PERGAMON

Pattern Recognition 32 (1999) 151–165

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

Discriminative wavelet shape descriptors for recognition of 2-D patterns

Dinggang Shen[1], Horace H.S. Ip*

Image Computing Group, Department of Computer Science, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

Received 11 February 1997; received in revised form 3 August 1998

Abstract

In this paper, we present a set of wavelet moment invariants, together with a discriminative feature selection method, for the classification of seemingly similar objects with subtle differences. These invariant features are selected automatically based on the discrimination measures defined for the invariant features. Using a minimum-distance classifier, our wavelet moment invariants achieved the highest classification rate for all four different sets tested, compared with Zernike's moment invariants and Li's moment invariants. For a test set consisting of 26 upper cased English letters, wavelet moment invariants could obtain 100% classification rate when applied to 26×30 randomly generated noisy and scaled letters, whereas Zernike's moment invariants and Li's moment invariants obtained only 98.7 and 75.3%, respectively. The theoretical and experimental analyses in this paper prove that the proposed method has the ability to classify many types of image objects, and is particularly suitable for classifying seemingly similar objects with subtle differences.

Keywords: Invariant feature; Wavelet transform; Zernike's moments; Hu's moments; Li's moments; Rotation invariant; Feature selection; nearest-neighbor classifier; Character classification; Document analysis and recognition

1. Introduction

The use of moment invariants as features for identification and inspection of 2D shape has received much attention [1]. Since Hu [2] presented the first paper on the use of image moments for 2D pattern recognition, moment-based techniques have found wide applications [3, 4]. Resis [5] revised some of the theoretical proofs in [2]. By comparing the relationship between the theory of moment invariants and the Fourier–Mellin transformation, Li [6] not only re-formulated Hu's seven moment invariants but also introduced some new high-order moment invariants. However, Hu's moment invariants have some drawbacks. One drawback is their dramatic increase in complexity with increasing order. The second is that they are not derived from a family of orthogonal functions, and so contain much redundant information about an object's shape.

Teague [7] suggested the use of orthogonal moments based on the theory of orthogonal polynomials to overcome the problems associated with the regular moments. Various types of image moments, including geometrical moments, Legendre moments, Zernike moments, pseudo-Zernike moments, Fourier–Mellin moments, and

*Corresponding author. Tel.: 0085227888641; Fax: 0085227888614; E-mail: cship@cityll.eduhk

[1]Current address: School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore.

0031-3203/99/$ – See front matter
PII: S0031-3203(98)00137-X

complex moments, have been evaluated in term of noise sensitivity, information redundancy, and capability of image description [8]. The result is that Zernike moments have the best overall performance. Khotanzad and Hong [9] have also compared Hu's moment invariants and the magnitudes of Zernike moments in rotational invariant recognition of characters and shapes. Their results clearly show that Zernike moments are superior to Hu's moments. Accordingly, the orthogonal moment features have been used for the recognition of handwriting Arabic numerals [10].

The geometrical moments of an image are integrals of the image function over space, and the image can be uniquely determined by its geometrical moments of all orders. Low-order moments are relatively more stable than high-order moments. However, low-order moments can be used to differentiate between images of real objects *only if* their shapes are *significantly* different. Since these moments are designed to capture global information about the image, they are not suitable for classifying similar objects when corrupted by a significant amount of random noise. In fact, the geometrical moments are sensitive to digitization error, minor shape deformations, camera non-linearity, and non-ideal position of camera. We may understand this conclusion by observing Fig. 1 which demonstrates the unstable property of Hu's moment invariants for various shapes of "1" and "|". In Fig. 1, only two second-order moment invariant features (ϕ_1, ϕ_2) are given, as they are regarded as the best discriminative features [11]. Attempting to overcome this problem, Sluzek [11] presented a method of using the second-order moment invariants to create more position-invariant descriptors and to improve the resolution of these descriptors. First, the object of interest was partially occluded by circles located in its center. Then, this object is represented by a family of different shapes, and the moment invariants of order 2 of all these shapes are regarded as shape descriptors of the object. We will show in Section 3 that the applicability of this method is limited.

In a shape recognition system, typically a set of numerical features are extracted from an image. The selection of discriminative features is a crucial step in the process, since the next stage sees only these features and acts upon them. In general, discriminative features must satisfy small intraclass variance and large interclass separation [12].

The main contributions of this paper are a set of wavelet moment invariants presented for capturing global and local information from the objects of interest, and a method of selecting discriminative features based on a set of discrimination measures defined for the features. Even for two seemingly similar but different objects, such as "1" and "|", our wavelet moments invariants still succeed in extracting discriminative features for separating these two classes of objects, whereas, Li's moment invariants and Zernike's moment invariants failed. Furthermore, a minimum-distance classifier is applied to four test sets, the set consisting of "1" and "|" (Fig. 4), the set consisting of two artificially generated objects (Fig. 8), the set consisting of 26 upper cased English letters (Fig. 12), and the set consisting of two similar wrenches (Fig. 15). We compared performances of the classifications with respect to different number of wavelet moment invariants, Zernike's moment invariants and Li's moment invariants (extended from Hu's moment invariants) [6]. For example, for the first test set containing "1" and "|", 100% recognition rate was obtained using our method *on the basis of only one feature*. However, for Zernike's moment invariants, the highest

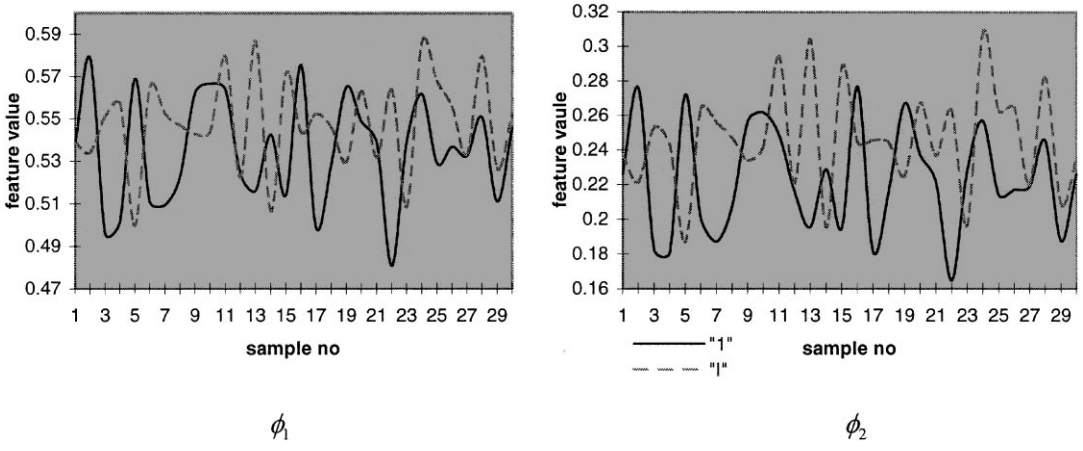


Fig. 1. The instability of Hu's first two moments (ϕ_1, ϕ_2) for various shapes of "1" and "|" (Since two feature values of the two shapes overlap significantly, it is impossible to discriminate the two shapes based on only these two features.)

recognition rate is 98% on the basis of at least 11 features. While for Li's moment invariants, the highest recognition rate is 83% when 29 features are used.

We will give a generalized expression for rotation invariant feature extraction in the following section, in order to show that wavelet moment invariants have much superiority in terms of discrimination powers and their low sensitivity to noise. In Section 3, we will apply wavelet transforms to feature extraction, and a set of wavelet moment invariants will also be given there. In Section 4, the discriminative feature selection method will be described. Section 5 presents some experimental results which show that wavelet moment invariants are better than Zernike's moment invariants and Li's moment invariants for object discrimination. This paper concludes in Section 6.

2. Generalized expression for obtaining moment-based rotation invariant feature

In this section, we will give a generalized expression for obtaining the rotation invariant features. After comparing several well-known moments which are included in the generalized expression, we will show that wavelet moment invariants have much superiority over other moments. In this paper, translation invariant and scaling invariant are achieved by using a normalization based on regular moments. Rotation invariant is achieved by extracting wavelet moment invariants, Zernike's moment invariants, or Li's moment invariants, since they are intrinsically rotation-invariant.

2.1. Translation and scaling normalization

Let $f(x, y)$ represent a 2-D binary image object in the (x, y)-coordinate, and its corresponding form in the polar coordinate be $f(r,\theta)$. The relationship between $f(x, y)$ and $f(r, \theta)$ is given as follows:

$$x = r\cos(\theta),$$
$$y = r\sin(\theta).$$

In a shape recognition system, the objects of interest are often represented by a set of numerical features with a goal to remove redundancy in the data and to reduce its dimensions. Moreover, the extracted features are expected to be invariant under translation, scaling and rotation to suit for different viewing directions. Translation invariant and scaling invariant can be achieved using the regular moments as follows.

