded images as control points. Note that the centroid of a region is very stable under random noise and gray level variations.[4] The second stage is utilized to evaluate the similarity. Two kinds of reference sets are defined. In the first kind, only one template is required for each class if the geometric discrepancy between the transformed test pattern and the related reference pattern is small. The second set consists of different numbers of templates for different classes. The scales and rotation angles of the templates are designed according to the intrinsic invariance properties of objects.

The organization of the paper is as follows. Section 2 derives the affine relation between two sets of centroids of thresholded images. The proposed algorithm of invariant pattern recognition is described in Sec. 3. Section 4 demonstrates the experimental results. The optical implementation is discussed in Sec. 5. Section 6 brings up the conclusions.

2 Affine Relation Between Two Sets of Centroids of Thresholded Images

Suppose that two images are given, with one image affine-transformed related to the other. The centroids of the partial images, which are obtained by specific gray level thresholds, can satisfy the same affine relation between the original images. Thus, these centroids can be used as the control points for estimating the affine parameters between these two images. The principle is shown as follows.

Figure 1(a) shows reference image I_1, which contains three regions with different gray levels. Region U_{11} has gray level 50 and its centroid is located at C_{11} with mark ''x.'' Region U_{12} has gray level 125 and its centroid is C_{12}. Region U_{13} has gray level 200 and its centroid is C_{13}. Figure 1(b) is image I_2, which is the rotated, translated and down-scaled version of I_1. The affine relation between I_1 and I_2 is expressed as

$$P_2=\Lambda P_1+T,$$

$$\Lambda=s\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix},\quad T=\begin{pmatrix}t_x\\ t_y\end{pmatrix}, \tag{1}$$

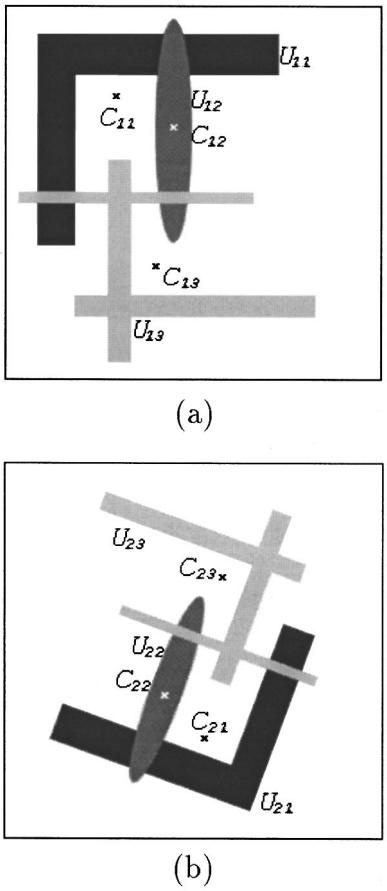


Fig. 1 (a) Original image with three gray levels and (b) the image after affine transform.

where P_1 denotes the coordinate (on the grid) of a point in I_1, and P_2, which may not be on the grid of I_2, is the mapping of P_1. Here Λ is the scale and rotation matrix and $\mathbf{T}$ is the translation vector. Image I_i can be divided into $(L+1)$ regions and be denoted as

$$I_i=\{U_{i1},U_{i2},\cdots,U_{iL},\eta_i\}\quad i=1,2, \tag{2}$$

where U_{ij} denotes each thresholded region and η_i is the uninteresting area.

Let $\xi_{1j}[l]\in U_{1j}$ and $\xi_{2j}[l]\in U_{2j}$ be the coordinate vectors of points, denoted by

$$\xi_{ij}[l]=\begin{pmatrix}\breve{x}_{ij}[l]\\ \breve{y}_{ij}[l]\end{pmatrix},\quad i=1,2;\quad j=1,\cdots,L;\quad l=1,\cdots,N_{ij}, \tag{3}$$

where N_{ij} denotes the pixel number of the j'th region in image I_i. Hence, U_{ij} consists of the point sets of thresholded images, which are represented as follows:

$$U_{ij}=\{\xi_{ij}[l]|t_{j1}<g(\xi_{ij}[l])<t_{j2},l=1,\cdots,N_{ij}\}\quad i=1,2;\quad j=1,\cdots,L, \tag{4}$$

where $g(x)$ is the gray level at point x. The point sets U_{1j} and U_{2j} are obtained by thresholding the two images using the threshold interval z_j in the threshold set Z,

$$Z=\{z_j=(t_{j1},t_{j2})|j=1,\cdots,L\}, \tag{5}$$

where t_{j2} is the upper bound of z_j and t_{j1} is the lower bound.

Note that the number N_{1j} may not be equal to N_{2j} owing to the scaling effect. Let C_{ij} denote the centroid of area U_{ij}. C_{1j} and C_{2j} can be evaluated as follows:

$$C_{1j}=\frac{\Sigma_{\xi_{1j}\in U_{1j}}\xi_{1j}}{N_{1j}}, \tag{6}$$

$$C_{2j}=\frac{\Sigma_{\xi_{2j}\in U_{2j}}\xi_{2j}}{N_{2j}}\simeq\frac{\Sigma_{\xi_{2j}\in\Omega}\xi_{2j}}{N_{1j}}, \tag{7}$$

where Ω is the point set that contains the points of U_{2j}. The point in Ω is the corresponding nearest mapping of each point in U_{1j}; thus, the set Ω has N_{1j} points. If $N_{2j}<N_{1j}$, U_{2j} is a subset of Ω. If $N_{2j}>N_{1j}$, Ω is a subset of U_{2j}. By using Eq. (1) we have

$$\xi_{2j}\simeq\Lambda\xi_{1j}+T. \tag{8}$$

Substituting Eq. (8) into Eq. (7), we have

$$C_{2j}\simeq\frac{\Sigma_{\xi_{1j}\in U_{1j}}[\Lambda\xi_{1j}+T]}{N_{1j}}=\Lambda\frac{\Sigma_{\xi_{1j}\in\Omega}\xi_{1j}}{N_{1j}}+T=\Lambda C_{1j}+T. \tag{9}$$

From Eq. (9), it is clear that the centroid pair (C_{1j},C_{2j}) also satisfies the same affine relation [i.e., Eq. (1)] approximately. Hence, by using the centroid pairs, the affine parameters between two images can be estimated. It is important to note that these centroid pairs (C_{1j},C_{2j}) are, so to speak, ''*labeled points*'' because they are generated from the same process, i.e., C_{1k} and C_{2k} can be automatically paired as soon as these two centroids are available. Hence, the design of the pairing algorithm is not required in the proposed method. In general, finding the pairing of two sets of control points is a crucial and difficult task since the invariant features (e.g., moment invariants) of the small area around each control point must be evaluated. It is usually time-consuming to some extent. However, this process can not be omitted in the conventional registration methods.

