Neural Comput & Applic
DOI 10.1007/s00521-013-1356-4

ORIGINAL ARTICLE

Human face recognition based on ensemble of polyharmonic extreme learning machine

Jianwei Zhao · Zhenghua Zhou · Feilong Cao

Received: 21 February 2012 / Accepted: 29 January 2013

Abstract This paper proposes a classifier named ensemble of polyharmonic extreme learning machine, whose part weights are randomly assigned, and it is harmonic between the feedforward neural network and polynomial. The proposed classifier provides a method for human face recognition integrating fast discrete curvelet transform (FDCT) with 2-dimension principal component analysis (2DPCA). FDCT is taken to be a feature extractor to obtain facial features, and then these features are dimensionality reduced by 2DPCA to decrease the computational complexity before they are input to the classifier. Comparison experiments of the proposed method with some other state-of-the-art approaches for human face recognition have been carried out on five well-known face databases, and the experimental results show that the proposed method can achieve higher recognition rate.

Keywords Human face recognition · Ensemble of polyharmonic extreme learning machine · Fast discrete curvelet transform · 2-Dimension principal component analysis

1 Introduction

Human face recognition, as one of the most important methods for the identification, has been widely studied recently. Compared with some other approaches in biometrics, such as fingerprint, iris, and DNA recognition, human face recognition may not have a better level of accuracy, but it is inexpensive, convenient, and hassle-free. Therefore, human face recognition has widespread applications, such as smart cards, telecommunication, database security, medical records, digital libraries, and so on [1].

Human face recognition system includes two key steps: feature extraction and classification. Feature extraction is to give an effective representation of the face images to decrease the computational complexity of the classifier, which can greatly enhance the performance of the human face recognition system. While classification is to distinguish those features with a good classifier. Therefore, in order to improve the recognition rate of a human face recognition system, it is crucial to find a good feature extractor and an effective classifier.

Recently, many methods for dimensionality reduction have been proposed to improve the accuracy and speed of a human face recognition system [2–5]. Kirby et al. [6] and Turk et al. [7] employed principal component analysis (PCA) to obtain a lower dimensional representation of the human face image. As one of classical methods for dimensionality reduction, PCA provides effective approximation, but it suffers from high combinational load and poor discriminatory power [8]. To eliminate these limitations of PCA, researchers have proposed some other techniques for dimensionality reduction, such as independent component analysis (ICA) [9], linear discriminant analysis (LDA) [10], Kernel principal component analysis (KPCA) [11], Kernel linear discriminant analysis (KLDA) [12], and so on. However, converting matrices into vectors often results in a high-dimensional vector space, in which it is difficult to evaluate the covariance matrix accurately due to its large size and the relatively small number of training samples. Furthermore, it is time-consuming to compute the eigenvectors of a large size covariance matrix. In order to

J. Zhao · Z. Zhou · F. Cao (✉)
Department of Mathematics, China Jiliang University,
Hangzhou 310018, Zhejiang Province,
People's Republic of China
e-mail: feilongcao@gmail.com

Published online: 15 February 2013

overcome these limitations and flaws, a new approach called two-dimensional principal component analysis (2DPCA) was proposed by Yang et al. [13], which directly computes eigenvectors of the so-called image covariance matrix without matrix-to-vector conversion. Compared to the traditional PCA, 2DPCA is completely based on the matrix rather than vectors. The obvious advantage of 2DPCA is that the image matrix does not need to be vectorized prior to feature extraction [14].

At the same time, many transform-based methods have been proposed to enhance the performance of the human face recognition system. These methods transform the face images into a new domain to achieve a higher level of class separability. At the early stage, wavelet transform was often applied in human face recognition system, including two-dimensional discrete wavelet transform [15], wavelet and curvelet moments [16], fast beta wavelet networks [17], discrete wavelet transform [18], and wavelet-based feature selection [19–22]. With the emergence of curvelet transform [23], many researchers propose some improved methods for human face recognition based on the curvelet transform, such as curvelet-based moment method [16], fast discrete curvelet transform (FDCT) [24], curvelet-based image fusion [25–27], and so on.