The definition for a regular moment m_{pq} is

$$m_{pq} = \int\int x^p y^q f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

Since the center of the shape is invariant to translation, rotation and scaling, the method of solving the translation problem is to locate the centroid of the shape. The coordinates of the center of the shape are

$$X_0 = \frac{m_{10}}{m_{00}},\ Y_0 = \frac{m_{01}}{m_{00}}$$

In Hu's moments, scaling invariant was obtained by normalizing the central moments by m_{00}. To simplify discussion, we extract the scaling factor based on Hu's idea. The scaling factor of the present object size, compared with the expected size, is

$$\alpha = \sqrt{\frac{m_{00}}{AREA}},$$

where $AREA$ is a constant, corresponding to the expected size of the object. This way, we can obtain the translation and scaling normalized shape by changing the coordinates according to the following transformation:

$$\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} (x - X_0)/\alpha \\ (y - Y_0)/\alpha \end{pmatrix}.$$

Notice that if $f(x, y)$ is a binary image object free of noise, then the object area is equal to m_{00} and hence α will be equal to 1. For the rest of this paper, we let $f(x, y)$ and $f(r, \theta)$ represent the translation and scaling normalized image object.

2.2. Generalized expression

To get rotation invariant moments, typically the following generalized expression is used:

$$F_{pq} = \int\int f(r, \theta) g_p(r) \mathrm{e}^{\mathrm{j}q\theta} r\,\mathrm{d}r\,\mathrm{d}\theta, \tag{1}$$

where F_{pq} is the pq-order moment, $g_p(r)$ is a function of radial variable r, and p and q are integer parameters. It is easy to prove that the value of $\|F_{pq}\|$ is rotation invariant and the combined moments, such as $F_{p_1q} \cdot F^*_{p_2q}$, are also rotation invariant, where $\|x\| = \sqrt{x \cdot x^*}$ and symbol * denotes conjugate of complex number. The proof of the rotation invariant property of $\|F_{pq}\|$ can be briefly given as follows. If an image object $f(r, \theta)$ is rotated by an angle of β, its corresponding moment will become $F^{Rotated}_{pq} = F_{pq}\mathrm{e}^{\mathrm{j}q\beta}$. Since $\|F^{Rotated}_{pq}\| = \sqrt{F^{Rotated}_{pq}(F^{Rotated}_{pq})^*} = \|F_{pq}\|$, the rotation invariant property of $\|F_{pq}\|$ is thus proven. The other combined moments can be obtained using methods described in [6, 10], such as $F_{p_1q} \cdot ((F_{p_2}1)^q)^*$. Notice that the definition of the combined moments expressed in terms of $F_{p_1q} \cdot F^*_{p_2q}$ has more advantages than that of the combined moments in [6,10], i.e. $F_{p_1q} \cdot ((F_{p_2}1)^q)^*$. One advantage is that the magnitudes of

all the combined moments $F_{p_1q} \cdot F^*_{p_2q}$ are comparable, whereas the magnitudes of all the combined moments $F_{p_1q} \cdot ((F_{p_2 1})^q)^*$ may not be comparable since the magnitude of $F_{p_1q} \cdot ((F_{p_2 1})^q)^*$ changes with q. Another advantage is that, for the rotationally symmetric shapes with fold number larger than 1, the values of all the combined moments $F_{p_1q} \cdot ((F_{p_2 1})^q)^*$ are zeroes, whereas the values of all the combined moments $F_{p_1q} \cdot F^*_{p_2q}$ may be non-zeroes for any q larger than the fold number [15].

In order to reduce the problem of feature extraction from a 2D image object to that from a 1D sequence, expression (1) is rewritten as follows:

$$F_{pq} = \int S_q(r) \cdot g_p(r) r \, \mathrm{d}r \tag{2}$$

where $S_q(r) = \int f(r, \theta) \mathrm{e}^{\mathrm{j}q\theta} \, \mathrm{d}\theta$. Note that $S_q(r)$ is now a 1D sequence of variable r. It is important to note from Eq. (2) that if $g_p(r)$ is defined on the whole domain of variable r, then F_{pq} is a *global feature*, on the other hand, if the function of $g_p(r)$ is locally defined, then F_{pq} may be seen as a *local feature*.

Based on expression (2), we can easily show that the original definitions for Hu's moments, Li's moments and Zernike moments are special cases of expression (1) and the extracted features are global features.

(i) When setting $g_p(r) = r^p$ and some constraints on p and q, then Hu's moments [2] and Li's moments [6] can be obtained.
(ii) When setting $g_p(r)$ to be the following orthogonal polynomials,

$$g_p^{Zernike}(r) = \sum_{s=0}^{(p-|q|)/2} (-1)^s \times \frac{(p-s)!}{s! \left(\frac{p+|q|}{2} - s\right)! \left(\frac{p-|q|}{2} - s\right)!} r^{p/2-s}$$

and some constraints on p and q, then we can obtain Zernike's moment invariants $\| F_{pq}^{Zernike} \|$, which are the magnitude of Zernike moments. That is

$$\| F_{pq}^{Zernike} \| = \left\| \int S_q(r) \cdot g_p^{Zernike}(r) r \, \mathrm{d}r \right\| \tag{3}$$

Furthermore, Hu's moment invariants, Li's moment invariants and Zernike's moment invariants are calculated on the global image space. Fig. 1 indicates that the values of Hu's moment invariants are sensitive to noise. This observation is also true for Li's moment invariants and Zernike's moment invariants. Thus it is not easy to correctly classify similar image objects with subtle differences based on such global moment invariants. We further elaborate this point as follows. Suppose we have two similar objects and their corresponding moments are $\| F_{pq} \|$ and $\| F'_{pq} \|$, respectively. Since they are similar, then in the absence of noise they may be related as:

$$\| F_{pq} \| = \| F'_{pq} \| + \Delta_{pq},$$

where Δ_{pq} is a small deviation. If there exists any noise coming from digitization or randomly-added Gaussian noise, then their relationship may include the noise term:

$$\| F_{pq} \| = \| F'_{pq} \| + \Delta_{pq} + noise_{pq}.$$

Since the two images are similar, Δ_{pq} is always very small for Hu's moment invariants, Li's moment invariants and Zernike's moment invariants. However, the magnitude of $noise_{pq}$, which is integrated from the entire image space, may be greater than the magnitude of Δ_{pq}. Consequently, the values of $\| F_{pq} \|$ and $\| F'_{pq} \|$ may overlap and oscillate from sample to sample, as shown in Fig. 1.

If $g_p(r)$ is a locally defined function of the radial variable, then the features, $\| F_{pq} \|$ and $\| F'_{pq} \|$, will be extracted from a local area of the image space and the difference, Δ_{pq}, between the corresponding features of different shapes will be larger. Similarly, the noise effect is also computed from the local area, the value of $noise_{pq}$ will be smaller. It is likely that the values of $\| F_{pq} \|$ and $\| F'_{pq} \|$ will not be overlapping, such as Figs. 6a1 and b1. This is the key idea that underlies the development presented in this paper and that leads us to develop a rank order procedure of extracting suitable discriminative features based on the wavelet transform.

3. Wavelet moment invariants

Sluzek [11] described a prototype object by means of a family of shapes, which are created through occluding the object by circles (of different radius) located at the object's center. The Hu's moment invariants of such shapes, which are functions of the radius of an occluding circle, were used as features for quality detection of industrial parts. Fig. 2 illustrates the effect of occluding the object by a circle, where $Th(r, c)$ is an occluding function and c represents the radius of the occluding circle. If the function $g_p(r)$ is defined as $r^p \cdot Th(r, c)$ and the relationship between p and q is constrained as in Hu's moments, then the combined features are identical to the features suggested by Sluzek [11]. However, $Th(r, c)$ is a step function, which is not a localized window-liked function. It helps to extract the object information in the region of $r > c$, but not within $r < c$. Hence, this method may not be suitable for all types of image objects.

In this section, we will introduce the application of wavelet basis functions for feature extraction. Wavelet transform is a method for accomplishing localized analysis [13, 14, 16–18]. Different from the traditional short-time Fourier transform, wavelet transform is capable to

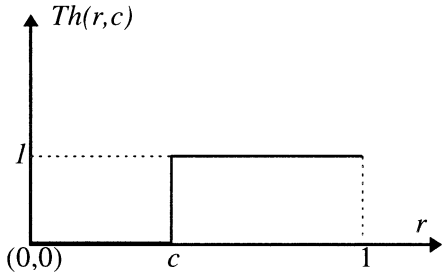


Fig. 2. The meaning for occluding the object by circles.

provide both time and frequency localization. The characteristic of wavelet transform is particularly suited to extracting local discriminative features.