3 Classification System Based on Registration Strategy

The proposed registration-based classification system is illustrated in Fig. 2. This system is divided into two stages. The first stage is for the affine transform of a test pattern. After affine transform, the resultant pattern is expected to be close to its related reference in geometry; thus, the pattern can be used directly to match with reference pattern. The second stage is for evaluating similarities between the transformed test pattern and the reference templates.

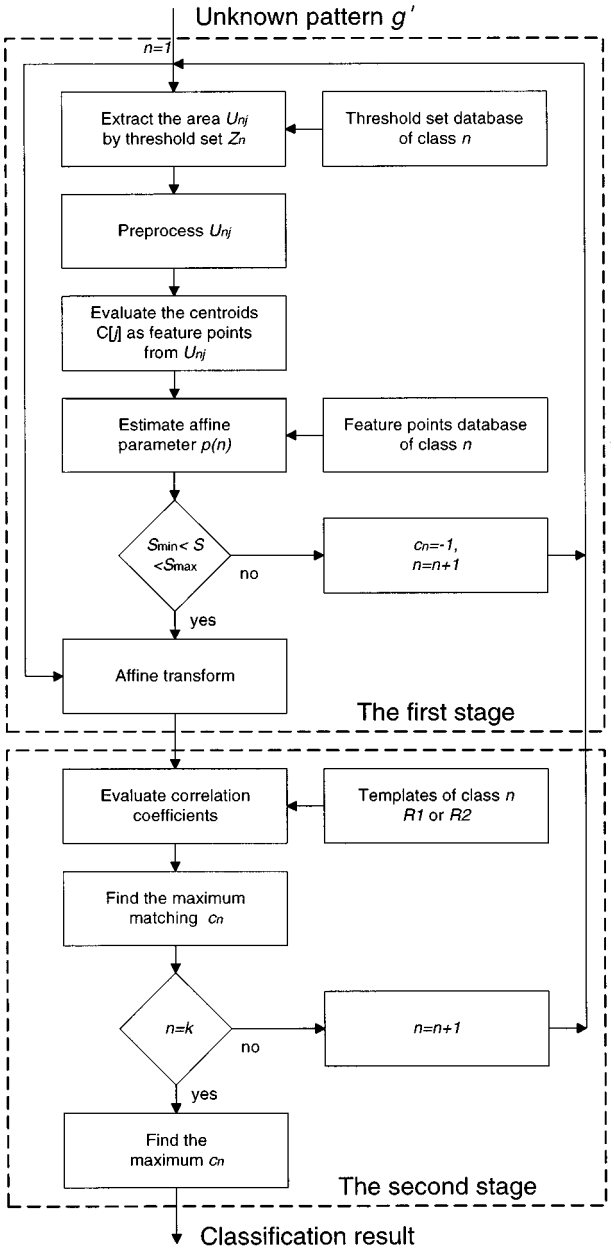


Fig. 2 Proposed system for object classification invariant to SRT.

The first stage can be considered as performing a preprocessing function for the second stage to reduce the geometric discrepancy between the test pattern and the reference pattern of the same class. Via the first stage, the conventional complicated scale and rotation invariant feature extraction is thus avoided; instead, a simple similarity metric can be employed. In another point of view, while compared with Caelli and Liu's method,[16] the amount of reference templates can be much reduced. In Caelli and Liu's method, reference templates must be prepared on every Δ_S scale in the whole possible scale range of a test pattern and on every Δ_Θ in the range (0 deg, 360 deg), where Δ_S and Δ_Θ are acquired from the analysis of objects' intrinsic invariance properties. Our proposed method merely provides the reference templates on every Δ_S scale

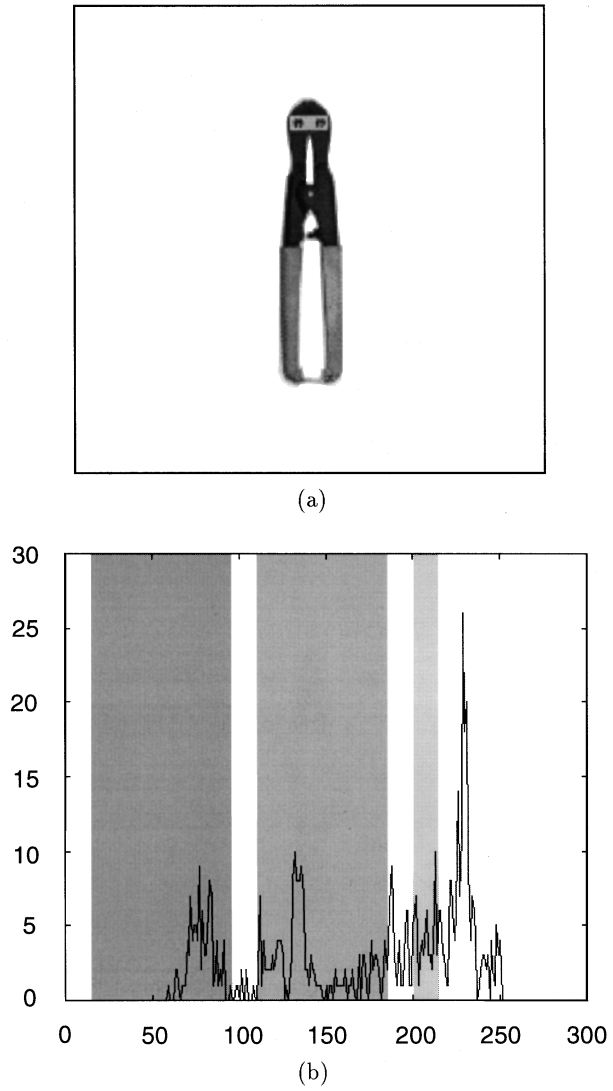


Fig. 3 (a) Reference pattern $M1$ and (b) histogram of pattern $M1$.