Human face recognition is in fact a multi-classification problem. The main difficulty in it is to find a classifier with good generalization performance. In this work, we aim to find an image classifier with a better generalization capability. In the work of human face recognition, there have been various classifiers, both linear and nonlinear, such as artificial feedforward neural networks (FNNs), support vector machine (SVM), polynomial classifier, and fuzzy rule-based system. In those classifiers, FNN seems one of the most popular techniques. The classifier for human face recognition of this paper falls into the category of FNNs. Although the FNNs have being intensively studied for many years, most of them may be regarded as different variations of the perception to recognize human face features [28–32].

It is well known that traditional FNNs approaches, such as perceptron, backpropagation (BP) algorithm, and radial basis function (RBF) network, often face difficulties in tuning parameters manually, which results in the slow learning speed. Fortunately, Huang et al. [33, 34] has proposed a fast learning algorithm, named Extreme Learning Machine (ELM), for training FNN with single hidden layer, and some other people discuss the improvements and applications of ELM, such as generalization performance of learning system [35, 36], image processing [37–39], fuzzy integral determinant [40], fuzzy rule learning [41], and so on. In this paper, we will propose a noval classifier, called ensemble of polyharmonic extreme learning machine (EP-ELM), for improving the accuracy and speed of recognition based on polyharmonic extreme learning machine (P-ELM) [39]. P-ELM added a polynomial with low degree in the FNN, and the main function of which is to reconcile the balance between the FNN and the polynomial. The proposed EP-ELM improve the stability of P-ELM by comprising several P-ELM.

This paper presents a method for human face recognition, which is a combination of FDCT, 2DPCA, and the proposed EP-ELM classifier. At first, the FDCT is used to be a feature extractor to obtain facial features. Then by using 2DPCA, those features are dimensionality reduced to decrease the computational complexity. Finally, the EP-ELM classifier is proposed to classify the features to complete the face recognition. Compared with some other state-of-the-art approaches in five well-known face database, the proposed method for human face recognition achieves the higher recognition rate.

We organize the study in the following manner. The starting point is a discussion on the proposed method which combines FDCT, 2DPCA, and the proposed EP-ELM classifier. These issues are covered in Sect. 2. Section 3 provides some comparison experiments of the proposed method with some other approaches for human face recognition. Conclusions are given in Sect. 4.

2 Proposed human face recognition method

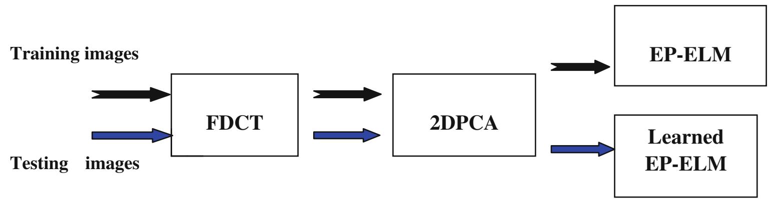
In this section, we will propose a novel method for human face recognition. The proposed method is composed of three stages as follows: feature extraction by FDCT, dimensionality reduction by 2DPCA, and classification by the proposed EP-ELM classifier. A block schematic diagram of the proposed method is shown in Fig. 1.

Figure 1 displays the process of our proposed method for human face recognition. At first, some features of training images are extracted with FDCT, and then they are dimensionally reduced by means of 2DPCA, and finally, these reduced features are taken as sample data to train EP-ELM classifier. With the learned EP-ELM classifier, we can recognize human face images. That is, we take the same pretreatment in the first two steps for testing images and training images, then use the learned EP-ELM to classify the corresponding features of testing images to realize human face recognition.

2.1 Feature extraction by FDCT

In the first step of the proposed method, curvelet transform is used to generate initial feature vectors for human face recognition. Although images can be directly input to a classifier to realize the classification, it will result in huge computational cost for the large scale of image data. Therefore, it is necessary to extract some key information

Fig. 1 Diagram of the proposed method for human face recognition



(features) from the original image data with some transforms. In our proposed method, we will extract features from the original facial images with the FDCT proposed by Candès et al. in the paper [24].

Let ω denote the frequency variant, r and θ represent polar-coordinate, then there exist two smooth and non-negative windows, radial window W, and angular window V, satisfying the following conditions

$$\sum_{j=-\infty}^{+\infty} (W(2^j r))^2 = 1,\ r \in (0.75, 1.5) \tag{2.1}$$

and

$$\sum_{l=-\infty}^{+\infty} (V(t-l))^2 = 1,\ t \in (-0.5, 0.5), \tag{2.2}$$

respectively. The frequency window U_j of Fourier frequency is defined by

$$U_j(r,\theta) = 2^{-\frac{3}{4}j} W(2^{-j} r) V(\frac{2^{\lfloor \frac{j}{2} \rfloor}\theta}{2\pi}), \tag{2.3}$$

where $\lfloor \frac{j}{2} \rfloor$ denotes the integer part of $\frac{j}{2}$. From formula (2.1), U_j is a wedge window generated by the radial window and the angular window.