In this paper, we treat $\{g_p(r)\}$ in Eq. (2) as wavelet basis functions, and consider the family

$$\psi^{a,b}(r) = \frac{1}{\sqrt{a}}\,\psi\left(\frac{r-b}{a}\right),$$

where a ($a \in \Re_+$) is a dilation parameter and b ($b \in \Re$) is a shifting parameter. From now on, the basis functions $\{g_p(r)\}$ are replaced by wavelet basis functions $\{\psi^{a,b}(r)\}$. We consider using the cubic B-spline wavelets [17,18] which are optimally localized in space-frequency and are close to the forms of Li's (or Zernike's) polynomial moments. The mother wavelet $\psi(r)$ of the cubic B-spline in Gaussian approximation form [17] is

$$\psi(r) = \frac{4a^{n+1}}{\sqrt{2\pi(n+1)}}\,\sigma_w \cos(2\pi f_0(2r-1)) \times \exp\left(-\frac{(2r-1)^2}{2\sigma_w^2(n+1)}\right),$$

where $n = 3$, $a = 0.697066$, $f_0 = 0.409177$ and $\sigma_w^2 = 0.561145$. This function is plotted in Fig. 3. The values of parameters a and b are usually discrete. The discretization of the dilation parameter is done by choosing $a = a_0^m$, where m is an integer and $a_0 > 1$ or $a_0 < 1$. b is discretized by taking the integer (positive and negative) multiples of $b_0 a_0^m$, where $b_0 > 0$ is appropriately chosen so that $\psi(r-b)/a)$ "covers" the whole domain at different values of m. Thus,

$a = a_0^m$, m is integer,

$b = nb_0a_0^m$, n is integer.

Since the image size is always restricted in a domain $\{r \leqslant 1\}$, let both parameters a_0 and b_0 be set to 0.5, and the domains for m and n be restricted as follows:

$$\begin{cases} a = 0.5^m, & m = 0, 1, 2, 3, \\ b = 0.5 \cdot n \cdot 0.5^m, & n = 0, 1, \ldots, 2^{m+1}. \end{cases}$$

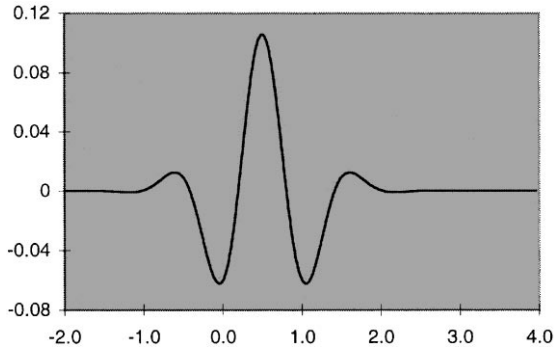


Fig. 3. The cubic B-spline mother wavelets.

Then the wavelet defined along a radial axis in any orientation is denoted by

$$\psi_{m,n}(r) = 2^{m/2}\psi(2^m r - 0.5n)$$

Let this function sweep across in all angular rotations in the moment computation, it will be able to extract either global or local information depending on the values of m and n. Notice that setting b_0 to 0.5 leads to oversampling. However, we will select wavelet-based features for object classification, *not for image reconstruction*; Any redundant and sensitive features can be screened out after feature selection.

Now let us introduce a set of wavelet moment invariants for classifying objects, they are defined as follows:

$$\| F_{m,n,q}^{wavelet} \| = \left\| \int S_q(r) \cdot \psi_{m,n}(r) r \, \mathrm{d}r \right\|, \tag{4}$$

where $\psi_{m,n}(r)$ replaces $g_p(r)$ in (2), $m = 0, 1, 2, 3$, $n = 0, 1, \ldots, 2^{m+1}$, and $q = 0, 1, 2, 3$. The above definition indicates that $F_{m,n,q}^{wavelet}$ is actually a wavelet transform of $S_q(r)r$. Since $F_{m,n,q}^{wavelet}$ is close to the generalization of the *first* moment of $S_q(r)$, $F_{m,n,q}^{wavelet}$ can also be regarded as the *first* moment of $S_q(r)$ at the mth scale level with shift index n. The last variable r in $S_q(r)r$ is derived from the transformation from the xy-coordinate to the *polar* coordinate (r, θ). For a fixed r, $S_q(r) = \int f(r, \theta) \mathrm{e}^{\mathrm{j}q\theta} \mathrm{d}\theta$ represents the qth frequency feature of the image object $f(r, \theta)$ in the phase domain $\{0 \leqslant \theta \leqslant 2\pi\}$. So $S_q(r)r$ expresses the feature distribution of the object $f(r, \theta)$ in the radial domain $\{0 \leqslant r \leqslant 1\}$. Furthermore, using different scale index m and shift index n, we ensure that the wavelets $\{\psi_{m,n}(r)\}$ cover the whole radial domain $\{0 \leqslant r \leqslant 1\}$ and thus the wavelet moment invariants, $\| F_{m,n,q}^{wavelet} \|$, can provide features of the object $f(r, \theta)$ at different scale levels. Notice that wavelet moment invariants are invariant to rotation of object. For its proof, please refer to the proof given for the generalized rotation invariant expression in the beginning of Section 2.2.

We will compare the quality of these wavelet moment invariants with that of Zernike's moment invariants and

also that of Li's moment invariants. Zernike's moment invariants, $\| F_{pq}^{Zernike} \|$, have been given in Eq. (3). Here, we constrain the orders p and q such that: $p \leqslant 11$, $(p-q)$ even, $0 \leqslant q \leqslant p$. Since the higher-order moments are too sensitive to noise, they cannot be used as the discriminative features of an object. The total number of Zernike's moment invariants that we used is 42. Li's moment invariants, containing Hu's seven moment invariants, are expressed as M_p [6]. We also used 42 of them $\{M_p, p = 1, 2, \ldots, 42\}$ in this paper for comparison.

4. Discriminative feature selection and classification

It is well known that the selection of discriminative features is a crucial step in any shape recognition system, since the next stage sees only these features and acts upon them. Invariant features which capture the global characteristics of an object are often sensitive to noise, thus it is important to design a rank order procedure of selecting discriminative features which have small intraclass variance and large interclass separation. Many feature selection methods have been developed [20], and most of the search methods except for the *Branch and Bound algorithm* are suboptimal. The *Branch and Bound algorithm* implicitly searches all of the combinations and guarantees a globally optimum feature set. In practical applications, one tends to employ more computationally efficient methods which are suboptimal. The suboptimal methods include Sequential Forward Selection (SFS), Generalized Sequential Forward Selection (GSFS), Sequential Backward Selection (SBS), and Generalized Sequential Backward Selection (GSBS). The feature selection method used in this paper is also suboptimal. We use standard variance-based feature discrimination techniques, such as between-to within-class variance ratio [18], for defining the discrimination measures of features. In this way, we rank features according to their discrimination measures and select the best feature set for object classification. It should be pointed out that the feature selection criterion is dependent on the classification method. Thus, the choice of between-to within-class variance ratio statistics would be better suited to the use of the minimum-distance classifier discussed in Section 4.2. In the following, we present a rank order procedure for selecting discriminative features from the set of wavelet moment invariants $\{\| F_{m,n,q}^{wavelet} \|\}$. The feature selection procedures for Li's moment invariants and Zernike's moment invariants used in our experiments are similar to the procedure suggested here.

4.1. An automatic discriminative feature selection algorithm

The mean of each invariant feature $\| F_{m,n,q}^{wavelet} \|$ for shape S_i, $m(S_i, \| F_{m,n,q}^{wavelet} \|)$, and the standard deviation $\sigma(S_i, \| F_{m,n,q}^{wavelet} \|)$ can be estimated from a sufficient number of samples of the shape S_i. In our experiments, the number of training samples, NTS, used for each shape S_i is $NTS = 30$. Some examples of the training samples for the shape "|" are shown in Fig. 4a, and those of the shape "1" are shown in Fig. 4b. Table 1 gives means and standard deviations of six wavelet moment invariants for shapes "1" and "|".

The between-to within-class variance ratio, or its inverse, may be used to describe the discrimination degree of a certain feature for discriminating two classes. We specifically use the following discrimination measure which evaluates the effectiveness of using the feature $\| F_{m,n,q}^{wavelet} \|$ to differentiate between two shapes S_i and S_j:

$$Q(\| F_{m,n,q}^{wavelet} \|, S_i, S_j) = \frac{\eta(\sigma(S_i, \| F_{m,n,q}^{wavelet} \|) + \sigma(S_j, \| F_{m,n,q}^{wavelet} \|))}{|m(S_i, \| F_{m,n,q}^{wavelet} \|) - m(S_j, \| F_{m,n,q}^{wavelet} \|)|},$$

where $\eta = 3.0$. This measure is similar to the inverse of the square root of the between-to within-class variance ratio. It is a probabilistic separability measure, if we consider the feature distribution as a Gaussian function. The reason for choosing η to be 3.0 is based on the property that the probability of a class conditional Gaussian variable distributed in the interval $[m - 3.0^* \sigma, m + 3.0^* \sigma]$ is about 99.8%. Thus, the smaller the

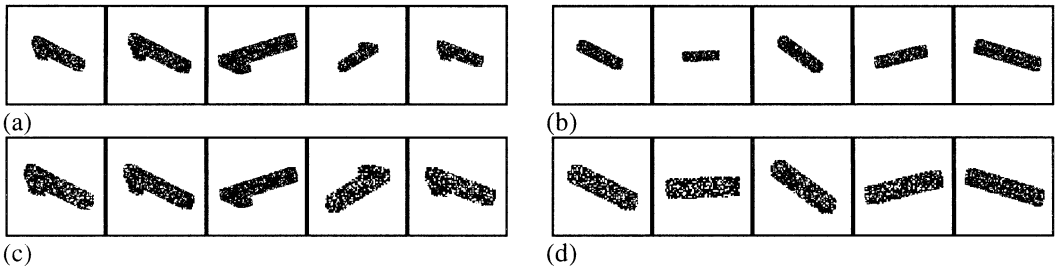


Fig. 4. Some training images of "1" and "|". (a) for "1" and (b) for "|", while (c) and (d) are the translation and scaling normalized versions.