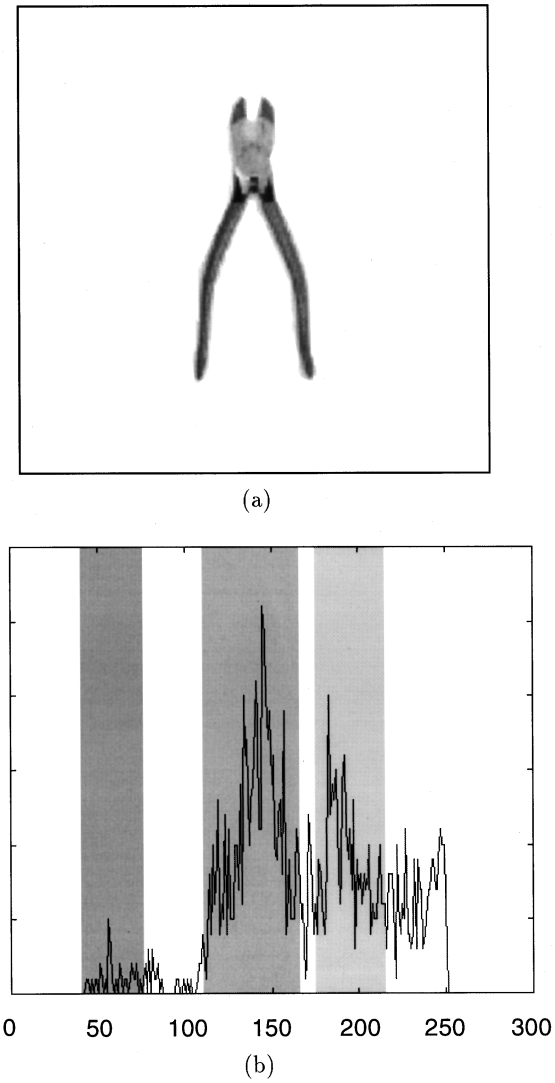


Fig. 4 (a) Reference pattern $M2$ and (b) histogram of pattern $M2$.

in the much reduced scale range and on every Δ_Θ in a small range of rotation angles.

It is assumed in this paper that (1) there are k sorts of objects for classification ($k=5$), which are named ''class i,'' $i=1,\cdots,k$; (2) there are k primary reference patterns: Mn for class n, $n=1,\cdots,k$; each Mn has scale=1.0 and rotation=0 deg, as shown in Figs. 3(a), 4(a), 5(a), 6(a), and 7(a), respectively; (3) each test pattern's scale is constrained in (0.5,1.5), while rotation is not restricted; (4) the test objects (e.g., industrial tools) are composed of several regions, each of which is related to a specific gray level distribution that can be distinguished from the gray level distribution of other regions and can be inspected in advance; (5) the light source is under control (e.g., in the factory) such that the gray level change of the test pattern is small; and (6) the gray level distribution of the background is not mixed up with that of any object's regions of interest.

3.1 *First Stage: Registration-Based Strategy*

If the test pattern g' belongs to class n, its affine-transformed pattern g will be similar to the reference pattern Mn. Hence, the correlation coefficient between g and Mn will approach 1. According to the principle described in Sec. 2, it is necessary to prepare the threshold set Z_n for class n. In this paper, the appropriate threshold intervals are determined by choosing the regions of interest of Mn. Figure 3(b) shows the image histogram of $M1$. First, it is required to analyze the gray level distribution of each region of interest. The three darkened rectangular areas in Fig. 3(b) are the selected threshold intervals. Each interval is related to a specific region of the primary reference $M1$. By the same principle, the threshold intervals of $M2$, $M3$, $M4$, and $M5$ are demonstrated in Figs. 4(b), 5(b), 6(b), and 7(b), respectively. After this procedure, the selected threshold sets are stored in a database to be used in the first stage. They are listed in Table 1. By using such threshold sets, the control points of the primary references can be acquired and stored in advance. Note that choosing just three regions for every class of object is simply for ease of explanation.

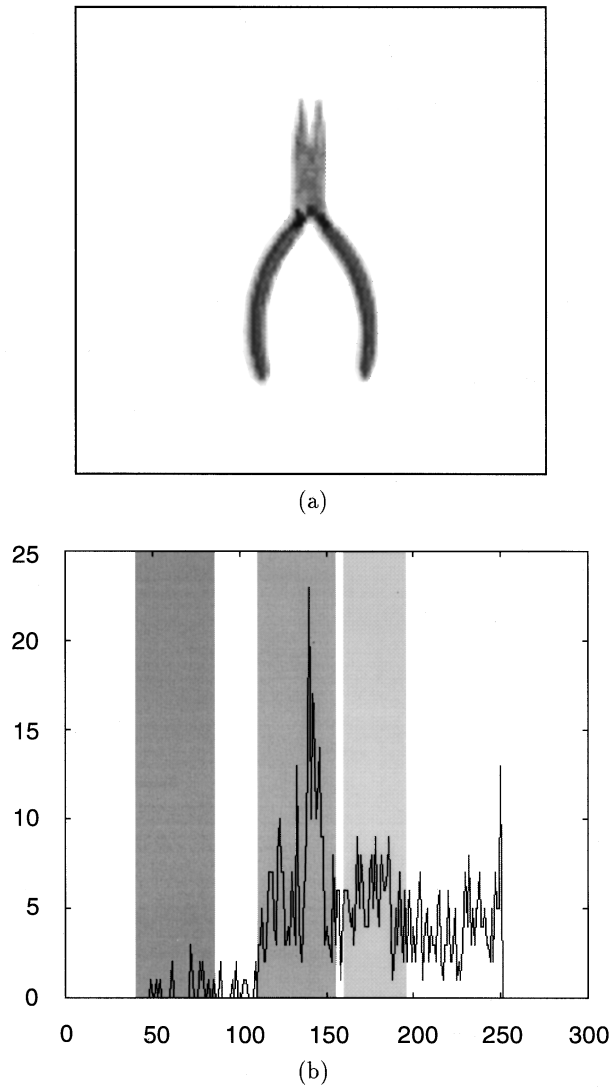


Fig. 5 (a) Reference pattern $M3$ and (b) histogram of pattern $M3$.