Choose a mother curvelet φ_j such that its corresponding Fourier transform $\hat{\varphi}_j$ is U_j, then all the curvelets at the scale 2^{-j} can be obtained by rotating and translating $\hat{\varphi}_j$. Let R_θ denote the 2×2 orthogonal rotation matrix with respect to the parameter θ ($0 \le \theta < 2\pi$), and $\{\theta_l\}_{l=0}^{+\infty}$ be the sequence of equispaced rotating angular defined by

$$\theta_l = 2\pi \times 2^{-\lfloor j/2 \rfloor} \times l, \tag{2.4}$$

then for each j, l, and the translation vector $\mathbf{k} = [k_1, k_2]^\top \in \mathbb{Z}^2$, the curvelet $\varphi_{j,l,\mathbf{k}}$ can be defined by

$$\varphi_{j,l,\mathbf{k}}(\mathbf{x}) = \varphi_j\left(R_{\theta_l}(\mathbf{x} - \mathbf{x}_{\mathbf{k}}^{(j,l)})\right), \mathbf{x} \in \mathbb{R}^2, \tag{2.5}$$

where $\mathbf{x}_{\mathbf{k}}^{(j,l)} = R_{\theta_l}^{-1}[k_1 2^{-j}, k_2 2^{-j/2}]^\top$.

Now the curvelet coefficients of a function $f \in L^2(\mathbb{R}^2)$ can be calculated by the inner product of the element f and the curvelets $\varphi_{j,l,\mathbf{k}}$, that is,

$$c(j,l,\mathbf{k}) = \langle f, \varphi_{j,l,\mathbf{k}} \rangle = \int_{\mathbb{R}^2} f(\mathbf{x}) \overline{\varphi_{j,l,\mathbf{k}}(\mathbf{x})} d\mathbf{x}. \tag{2.6}$$

Since digital curvelet transforms operate on the frequency domain, it is useful to apply Plancherel's theorem to express the inner product in formula (2.6) as an integral over the frequency plane like

$$c(j,l,\mathbf{k}) = \frac{1}{(2\pi)^2} \int \hat{f}(\omega) \overline{\hat{\varphi}_{j,l,\mathbf{k}}(\omega)} d\omega. \tag{2.7}$$

In order to apply the curvelet transform in human face recognition conveniently, we will adopt its corresponding discrete style. In our proposed method, we will use the FDCT via wrapping to extract features.

For each j and $l (l = -2^{\lfloor j/2 \rfloor}, \ldots, 2^{\lfloor j/2 \rfloor} - 1)$, let $\theta_l = l \times 2^{-\lfloor j/2 \rfloor}$ and

$$S_{\theta_l} = \begin{bmatrix} 1 & 0 \\ -tan\theta_l & 1 \end{bmatrix}. \tag{2.8}$$

Define a local window $\hat{U}$ as

$$\hat{U}_{j,l}(\omega) = W_j(\omega) V_j(S_{\theta_l}\omega), \tag{2.9}$$

then the discrete curvelet coefficients of a function $f \in L^2(\mathbb{R}^2)$ can be calculated by

$$c(j,l,\mathbf{k}) = \int \hat{f}(S_{\theta_l}\omega) \hat{U}_j(\omega) e^{i\langle b,\omega \rangle} d\omega, \tag{2.10}$$

where $b = (k_1 \times 2^{-j}, k_2 \times 2^{-j/2})$.

Now the FDCT via wrapping can be summarized as follows.

FDCT via wrapping 2.1 [24]

1. For a given bivariate function f (image datum can be regarded as a bivariate function), compute its coefficients with 2D fast Fourier transform (FFT) to get its representation samples $\hat{f}(n_1, n_2)$ on the Fourier frequency, where $-n/2 \le n_1, n_2 < n/2$.
2. For each scale j and angle l, form the product $\hat{U}_{j,l}[n_1, n_2]\hat{f}(n_1, n_2)$.
3. Wrap this product around the origin and obtain $\hat{f}_{j,l}[n_1, n_2] = W(\hat{U}_{j,l}\hat{f})[n_1, n_2]$, where the ranges of n_1, n_2 and θ are $0 \le n_1 \le L_{1,j}$, $0 \le n_2 \le L_{2,j}$ and $(-\pi/4, \pi/4)$, respectively.
4. Apply the inverse 2D FFT to each $\hat{f}_{j,l}$ and save discrete curvelet coefficients.