Table 1
Means and standard deviations of six wavelet moment invariants for shapes "1" and "|"

	"1"		"|"		
Features	Mean	Standard deviation	Mean	Standard deviation	Q("1", "|")
$\|F^{wavelet}_{0,0,0}\|$	88.9	2.4	72.8	3.5	1.1
$\|F^{wavelet}_{0,1,0}\|$	44.9	2.6	30.0	3.4	1.2
$\|F^{wavelet}_{0,2,0}\|$	4.2	1.3	10.6	1.6	1.3
$\|F^{wavelet}_{1,0,0}\|$	17.5	3.1	32.5	2.8	1.2
$\|F^{wavelet}_{1,1,0}\|$	21.1	2.6	8.7	3.3	1.4
$\|F^{wavelet}_{1,2,0}\|$	19.5	2.9	24.2	2.4	3.4

value of the discrimination measure $Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j)$ is, the better the feature $\|F^{wavelet}_{m,n,q}\|$ can resolve shapes S_i and S_j. If $Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j)$ is smaller than 1, then the feature $\|F^{wavelet}_{m,n,q}\|$ is guaranteed to be able to differentiate between shapes S_i and S_j. Table 1 also shows values of $Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j)$ for the six wavelet moment invariants. Five of which are quite close to 1. It has been noted that there exist some wavelet moment invariants having discriminative measures $Q(\|F^{wavelet}_{m,n,q}\|$, "1", "|") smaller than 1; however none of Li's moment invariants or Zernike's moment invariants has a discriminative measure smaller than 1.

Suppose that there are N_{class} classes in the set of objects to be discriminated. In order to select discriminative features, we calculate the discrimination measures for each feature $\|F^{wavelet}_{m,n,q}\|$ and select a set $\{Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j), 1 \leqslant i, j \leqslant N_{class}\}$ whose values are smaller than 1. The discrimination number, $NDD(\|F^{wavelet}_{m,n,q}\|)$ which reflects the feature's capability of providing pairwise class separation, is defined as follows:

$$NDD(\|F^{wavelet}_{m,n,q}\|) = \sum_{i=1}^{N_{class}} \sum_{j=1, j\neq i}^{N_{class}} w(Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j)),$$

where $w(x) = 1$ if $x < 1$, otherwise $w(x) = 0$. For each feature $\|F^{wavelet}_{m,n,q}\|$, we define the worst overall discriminative measure, that is,

$$Q^{worst}(\|F^{wavelet}_{m,n,q}\|) = \max_{1\leqslant i,j\leqslant N_{class}, i\neq j} \{Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j) \cdot w(Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j))\}$$

from the set $\{Q(\|F^{wavelet}_{m,n,q}\|, S_i, S_j) < 1, 1 \leqslant i, j \leqslant N_{class}$ and $i \neq j\}$. Using these measures, we propose a procedure for selecting discriminative features as follows:

Discriminative feature selection algorithm:

(1) According to the discrimination number $NDD(\|F^{wavelet}_{m,n,q}\|)$ and the worst overall discriminative measure $Q^{worst}(\|F^{wavelet}_{m,n,q}\|)$, rank all wavelet moment invariants $\|F^{wavelet}_{m,n,q}\|$ for $m = 0, 1, 2, 3$, $n = 0, 1, \ldots, 2^{m+1}$ and $q = 0, 1, 2, 3$. The feature with the largest discrimination number is ranked first. For features with the same discrimination number, we rank them in the ascending order according to their worst overall discriminative measures.
(2) Select the *top* $N_{feature}$ features from the ordered feature list as discriminative features.

For shapes "1" and "|" shown in Fig. 4, the selected first and second best wavelet moment invariants are $\|F^{wavelet}_{1,2,1}\|$ and $\|F^{wavelet}_{1,2,2}\|$ respectively. Notice that these two wavelet moment invariants come from the same wavelet $\psi_{1,2}(r)$ at scale level 1 with shift index 2, which is shown in Fig. 5. Since the radial distance is in the domain $\{0 \leqslant r \leqslant 1\}$, only the part of $\psi_{1,2}(r)$ in this domain is displayed in Fig. 5. It is important to indicate that the maxima of this wavelet is located near an area where the difference of two shapes occur. For the training samples of the shapes "1" and "|", their corresponding invariant moments are shown in Figs. 6a1 and b1. The horizontal axis denotes the labeling of the training samples, while the vertical axis denotes the feature values. The solid curve represents the feature values for the shape "1", while the dotted curve represents the feature values for the shape "|".

For comparison purpose, we applied the same feature selection method to Li's moment invariants of the shapes

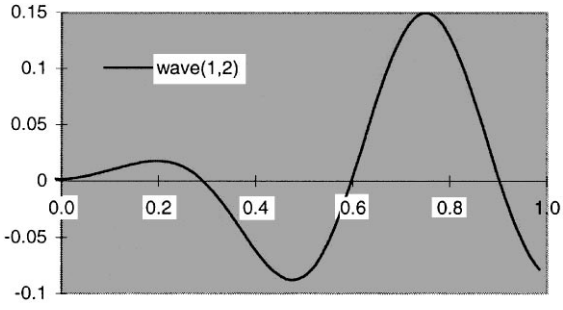


Fig. 5. The wavelet $\psi_{1,2}(y) = \text{wave}(1,2)$ involved in the selected best two discriminating wavelet moment invariants for shapes "1" and "|".

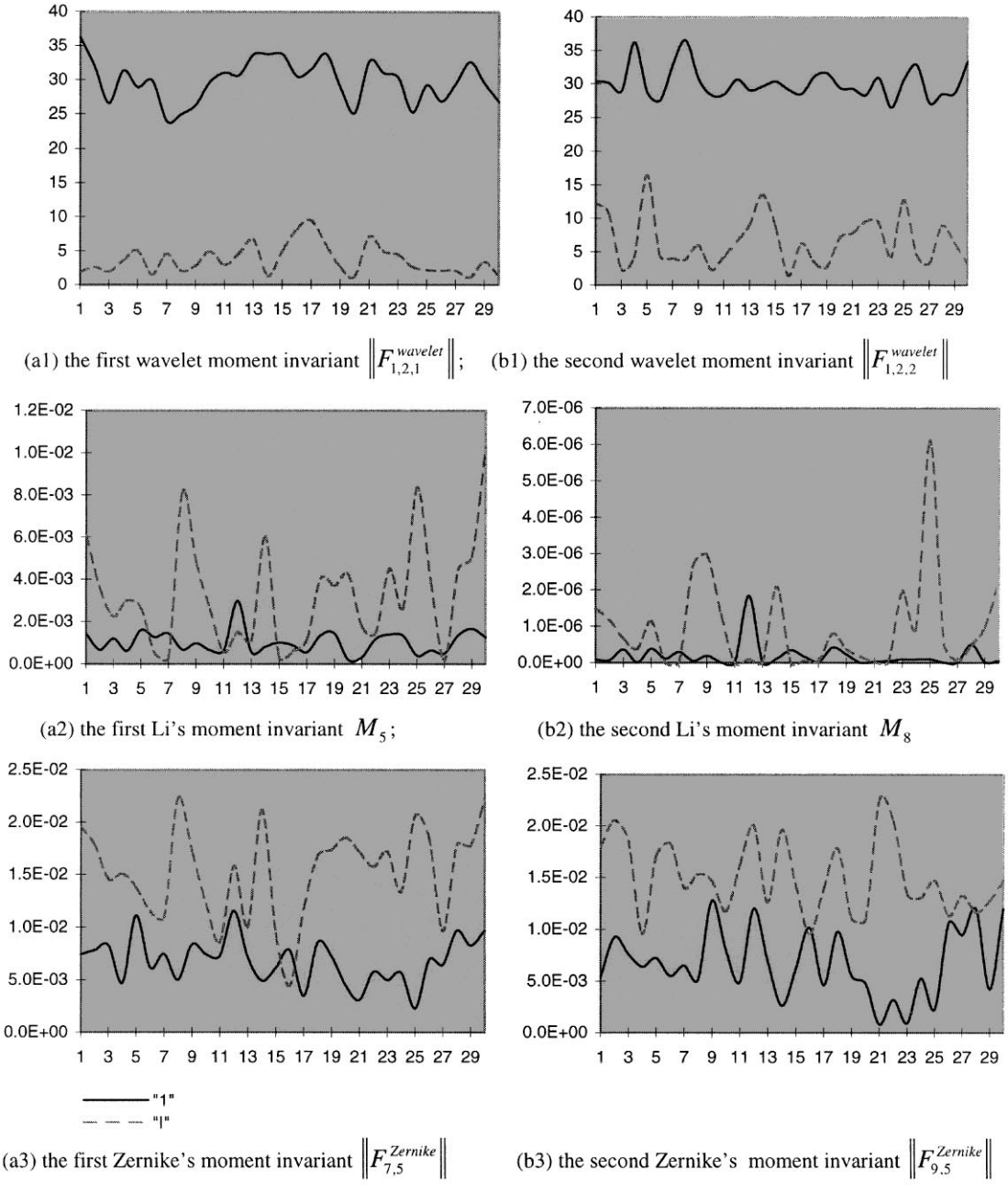


Fig. 6. Variations of values of the first two features with the samples of shapes "1" and "|": (a1) and (b1) are respectively for the two selected wavelet moment invariants, (a2) and (b2) for the two selected Li's moment invariants, and (a3) and (b3) for the two selected Zernike's moment invariants.