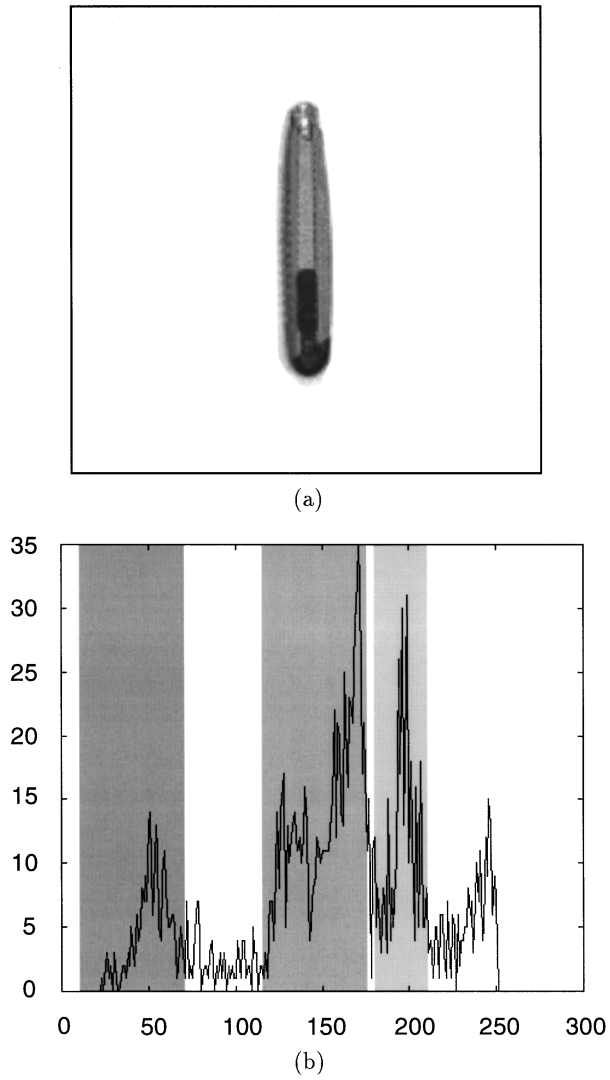


Fig. 6 (a) Reference pattern $M4$ and (b) histogram of pattern $M4$.

As shown in Fig. 2, as the unknown pattern g' enters, n is set to 1 and the procedures of the first stage can be described by the following steps.

Step 1. Extract each area U_{nj} by the j'th threshold interval of the threshold set Z_n (listed in Table 1).

Step 2. Preprocess U_{nj}. Seeing that a small amount of pixels, which have gray levels distributed in the threshold interval of U_{nj}, might be scattered around the background and some other regions and cause a small shift of the centroid of U_{nj}, preprocessing of U_{nj} is required. Let $N_8(x)$ denote the number of 8-neighbors of an element x in U_{nj}. If $N_8(x) \geqslant \rho$, the position x can be adopted in the region U_{nj}; ρ is a constant and is chosen as 3 in the experiments.

Step 3. Evaluate centroid $C[j]$ as a control point from U_{nj}.

$$C[j]=\begin{pmatrix} x[j] \\ y[j] \end{pmatrix}. \tag{10}$$

Step 4. Estimate affine parameters $p(n)=(s,\theta,t_x,t_y)$ by minimizing the following criterion,

$$J=\sum_{j=1}^{L}\{(x'[j]-x_n[j])^2+(y'[j]-y_n[j])^2, \tag{11}$$

where (x_n,y_n) are the control points of Mn and are stored in the database. Here L is the number of threshold intervals and (x',y') is the affine transform of (x,y) and is expressed as follows

$$\begin{aligned} x'[j]&=s(x[j]\cos(\theta)-y[j]\sin(\theta))+t_x, \\ y'[j]&=s(x[j]\sin(\theta)+y[j]\cos(\theta))+t_y, \end{aligned} \tag{12}$$

where s and θ are the scale and rotation-angle parameters to be estimated, respectively. (t_x,t_y) is the translation parameters to be estimated.

Step 5. Determine whether the estimated scale is in a proper range. Let $S=1/s$. If $S_{\min} < S < S_{\max}$, go ahead,

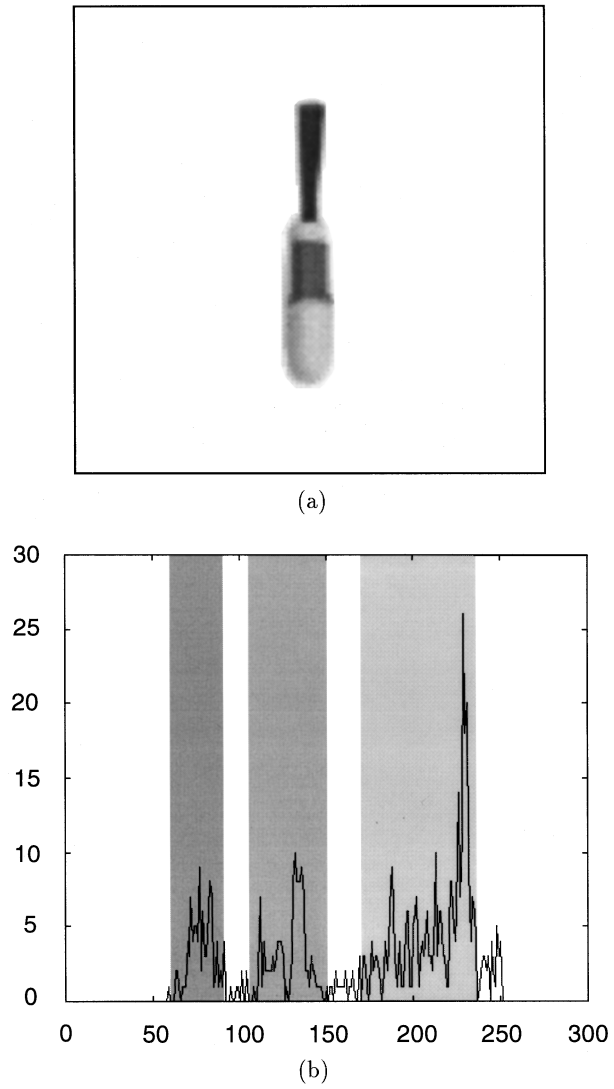


Fig. 7 (a) Reference pattern $M5$ and (b) histogram of pattern $M5$.

otherwise set $n=n+1$, $c_n=-1$ and go back to step 1. In the proposed system, $S_{\min}$ is set as 0.1 and $S_{\max}$ is set as 2.0 on account that the possible scale range of the test pattern is (0.5, 1.5). Hence, if S exceeds the range $(S_{\min}, S_{\max})$, the scale S is considered unreasonable; thus c_n is set as -1 and the affine transform can be omitted.