2.2 Dimensionality reduction by 2DPCA

Although FDCT can efficiently extract features from original facial images, the dimension of coarse curvelet coefficients matrix is still high, which will result in huge computational cost if these features are input to the classifier directly. Therefore, it is necessary to perform the work of dimensionality reduction to simplify the face recognition system. In our proposed method for human face recognition, we use the perfect 2DPCA proposed by Yang et al. [13] to reduce the dimension of the extracted features.

Let A be a $m \times n$ coarse subband coefficient matrix extracted from a facial image with FDCT, the goal of 2DPCA is to find the principal components of the matrix A. Now we begin to explain the steps of 2DPCA.

Let $\mathbf{x}$ be an $n \times 1$ unitary vector, then we can get an $m \times 1$ projected vector $\mathbf{y}$ by the linear transformation

$$\mathbf{y} = A\mathbf{x}. \tag{2.11}$$

So the covariance matrix of the projected vector $\mathbf{y}$ can be represented by

$$C_{\mathbf{x}} = E(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})^{\top} = E[(A - EA)\mathbf{x}][(A - EA)\mathbf{x}]^{\top}, \tag{2.12}$$

where E denotes the expectation.

Let $J(\mathbf{x})\!: = tr(C_{\mathbf{x}})$, called the generalized total scatter criterion, where $tr(C_{\mathbf{x}})$ is the trace of $C_{\mathbf{x}}$, then

$$J(\mathbf{x}) = \mathbf{x}^{\top}\left(E[(A - EA)^{\top}(A - EA)]\right)\mathbf{x} := \mathbf{x}^{\top}D_l\mathbf{x}, \tag{2.13}$$

where $D_l := E[(A - EA)^{\top}(A - EA)]$ is a non-negative definite $n \times n$ image covariance matrix.

Suppose that there are M coarse subband coefficient matrice $\{A_j | j = 1, 2, \ldots, M\}$ extracted from M facial images with FDCT, then the average image $\bar{A}$ of these matrice is

$$\bar{A} = \frac{1}{M}\sum_{j=1}^{M} A_j. \tag{2.14}$$

So the image covariance matrix D_l can be represented as follows:

$$D_l = \frac{1}{M}\sum_{j=1}^{M}(A_j - \bar{A})^{\top}(A_j - \bar{A}) \tag{2.15}$$

and the criterion $J(\mathbf{x})$ can be expressed by

$$J(\mathbf{x}) = \mathbf{x}^{\top}D_l\mathbf{x}. \tag{2.16}$$

Let $\mathbf{x}_{\text{opt}}$ be the unitary vector such that it maximize the generalized total scatter criterion $J(\mathbf{x})$, that is,

$$\mathbf{x}_{\text{opt}} = \arg\max_{\mathbf{x}} J(\mathbf{x}). \tag{2.17}$$

We call $\mathbf{x}_{\text{opt}}$ the optimal projection axes.

In general, there are more than one optimal projection axes. We usually select a set of optimal projection axes $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$ subjected to the orthonormal constraints and the maximization of the criterion $J(\mathbf{x})$. Actually, the optimal projection axes are the orthonormal eigenvectors of D_l corresponding to d largest eigenvalues.

Now for each coarse subband coefficient matrix A, we can compute the principal component of the A as follows

$$\mathbf{y}_i = A\mathbf{x}_i, \quad i = 1, 2, \ldots, d. \tag{2.18}$$

These principal component vectors form an $m \times d$ matrix $Y = [\mathbf{y}_1, \ldots, \mathbf{y}_d]$, which is called feature matrix. Up to now, each facial image matrix has been turned into a feature matrix.

2.3 Classification by the proposed EP-ELM classifier

After each facial image is turned into a feature matrix, we begin to recognize them with the proposed EP-ELM classifier. In this subsection, we will propose the EP-ELM classifier for classifying features based on P-ELM presented in the paper [39]. P-ELM added a polynomial with low degree in the FNN to reconcile the balance between the FNN and the polynomial. Our proposed EP-ELM improves the stability of P-ELM by comprising several P-ELM.