"1" and "|" and obtained the first and second best features as M_5 and M_8 respectively. Their corresponding values of 30 training samples of each shape are given in Figs. 6a2 and b2, respectively. Similarly, for Zernike's moment invariants, the selected first and second best features are $\|F_{7,5}^{Zernike}\|$ and $\|F_{9,5}^{Zernike}\|$, and their corresponding values of the training samples are shown in Figs. 6a3 and b3. From all the data shown in Fig. 6, it is clear that the values of the wavelet moment invariants for the two different shapes are well separated, while the values of Li's moment invariants (or Zernike's moment invariants) overlap to a certain extent for these two

shapes. This is demonstrated more clearly by scattergrams of the two shapes in the 2D feature space as shown in Fig. 7. Only the scattergrams of the two shapes in the feature plane of the wavelet moment invariants are well separated, the scattergrams of the two shapes in other feature planes (of either Li's moment invariants or Zernike's moment invariants) are overlapping. In summary, Figs. 6 and 7 demonstrate that the wavelet moment invariants are better features to use for discriminating shapes that are seemingly similar, and that our feature selection procedure *automatically* selects these discriminative features.

4.2. Minimum distance classification

We will evaluate the selected discriminative features with the minimum-distance classification rule in all the experiments presented in this paper. We compare

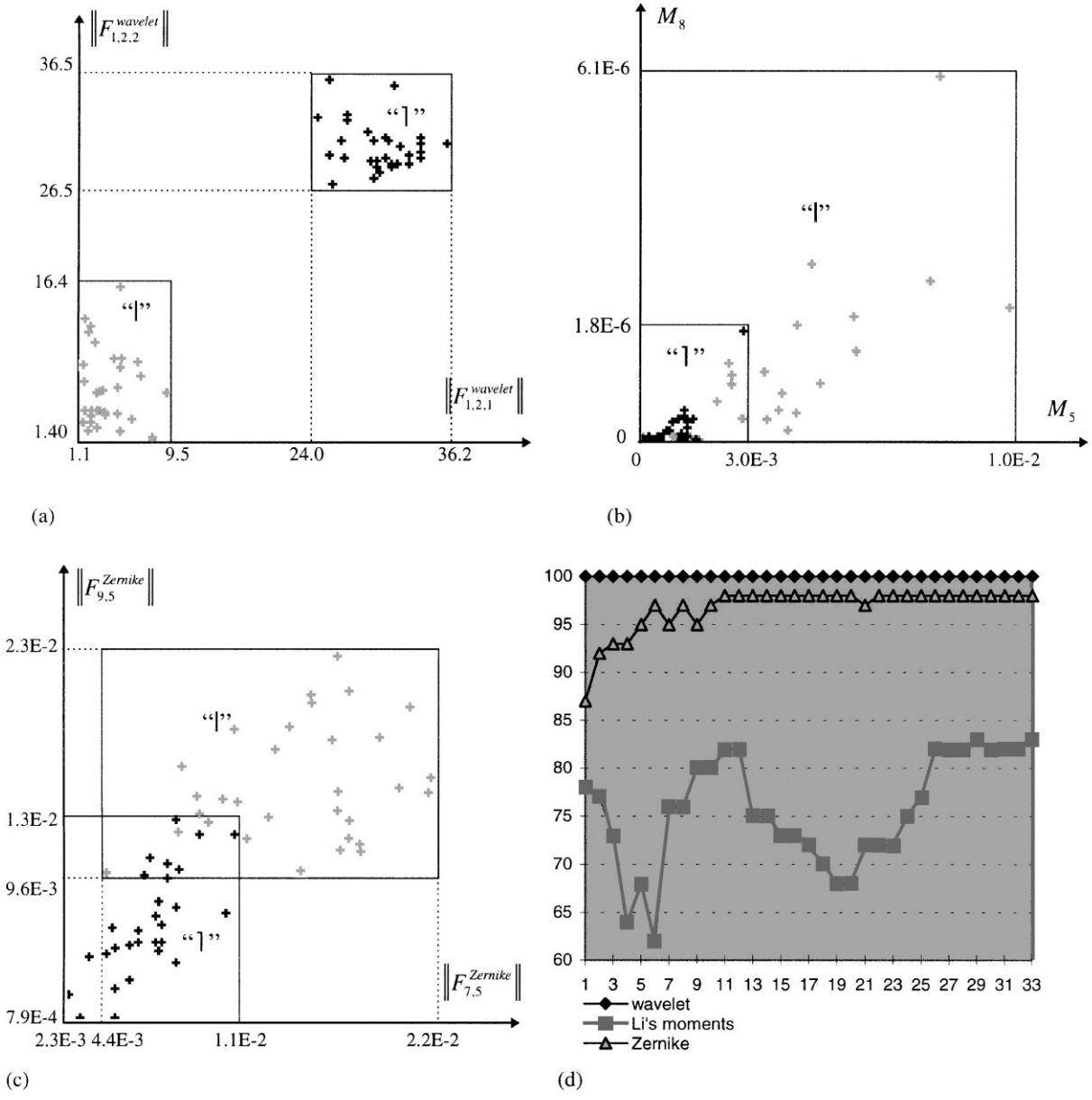


Fig. 7. Scattergrams of shapes "1" and "|" in the selected 2D feature spaces: (a) of the two best wavelet moment invariants, (b) of the two best Li's moment invariants, and (c) of the two best Zernike's moment invariants; While (d) shows classification rate of 100 testing samples varies with the number of selected features.

classification rates based on features derived from the selected wavelet moment invariants, the selected Li's moment invariants, and the selected Zernike's moment invariants. Let $\{F_k^{selected} | k = 1,2, \ldots, N_{feature}\}$ denote the selected features via our feature selection method. Let the number of the selected features be $N_{feature}$, the number of object classes be N_{class}, and the number of the training samples for every class be N_{sample}. Furthermore, let $Data(F_k^{selected}, Sample_j^{(i)})$ denote the kth selected feature of the jth training sample of the ith class. Then for each class of samples we can obtain the sample average, $Average_k^{(i)}$, and sample variance, $Variance_k^{(i)}$, of every selected feature. That is

$$Average_k^{(i)} = \frac{1}{N_{sample}} \sum_{j=1}^{N_{sample}} Data(F_k^{selected}, Sample_j^{(i)}),$$

$$Variance_k^{(i)} = \frac{1}{N_{sample}} \sum_{j=1}^{N_{sample}} (Data(F_k^{selected}, Sample_j^{(i)}) - Average_k^{(i)}))^2.$$

Each class is then represented by a sample mean feature vector $[Average_k^{(i)};\ k = 1, 2, \ldots, N_{feature}]$. When an unknown object with feature vector $X = [X_1, X_2, \ldots, X_{N_{feature}}]$ is to be classified, the nearest neighbor of X is sought among all N_{class} sample mean feature vectors and its class label is assigned to X. The distance between X and the ith class sample mean feature vector is measured using the square of the normalized Euclidean distance. That is

$$d(X, Class^{(i)}) = \sum_{k=1}^{N_{feature}} \frac{(X_k - Average_k^{(i)})^2}{Variance_k^{(i)}}.$$

The object X is classified into class i^*, where i^* satisfies

$$d(X, Class^{(i^*)}) = \min_i d(X, Class^{(i)}),\ i = 1, 2, \ldots, N_{class}.$$

Based on our discriminative feature extraction method, the classification rates for two shapes "1" and "1" using various number of selected features via three different approaches of moment invariants are shown by three curves in Fig. 7d. The classification rates are obtained by using 50 testing samples for each shape. The horizontal axis represents the number of selected features, $N_{feature}$, and the vertical axis represents the percentage of correct classifications. The classification rate of using *any number* (*even only one*) *of wavelet moment features is* 100%. Applying Li's moment invariants as features, the highest classification rate obtained is 83% when 29 features are used. It is noted that Li's moment features in the experiment are unstable as the classification rate oscillates when the number of features increases; Some features introduce contradictory information. This probably is due to insufficient number of training samples used in feature selection process, while the testing samples are relatively large. Using Zernike's moment invariants, the highest classification rate obtained is 98% and the minimum number of features needed is 11. We can conclude that the wavelet moment invariants are better than Zernike's moment invariants, and much better than Li's moment invariants for classifying these two seemingly similar shapes with subtle differences.