Step 6. Apply affine transform to g' with the estimated affine parameters $p(n)$.

Step 7. Go to the second stage to evaluate correlation coefficients.

Table 1 Threshold intervals of each class.

Z_1	Z_2	Z_3	Z_4	Z_5
(15,95)	(40,75)	(40,85)	(10,70)	(60,90)
(110,185)	(100,165)	(110,155)	(115,175)	(105,150)
(200,215)	(175,215)	(160,195)	(180,210)	(170,235)

The estimated scale (i.e., $S=1/s$) and rotation (i.e., $\Theta=-\theta$) are shown in Table 2 (step 2 is not used) where the test pattern is of class 1, and s and θ are estimated at the step 4. The true scale and rotation angle of the test patterns, i.e., $(\hat{S},\hat{\Theta})$, are given in the first two columns. There are 11 test patterns of each class used for classification. It is clear that the third and fourth columns have the closest scales and rotation angles to the true values. The estimated (S,Θ) by other threshold sets cannot give a reasonable estimations because the transformed version g does not match the Mn where n is not equal to 1. Even if (S,Θ) have values close to the true $(\hat{S},\hat{\Theta})$, the similarity between g and Mn $(n \neq 1)$ will not be high owing to the fact that the two patterns are not of the same class. Table 3 shows the average estimated errors. $(\tilde{S},\tilde{\Theta})$ are the average of $(|\hat{S}-S|,|\hat{\Theta}-\Theta|)$, where (S,Θ) are the results in case that g and Mn originate from the same class.

3.2 *Second Stage*

The second stage is employed to evaluate the correlation coefficients of the transformed test pattern g and the reference templates. The procedure of the second stage is described as follows.

Step 1. Evaluate the correlation coefficient of g and the (m)th template of class n as

$$c_{nm}=\Sigma_i\Sigma_j[g(i,j)-\overline{G}][\,f_{nm}(i,j)-\overline{F_{nm}}]'\{\Sigma_i\Sigma_j[g(i,j)-\overline{G}]^2\}^{1/2\prime}\{\Sigma_i\Sigma_j[\,f_{nm}(i,j)-\overline{F_{nm}}]^2\}^{1/2}, \quad (13)$$

where $\overline{X}$ denotes the mean of a pattern x.

Step 2. Find the maximum of c_{nm} and denote it by c_n, which determines the similarity between g' and class n.

Step 3. If $n=k$, go ahead, otherwise, go back to the first stage.

Step 4. Find the maximum of c_n to lead to the classification result.

Two types of template sets are proposed. The first type is the set $R1$,

$$R1=\{M1,M2,M3,M4,M5\}. \quad (14)$$

According to the experiments, if the test patterns are not degraded by noise, $R1$ is able to accomplish the classification task without failure. The second type is the set $R2$,

$$R2=\{f_{11},\cdots,f_{1N_1},\cdots,f_{n1},\cdots,f_{nNn}\}, \quad (15)$$

in which each class contains different number (N_n) of templates whose scale and rotation intervals are determined according to each class of objects intrinsic invariance property.[16] Figure 8(a) shows the correlation coefficients of pattern $f_n(S)$ and Mn. The threshold is chosen as 0.85. For example, $f_1(0.9)=0.86$; thus, the scale interval is chosen as $\Delta_S=2|1.0-0.9|=0.2$, i.e., class 1 must prepare reference templates on every Δ_S scale interval to guarantee the correlation coefficients not below the threshold 0.85 when the scale of the test pattern ranges from 0.5 to 1.5. By the

Table 2 Estimated scale S and the rotation angle Θ of the test patterns g' of class 1.

		M_1		M_2		M_3		M_4		M_5	
$\hat{S}$	$\hat{\Theta}$	S	Θ	S	Θ	S	Θ	S	Θ	S	Θ
0.5	30	0.44	31.17	0.81	31.35	0.89	31.25	0.37	203.96	0.39	26.72
0.6	60	0.50	59.73	0.93	59.32	1.02	58.85	0.42	236.50	0.44	60.14
0.7	90	0.74	90.52	1.45	91.40	1.65	90.75	0.40	262.60	0.37	84.67
0.8	120	0.67	120.24	1.25	120.16	1.38	119.83	0.55	295.14	0.58	118.37
0.9	150	0.72	150.39	1.29	149.57	1.40	149.37	0.71	326.75	0.78	150.25
1.0	180	1.00	179.99	1.94	180.59	2.19	179.99	0.60	353.50	0.58	176.24
1.1	210	0.94	209.61	1.76	209.22	1.94	208.64	0.74	27.07	0.78	210.90
1.2	240	0.99	240.18	1.80	239.54	1.96	239.24	0.91	56.62	0.99	240.16
1.3	270	1.19	269.70	2.26	269.85	2.52	269.25	0.81	85.44	0.82	268.89
1.4	300	1.14	299.82	2.08	298.92	2.26	298.51	1.07	117.48	1.16	301.25
1.5	330	1.35	329.53	2.52	329.07	2.78	328.46	1.07	147.44	1.12	331.38

same principle, the rotation interval of each class can be determined by Fig. 8(b). Therefore, by the analysis of the intrinsic invariance property from Fig. 8, the scale and rotation intervals Δ_S and Δ_Θ of reference templates of each class can be obtained and shown in Table 4.

4 Experimental Results

The test patterns are arranged with scale and rotation angle at (0.5,30 deg), (0.6,60 deg), (0.7,90 deg), (0.8,120 deg), (0.9,150 deg), (1.0,180 deg), (1.1,210 deg), (1.2,240 deg), (1.3,270 deg), (1.4,300 deg), (1.5,330 deg), respectively. Therefore, each class has eleven test patterns and the total number of test patterns is 55.

Table 5 shows the average errors of the estimated scales and rotations by using the first stage (step 2 is employed) in case that the test patterns and the reference patterns are of the same class. The total average rotation error is 0.33 deg, which is smaller than 0.71 deg, as shown in Table 3 where step 2 is omitted. Table 5 also shows the scale and rotation range of the transformed image g. It is clear that the scale range of g is around 1.0 and the rotation range is around

Table 3 Average errors of the estimated scales and the rotation angles. For example, $\tilde{S}=0.13$ is the average of the differences between the first and the third columns of Table 2.

Class	$\tilde{S}$	$\tilde{\Theta}$
1	0.13	0.38
2	0.03	0.68
3	0.03	2.07
4	0.02	0.26
5	0.01	0.18
Average	0.04	0.71

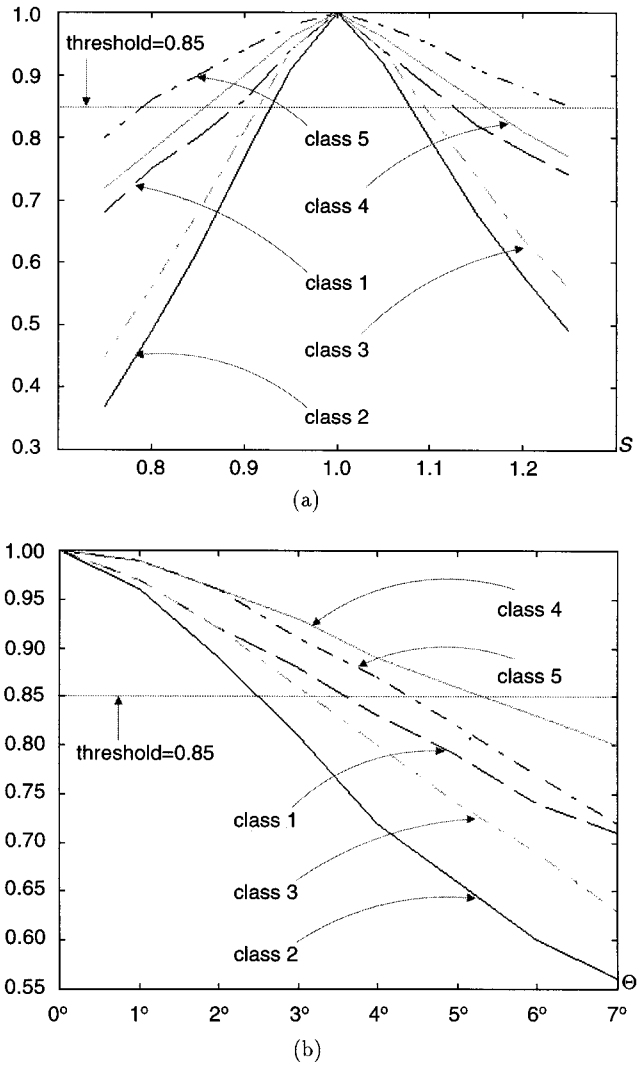


Fig. 8 (a) Correlation coefficients for scale variation and (b) correlation coefficients for rotation variation.

Table 4 Scale and the rotation intervals of each class's templates.

Class	Δ_S	Δ_Θ
1	0.20	7
2	0.15	5
3	0.20	6
4	0.30	10
5	0.40	9

0 deg, which is the goal of the first stage to reduce the geometric discrepancy between the test pattern and its related reference template.

If the test patterns are degraded by zero mean Gaussian noise with variance 100, the average scale and rotation errors will be larger, as shown in Table 6. The total average scale and rotation errors are 0.07 and 1.45 deg, respectively. The scale and rotation ranges are also enlarged. Especially for classes 1 and 5, the scale ranges are extended to 1.5. That is so because the added noise affects the scope of the thresholded regions to some extent; also, the selected threshold intervals are one of the reasons to affect the variations of the positions of the control points. Table 7 shows the results if the test patterns are degraded by zero mean Gaussian noise with variance 150. Figure 9 shows the test pattern of class 1, which is degraded by noise. The average scale error is 0.18, which is expected to bring about poor classification results. The classification results (using the reference set $R1$ defined in Sec. 3.2) for the test patterns, which are degraded by noise, are given in Table 8. The first column represents the true class number of the test patterns. For instance, the 11 test patterns of class 1 are classified as eight of class 1 and three of class 4. The correct ratio is 0.727.

Table 9 shows the average scale and rotation ranges by averaging the ranges as shown in Tables 5, 6 and 7. The scale and rotation ranges of g are empirical results, which are related to the selected threshold sets and the level of the noise degradation. However, it is reasonable to find a range that is large enough to design the reference set $R2$. According to Table 9, the reference set $R2$ is designed by using the scale and rotation intervals, as shown in Table 4. The desired $R2$ is given in Table 10. For example, class 1 consists of two templates with rotation angle -1 deg: f_{11} and f_{12}. The scales of f_{11} and f_{12} are 1.1 and 1.3, respectively. Hence, there are eight reference templates in $R2$. The classification results by using the reference set $R2$ are given in Table 11. The correct ratio is 0.927, which is much higher than those that resulted from $R1$, while the number of the reference templates does not increase too much. Thus, the analysis of object's intrinsic invariance property can facilitate the design of the reference set $R2$.

In other words, the registration-based strategy is proposed to reduce the geometric discrepancy between the test pattern and the related reference pattern. The estimated scale and rotation are very accurate, i.e., the test pattern can be transformed to be close to the related reference template in geometry. Two types of reference template sets can be utilized according to the requirements. $R1$ contains k primary reference templates and $R2$ consists of different number of templates for each class according to the analysis of the intrinsic invariance property of each class. Hence, $R1$ is sensitive to noise degradation. However, it is reasonable to realize that $R2$ can maintain a satisfactory correct classification ratio by designing the reference templates appropriately.