5. Experimental results on *N*-class problems with added noise

So far, we have described the wavelet-based method in relation to a 2-class problem and made use of the two shapes "1" and "|". In this section, we present our experiments on other 2-class and N-class problems to assess the robustness of the method in the presence of noise. Two artificially generated shapes (Fig. 8) were used to illustrate that the wavelet-based moments, together with our feature selection method described before, can provide a set of discriminative features. We also used a set of 26 upper cased English letters (Fig. 12), although no two shapes are seemingly similar, to show that the classification performance by using wavelet moment invariants is again better than that by Zernike's moment invariants, and much better than that by Li's moment invariants. Finally, we applied our method to classify two similar wrenches as shown in Fig. 15. From these experiments, we found that when the number of feature reaches a certain limit, the classification accuracy may not be increased with additional features. The effectiveness of some lower ranked features becomes questionable. The discrimination measure Q of a feature was computed based on a small set of training samples which may be insufficient leading to an unreliable ranking. The main reason is probably due to the fact that our feature selection method is suboptimal. The features were *first* ranked according to their individual discriminative powers thus computed, and *then* a subset of the *top* ranking features were sought and used in testing. For the latter, *Branch and Bound algorithm* [20] may be employed to search for all possible feature combinations and guarantee an "optimum" feature set which will be computationally much more demanding.

Set 1: *two artificially generated shapes.* The training set of the two artificially generated shapes illustrated in Fig. 8 consists of 60 images (30 for each shape), which are the *scaled* and *rotated* versions of the two shapes corrupted by random noise. The testing set consists of 100 images (50 for each shape), which are also randomly transformed versions of the two shapes. Fig. 9 shows some noisy samples in various scales and orientations, and their corresponding translation and scaling normalized versions. Based on the proposed method of discriminative feature selection described in Section 4, the first and second best wavelet moment invariants are $\| F_{1,3,0}^{wavelet} \|$ and $\| F_{1,2,0}^{wavelet} \|$ respectively. The corresponding

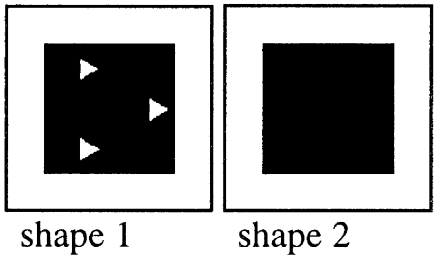


Fig. 8. Two synthetic shapes

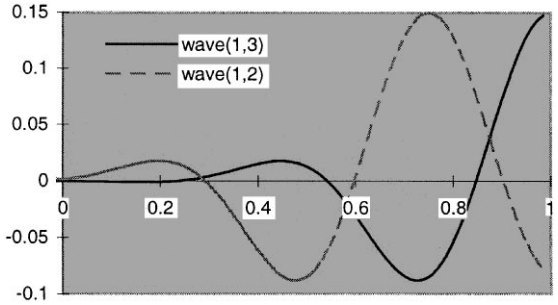


Fig. 10. Two wavelets $\psi_{1,3}(r) = \text{wave}(1,3)$ and $\psi_{1,2}(r) = \text{wave}(1,2)$ involved in the selected best two discriminating wavelet moment invariants for shapes in Fig. 8.

wavelets, $\psi_{1,3}(r)$ and $\psi_{1,2}(r)$, are shown in Fig. 10. Since the radial domain is in $\{0 \leqslant r \leqslant 1\}$, only the parts of $\psi_{1,3}(r)$ and $\psi_{1,2}(r)$ in this domain are displayed. This figure indicates that the maxima of these functions are localized in the region near the three small triangles of shape 1 as shown in Fig. 8, which provide the distinction between these two shapes. As to Li's moment invariants, the first and second best features are M_{12} and M_{33} respectively. While for Zernike's moment invariants, the first and second best features are $\|F_{7,7}^{Zernike}\|$ and $\|F_{9,3}^{Zernike}\|$ respectively. The scattergrams in these three feature planes are shown in Fig. 11. The scattergrams in Fig. 11a corresponding to wavelet moment invariants are separated, while other pairs of scattergrams shown in Figs. 11c and b respectively for Zernike's moment invariants and Li's moment invariants are significantly overlapping. This leads to the superior classification rate by wavelet moment invariant features, as compared with Zernike's moment invariants and Li's moment invariants.

Fig. 11d shows three curves of classification rates corresponding to the use of three different feature sets. Given the same number of features used, the correct classification rate by wavelet moment invariants is higher than that by Zernike's moment invariants and that by Li's moment invariants respectively. When the first one *to* four features were used, the correct classification rate was 100% by wavelet moment invariants. The highest classification rate obtained by using Zernike's moment invariants was 95% when 14 features were used, while the highest classification rate obtained by using Li's moment invariants was only 87% when 24 features were used.

Set 2: *26 upper cased English letters.* The second experimental set consists of 26 upper cased English letters from "A" to "Z". For each letter, there are 30 randomly generated versions in the training set, and other 30 randomly generated versions in the testing set (see Fig. 12). Some of the randomly generated samples of "A" and their scaling normalized images are shown in Fig. 13. Feature selections from three types of features were considered: those derived from wavelet moment invariants, Zernike's moment invariants and Li's moment invariants. Again, the minimum-distance classifier was used. Three curves of classification rates obtained from the testing samples are shown in Fig. 14 for three types of features respectively. The highest classification rate obtained by wavelet moment invariants was 100% using 37 features. Zernike's moment invariants gave the highest classification rate of 98.7% when 26 features were used, while with that number of features the classification rate of wavelet moment invariants was 99.5%. For achieving 98.7% classification rate, only 15 wavelet moment features were required. Using Li's moment invariants, the

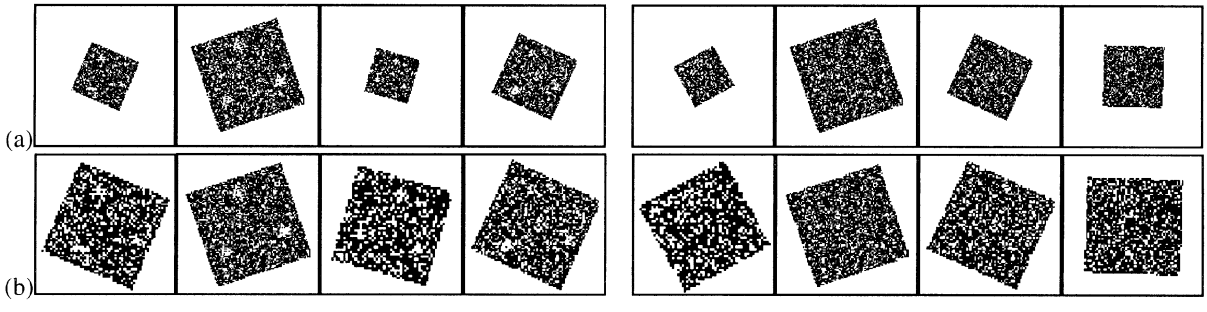
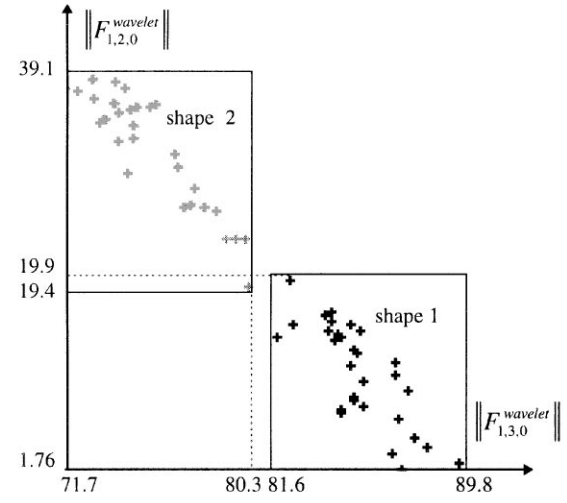
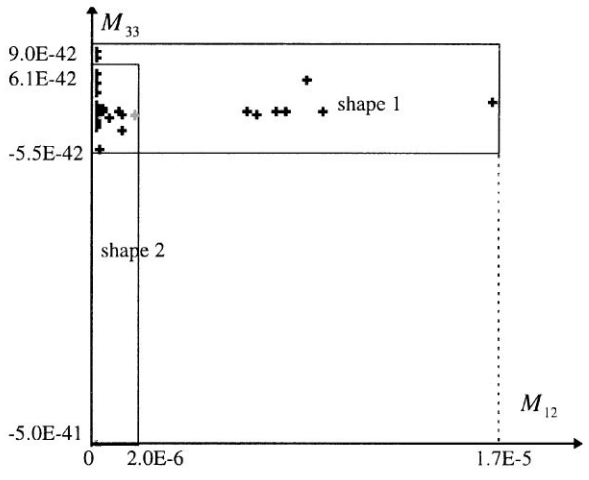


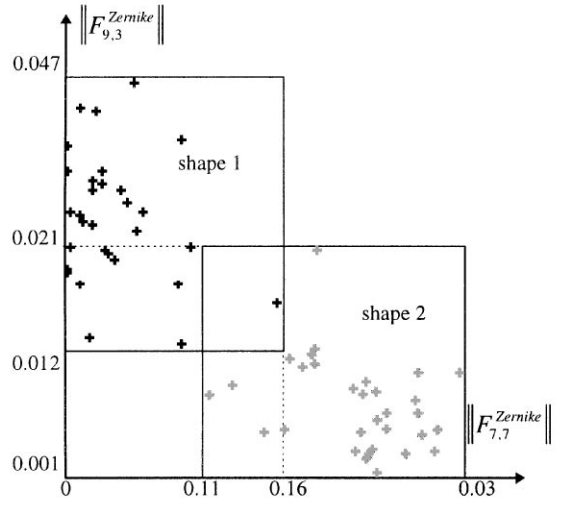
Fig. 9. Some training images of two synthetic shapes (a) the noise-added, scaled and rotated images; (b) the scaling normalized images.