The proposed method registers the test pattern with the reference patterns before matching. It is different from Sheng and Arsenault's method[17] (a normalization approach), which normalizes both the test and the reference patterns. Although only SRT is discussed in this paper, the skewing problem can also be solved by the proposed registration-based method owing to the utilization of the affine transform. However, Sheng and Arsenault's method cannot deal with the skewing problem because the normalized angle (the direction of the principal axis) will change when skewing occurs. Regarding the issue of intensity invariance, Sheng and Arsenault normalize the intensity and the image scale together by low-order moments after orien-

Table 6 Average errors and ranges of the estimated scales and rotation angles. The test patterns are degraded by zero mean Gaussian noise with variance 100.

Class	$\tilde{S}$	Range	$\tilde{\Theta}$	Range
1	0.12	(1.00,1.56)	0.89	(−3.74,1.22)
2	0.05	(0.93,1.14)	0.65	(−0.65,1.97)
3	0.03	(0.94,1.05)	1.75	(−1.72,5.57)
4	0.03	(0.95,1.00)	0.25	(−0.36,0.50)
5	0.13	(0.94,1.54)	3.73	(−5.29,7.45)
Average	0.07		1.45	

Table 5 Average errors and ranges of the estimated scales and the rotation angles with preprocessing of each thresholded regions.

Class	$\tilde{S}$	Range	$\tilde{\Theta}$	Range
1	0.07	(0.99,1.25)	0.19	(−0.37,0.54)
2	0.02	(0.97,1.04)	0.16	(−0.30,0.25)
3	0.02	(0.96,1.00)	0.96	(−0.44,1.56)
4	0.03	(0.95,1.00)	0.10	(−0.26,0.10)
5	0.02	(0.91,1.03)	0.23	(−0.65,0.39)
Average	0.03		0.33	

Table 7 Average errors and ranges of the estimated scales and rotation angles. The test patterns are degraded by zero mean Gaussian noise with variance 150.

Class	$\tilde{S}$	Range	$\tilde{\Theta}$	Range
1	0.20	(1.03,1.43)	2.23	(−5.47, 4.10)
2	0.18	(1.03,1.39)	5.32	(−33.80,4.25)
3	0.16	(1.11,1.28)	2.69	(−0.43, 6.41)
4	0.03	(0.99,1.17)	1.40	(−0.79, 6.53)
5	0.35	(1.44,1.90)	2.11	(−3.69, 3.50)
Average	0.18		2.75	

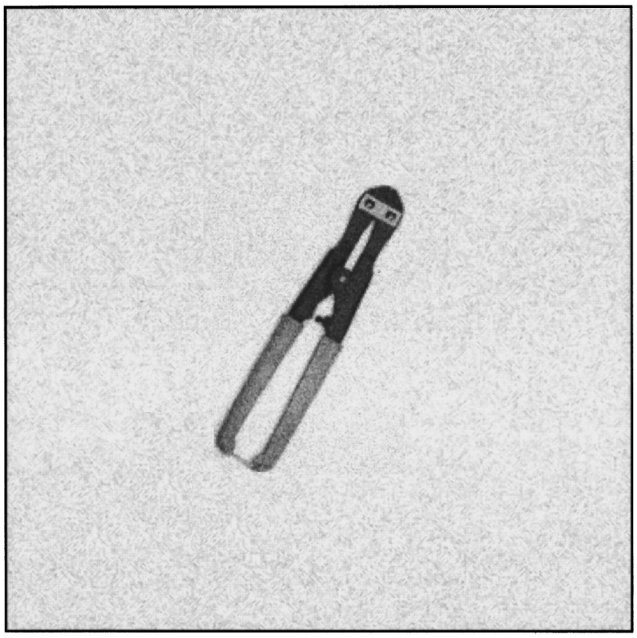


Fig. 9 Pattern of class 1 degraded by zero mean Gaussian noise with variance 150 (scale=1.5, angle=330 deg).

Table 8 Classification result using the reference set $R1$. The test patterns are degraded by zero mean Gaussian noise with variance 150. The correct ratio is 0.727.

	Class 1	Class 2	Class 3	Class 4	Class 5
1	8	0	0	3	0
2	0	10	0	1	0
3	0	0	11	0	0
4	0	0	0	11	0
5	0	0	0	11	0

Table 9 Scale and rotation ranges of the templates for each class.

Class	Scale Range	Rotation Range
1	(1.01,1.41)	(−3.19,1.95)
2	(0.97,1.19)	(−11.58,2.16)
3	(1.00,1.11)	(−0.87,4.52)
4	(0.96,1.05)	(−0.47,2.38)
5	(1.10,1.49)	(−3.21,3.78)

Table 10 Scales and rotations of the reference templates in $R2$.

Class	Scales	Rotations	Template Number
1	1.10,1.30	−1	2
2	1.10	−7,−2	2
3	1.05	2	1
4	1.00	1	1
5	1.10, 1.50	0	2

Table 11 Classification result using the reference set $R2$. The test pattern are degraded by zero mean Gaussian noise with variance =150. The correct ratio is 0.927.

	Class 1	Class 2	Class 3	Class 4	Class 5
1	10	0	0	1	0
2	0	10	0	1	0
3	0	1	10	0	0
4	0	0	0	11	0
5	0	0	0	1	10

tation normalization. The proposed method also requires an intensity change relation for the threshold sets. This other work will be accomplished in the near future.

At the end of this section, the computational cost of the proposed method is compared with the cost of the log-polar mapping method.[14] The matching procedure by log-polar mapping is given in Fig. 10. It is clear that the computational cost of the proposed method is much smaller than the cost of the log-polar mapping method. The main reason lies in the fact that the second stage of the proposed method computes only ''one'' value of the similarity metric because the translation between g (dimension=$N\times N$) and the related reference pattern is very small. Nonetheless, the usual similarity metric must compute the whole $N\times N$ values by shifting the template to search the best match.

5 Optical Implementation

The conventional matched filter (MF, i.e., VanderLugt correlator) cannot handle the geometric deformation. By using the proposed method, a bank of filters can be designed and used in the optical system. The filter design can be accomplished off-line. For example, the phase-only filter (POF) can be chosen as the similarity metric because the POF can eliminate light attenuation in the Fourier plane. It has the maximum optical transmittance. Hence, it can sharpen the correlation signal and result in good discrimination ability. However, it is noted[18] that POF is even more sensitive to object rotation than the MF, i.e., the number of filters for rotation variation will increase. Other filter variations[19] are POC, binary phase only filter (BPOF), and binary phase only correlator (BPOC), etc. An optical correlator can be designed by a real-time spatial light modulator (SLM) to sequentially introduce the multiple filters. Figure 11 depicts a schematic diagram of simple optical correlator. The filter bank contains all of the designed filters for each class. The correlation output pattern can be detected by a CCD camera. Since only a negligible amount of lateral translation

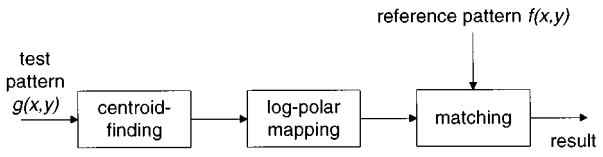


Fig. 10 Matching method of the log-polar mapping.

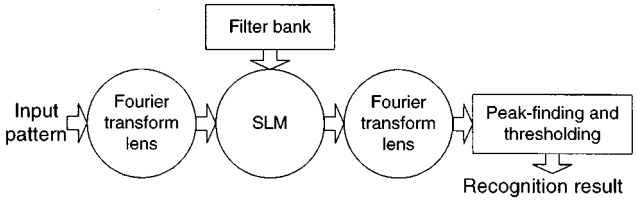


Fig. 11 Optical system.

still exists between the reference pattern and the transformed test pattern, peak-finding is not necessary. After thresholding, the recognition is achieved.

6 Conclusions

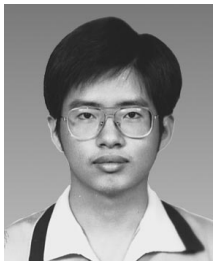
A new approach for SRT invariant pattern recognition has been proposed in this paper. The method is based on a registration strategy so that the geometric discrepancy between the test pattern and the related reference template can be much reduced before matching. The centroids of regions of interest are extracted as control points and automatically paired. The pairing procedure is not required, in contrast to the conventional registration method. By means of the proposed system, the extraction of scale and rotation invariant features is also not required and a simple similarity metric becomes satisfactory. In another point of view, the number of reference templates is much reduced compared with Caelli and Liu's method. Therefore, the proposed system is novel and feasible.

Acknowledgments

The paper was supported by National Science Council of the Republic of China under grant No. NSC 85-2612-E-007-002.

References

1. J. Flusser, ''Object matching by means of matching likelihood coefficients,'' *Pattern Recog. Letters* **16**, 893–900 (1995).
2. Y. Li, ''Reforming the theory of invariant moments for pattern recognition,'' *Pattern Recog.* **25**(7), 723–730 (1992).
3. A. Goshtasby, ''Template matching in rotated images,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-7**(3), 338–344 (1985).
4. J. Flusser and T. Suk, ''A moment-based approach to registration of images with affine geometric distortion,'' *IEEE Trans. Geosci. Remote Sens.* **32**(2), 382–387 (1994).
5. Y. Sheng and L. Shen, ''Orthogonal Fourier-Mellin moments for invariant pattern recognition,'' *J. Opt. Soc. Am. A* **11**(6), 1748–1757 (1994).
6. C.-H. Teh and R. T. Chin, ''On image analysis by the methods of moments,'' *IEEE Trans. Pattern Anal. Mach. Intell.* **10**(4), 496–513 (1988).
7. Y. Sheng and H. H. Arsenault, ''Object detection from a real scene using the correlation peak coordinates of multiple circular harmonic filters,'' *Appl. Opt.* **28**(2), 245–249 (1989).
8. M. T. Manry, ''Compound filter using the circular harmonic expansion,'' *Appl. Opt.* **26**(17), 3622–3627 (1987).
9. D. Mendlovic, E. Marom, and N. Konforti, ''Shift and scale invariant pattern recognition using Mellin radial harmonics,'' *Opt. Commun.* **67**(3), 172–176 (1988).
10. M. Kreutz, B. Völpel, and H. Janβen, ''Scale-invariant image recognition based on higher-order autocorrelation features,'' *Pattern Recog.* **29**(1), 19–26 (1996).
11. Y. Sheng, C. Lejeune, and H. H. Arsenault, ''Frequency-domain Fourier-Mellin descriptors for invariant pattern recognition,'' *Opt. Eng.* **27**(5), 354–357 (1988).
12. M. Fang and G. Hausler, ''Class of transforms invariant under shift, rotation, and scaling,'' *Appl. Opt.* **29**(5), 704–708 (1990).
13. S.-S. Wang, P.-C. Chen, and W.-G. Lin, ''Invariant pattern recognition by moment Fourier descriptor,'' *Pattern Recog.* **27**(12), 1735–1742 (1994).
14. W.-S. Shao and Y.-S. Chen, ''Pattern analysis on shift, rotation, and scaling,'' *Electron. Lett.* **28**(25), 2271–2272 (1992).
15. L. G. Brown, ''A survey of image registration techniques,'' *ACM Comput. Surv.* **24**(4), 325–376 (1992).
16. T. M. Caelli and Z.-Q. Liu, ''On the minimum number of templates required for shift, rotation, and size invariant pattern recognition,'' *Pattern Recog.* **21**(3), 205–216 (1988).
17. Y. Sheng and H. H. Arsenault, ''Noisy-image normalization using low-order radial moments of circular-harmonic functions,'' *J. Opt. Soc. Am. A* **4**(7), 1176–1184 (1987).
18. P. D. Gianino and J. L. Horner, ''Additional properties of the phase-only correlation filter,'' *Opt. Eng.* **23**(6), 695–697 (1984).
19. T. Nomura, K. Itoh, K. Matsuoka, and Y. Ichioka, ''Binary Fourier phase-only correlation,'' *Opt. Lett.* **15** (14), 810–811 (1990).



Wen-Hao Wang received a BS in electronics engineering from Fan Chia University, Taichung, Taiwan, in 1987 and an MS in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1992. He is currently a PhD student in the Electrical Engineering Department of the National Tsing Hua University. His research interests include pattern recognition, image tracking, image processing, and image registration.

Yung-Chang Chen received BS and MS degrees in electrical engineering from the National Taiwan University, Taipei, in 1968 and 1970, respectively, and a DrIng degree from the Berlin Technical University, Berlin, Germany, in 1978. Since 1978, he has been with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, where he is currently a professor. His research interests are image processing, pattern recognition, signal processing, medical image processing and video coding.