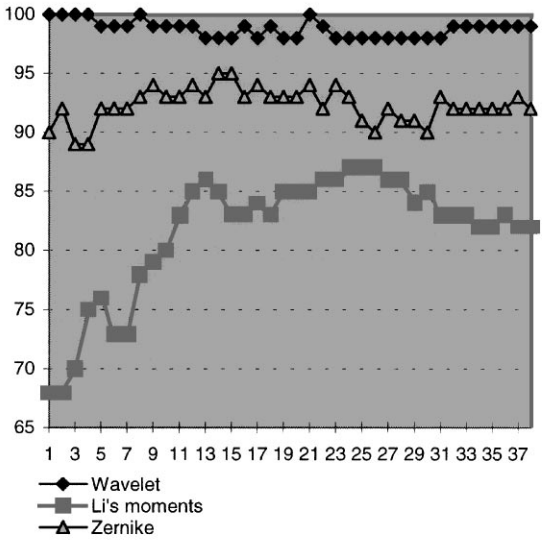
(a)



(b)



(c)



(d)

Fig. 11. Scattergrams of two shapes of Fig. 8 in the selected 2D feature spaces: (a) of the two best wavelet moment invariants, (b) of the two best Li's moment invariants, and (c) of the two best Zernike's moment invariants; (d) classification rates corresponding to three different feature sets used when 100 test samples are used.

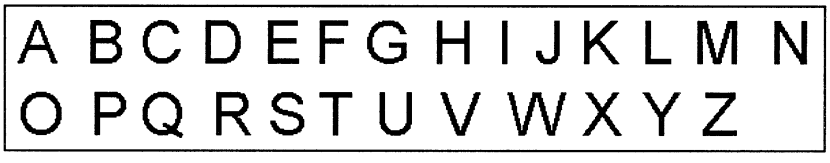


Fig. 12. 26 upper cased English letters used in an experiment.

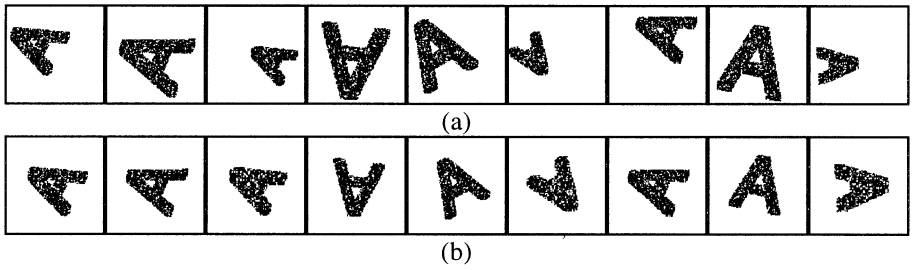


Fig. 13. Some randomly generated samples of "A"; (a) in different scales and orientations, (b) the scaling normalized versions.

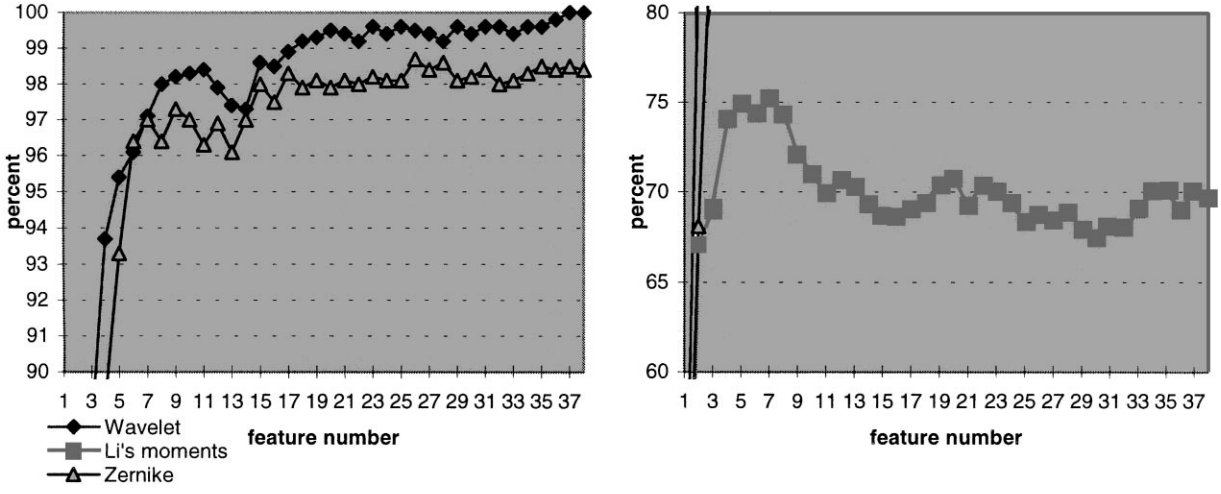


Fig. 14. Test results on classification rates of three different moment-based feature sets for 26 upper cased English letters.

classification rates were all lower than 75.3%. Hence, we are sure that the wavelet moment invariants are the best features among the three families of moment invariants for classifying the 26 upper cased English letters.

Set 3: Two similar wrenches. The third experimental set is on two similar wrenches shown in Fig. 15. There are 100 images (50 for each wrench) in the training set and the testing set respectively, they are noisy, randomly scaled and oriented versions of the two wrenches. According to the discriminative feature selection method, the first and second best wavelet moment invariants are $\|F_{0,1,0}^{wavelet}\|$ and $\|F_{0,1,2}^{wavelet}\|$ respectively. For Li's moment invariants and Zernike's moment invariants, the first and second best features are (M_1 and M_2) and ($\|F_{4,0}^{Zernike}\|$ and $\|F_{2,0}^{Zernike}\|$), respectively. The scattergrams of the wrenches corresponding to these three types of features are shown in Fig. 16. Only the scattergrams in Fig. 16a corresponding to wavelet moment invariants are well separated, while other pairs of scattergrams with Zernike's moment invariants and Li's moment invariants (Figs. 16b and c) respectively are overlapping. This is the

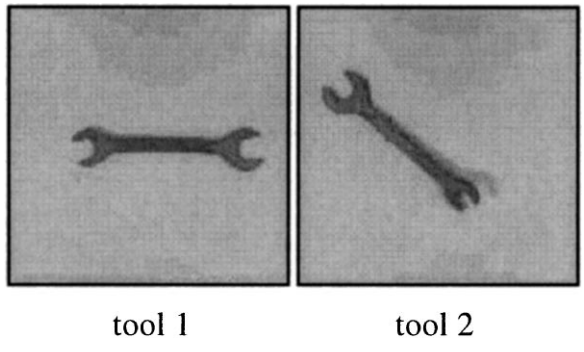


Fig. 15. Two similar wrenches.

reason why the classification rate obtained by wavelet moment invariants was higher than that by Zernike's moment invariants and that by Li's moment invariants (Fig. 16d). The classification rate by wavelet moment invariants is 100% for any number of features used. The highest classification rates achieved by Zernike's moment

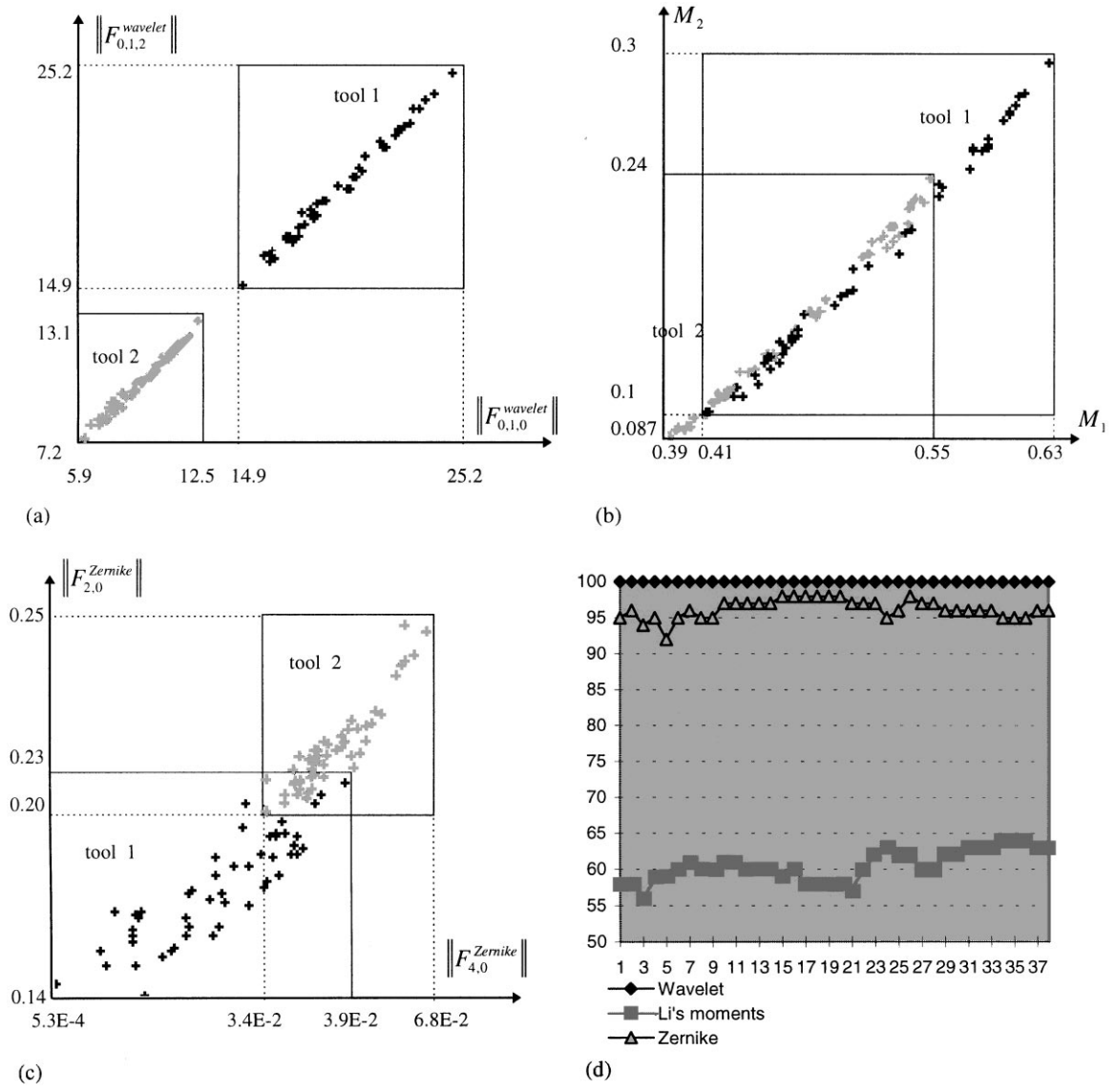


Fig. 16. Scattergrams of two wrenches of Fig. 15 in the selected 2D feature planes: (a) of the two best wavelet moment invariants, (b) of the two best Li's moment invariants, and (c) of the two best Zernike's moment invariants; (d) classification rates corresponding to three different feature sets.

invariants was 98% and that by Li's moment invariants was 64%; For obtaining these highest classification rates, at least 15 features were needed for Zernike's moment invariants and 34 features were needed for Li's moment invariants.

6. Conclusion

We have presented a set of wavelet moment invariants for capturing rotation invariant global and local shape information among the objects of interest. Compared with Zernike's moment invariants and Li's moment invariants, wavelet moment invariants can provide more discriminative features for a variety of shapes even when they are seemingly similar and in the presence of noise. A method of selecting discriminative features, based on a discrimination measure of each feature evaluated on the training samples, is also proposed.

Using a minimum-distance classifier, wavelet moment invariants gave the highest classification rate in all four experiments, as compared with Zernike's moment

invariants and Li's moment invariants. For the three sets of test patterns of Figs. 4, 8 and 15, wavelet moment invariants gave the highest classification rate when using any number of features. For the fourth set (26 upper cased English letters), wavelet moment invariants achieved 100% classification rate when using 37 features, whereas the highest classification rates obtained by Zernike's moment invariants and Li's moment invariants were only 98.7 and 75.3% respectively. Both analyses and experiments have shown that our method is suited to classifying many types of object shapes, particularly, when they are seemingly similar but actually different.

References

[1] R.J. Prokop, A.P. Reeves, A survey of moment-based techniques for unoccluded object representation and recognition, CVGIP: Graphical Models Image Process, Vol. 54(5) (1992) 438–460.

[2] M.K. Hu, Visual pattern recognition by moment invariants, IRE Trans. Inform. Theory 8 (1962) 179–187.

[3] S. Dudani, K. Breeding, R. McGhee, Aircraft identification by moment invariants, IEEE Trans. Comput. C-26 (1977) 39–45.

[4] Y.S. Abu-Mostafa, D. Psaltis, Recognitive aspects of moment invariants, IEEE Trans. PAMI PAMI-6 (1984) 698–706.

[5] T.H. Resis, The revised fundamental theorem of moment invariants, IEEE Trans. PAMI PAMI-13 (1991) 830–834.

[6] Y. Li, Reforming the theory of invariant moments for pattern recognition, Pattern Recognition 25 (1992) 723–730.

[7] M. Teague, Image analysis via the general theory of moments, J. Opt. Soc. Am. 70 (1980) 920–930.

[8] C.H. Teh, R.T. Chin, On image analysis by the methods of moments, IEEE Trans. PAMI 10 (1988) 496–513.

[9] A. Khotanzad, Y.H. Hong, Rotation invariant image recognition using features selected via a systematic method, Pattern Recognition 23(10) (1990) 1089–1101.

[10] R.R. Bailey, M. Srinath, Orthogonal moment featurs for use with parametric and non-parametric classifiers, IEEE Trans. PAMI 18(4) (1996) 389–399.

[11] A. Sluzek, Identification and inspection of 2-D objects using new moment-based shape descriptors, Pattern Recognition Lett. 16 (1995) 687–697.

[12] K. Tsirikolias, B.G. Mertzios, Statistical pattern recognition using efficient two-dimensional moments with applications to character recognition, Pattern Recognition 26(6) (1993) 877–882.

[13] I. Daubechies, Ten lecture on Wavelets, CBMS-NSF Regional Conf. Series in Applied Mathematics (00) 61, 1992.

[14] C.K. Chui, An Introduction to Wavelets, Academic Press, San Diego CA, 1992.

[15] D. Shen, H.H.S. Ip, Optimal axes for defining the orientations of shapes, IEE Electron. Lett. 32(20) (1990) 1873–1874.

[16] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Trans. Inform. Theory 36 (1990) 961–1005.

[17] M. Unser, A. Aldroubi, M. Eden, On the asymptotic convergence of B-spline wavelets to Gabor functions, IEEE Trans. Inform. Theory 38 (1992) 864–872.

[18] M. Unser, A practical guide to the implementation of the wavelet transform, in: A. Aldroubi, M. Unser, (Eds.) Wavelets in Medicine and Biology, CRC Press, Boca Raton, FL, 1996, pp. 37–73.

[19] R.O. Duda, P.E. Hart, Pattern Recognition and Scene Analysis, Wiley, New York, 1973.

[20] P.A. Devijver, J. Kittler, Pattern Recognition: a Statistical Approach, Prentice-Hall Englewood Cliffs, NJ, 1982.

About the Author—DINGGANG SHEN received his BS, MS, PhD degrees in electronic engineering from Shanghai JiaoTong University in 1990, 1992 and 1995, respectively. He worked as Research Assistant in the Department of Computer Science at the Hong Kong University of Science & Technology from December 1994 to June 1995. From September 1995 to February 1996, he was a Lecturer of Communication Engineering at Shanghai JiaoTong University. He was once a Research Fellow in the Department of Computer Science at City University of Hong Kong, from February 1996 to August 1997. Since June 1997, he works first as Post-Doctoral Fellow and then as Research Fellow in School of Electrical and Electronic Engineering at Nanyang Technological University, Singapore. His research interests are in the areas of computer vision, pattern recognition, image processing, neural network and image indexing & retrieval.

About the Author—HORACE H.S. IP received his B.Sc. degree in Applied Physics and Ph.D. degree in Image Processing from University College London, United Kingdom, in 1980 and 1983 respectively. Presently, he is the Head of the Department of Computer Science and also heads the Image Computing Research Group, at City University of Hong Kong. His research interests include image processing and analysis, pattern recognition, hypermedia computing systems and computer graphics. Prof. Ip is a member of the Editorial Board of the *Pattern Recognition* Journal (Elsevier), the *International Journal of Multimedia Tools and Applications* (Kluwer Academic), and the Chinese Journal of *CAD and Computer Graphics* (The Chinese Academy of Science) and a guest editor of the international journal of *Real-Time Imaging* (Academic Press) and *Real-time Systems* (Academic Press).

Prof. Ip serves on the IAPR Governing Board and co-chairs its Technical Committee (TC-12) on Multimedia Systems. He was the Chairman of the IEEE (HK) Computer chapter in 1996–97, and the Founding President of the Hong Kong Society for Multimedia and Image Computing. He has published over 80 papers in international journals and conference proceedings.