For a given set of sample data $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^{N} \subseteq \mathbb{R}^n \times \mathbb{R}^m$, the output of an FNN with L hidden nodes can be written by

$$f(\mathbf{x}) = \sum_{i=1}^{L}\beta_i G(\mathbf{a}_i, b_i, \mathbf{x}), \mathbf{x} \in \mathbb{R}^n, \tag{2.19}$$

where $\mathbf{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ are the hidden weights, $\beta_i \in \mathbb{R}^m$ is the output weight connecting the ith hidden node to the output node, and $G(\mathbf{a}_i, b_i, \mathbf{x})$ is the output of the ith hidden node with respect to the input $\mathbf{x}, i = 1, 2, \ldots, L.$

There are many computational hidden nodes for $G(\mathbf{a}_i, b_i, \mathbf{x})$, such as additive/RBF hidden nodes, multiplicative nodes, ridge polynomials, and so on. Additive and RBF hidden nodes are often used in applications. For additive hidden nodes, $G(\mathbf{a}_i, b_i, \mathbf{x})$ is taken to be

$$G(\mathbf{a}_i, b_i, \mathbf{x}) = g(\mathbf{a}_i \cdot \mathbf{x} + b_i); \tag{2.20}$$

while for the RBF hidden nodes, $G(\mathbf{a}_i, b_i, \mathbf{x})$ is taken to be

$$G(\mathbf{a}_i, b_i, \mathbf{x}) = g(b_i \parallel \mathbf{x} - \mathbf{a}_i \parallel), \tag{2.21}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is the activation function, $\mathbf{a}_i \cdot \mathbf{x}$ denotes the inner product of vectors $\mathbf{a}_i$ and $\mathbf{x}$ in $\mathbb{R}^n$, and $\| \cdot \|$ is the Euclidean norm on $\mathbb{R}^n$.

According to the theory of paper [42], an FNN plus a polynomial with a low degree can approximate the target functions better. Because FNN can handle scattered data that have rapid changes better, and the polynomial with a low degree can deal with the set of data with slow variations. So, in our proposed classifier, the approximate tool is taken to be as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{L} \beta_i G(\mathbf{a}_i, b_i, \mathbf{x}) + P(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n, \tag{2.22}$$

where $P(\mathbf{x})$ is a polynomial with low degree. In this paper, we take

$$P(\mathbf{x}) = x_1\gamma_1 + \cdots + x_n\gamma_n + \gamma_0, \tag{2.23}$$

where $\gamma_i \in \mathbb{R}^m$ and $\mathbf{x} = [x_1, \ldots, x_n]^\top$. Therefore, the network in (2.22) reconciles the balance between the FNN and the polynomial, which can integrate the advantages of FNNs and polynomials.

For a given set of training data $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N \subset \mathbb{R}^n \times \mathbb{R}^m$, the FNN with L hidden nodes in (2.22) best approximating these N training data means that there exist $\beta_i, \mathbf{a}_i, b_i, (i = 1, 2, \ldots, L), \gamma_k(k = 0, 1, 2, \ldots, n)$, and $\epsilon_j(j = 1, 2, \ldots, N)$ such that

$$\sum_{i=1}^{L} \beta_i G(\mathbf{a}_i, b_i, \mathbf{x}_j) + P(\mathbf{x}_j) = \mathbf{t}_j + \epsilon_j, \quad j = 1, 2, \ldots, N. \tag{2.24}$$

Equation (2.24) can be written compactly as

$$\mathbf{H}\beta = \mathbf{T} + \mathbf{E}, \tag{2.25}$$

where

$$\mathbf{H} = \begin{bmatrix} G(\mathbf{a}_1, b_1, \mathbf{x}_1) & \cdots & G(\mathbf{a}_L, b_L, \mathbf{x}_1) & x_{11} \cdots & x_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ G(\mathbf{a}_1, b_1, \mathbf{x}_N) & \cdots & G(\mathbf{a}_L, b_L, \mathbf{x}_N) & x_{N1} \cdots & x_{Nn} & 1 \end{bmatrix}, \tag{2.26}$$

$\beta = [\beta_1, \ldots, \beta_L, \gamma_1, \ldots, \gamma_n, \gamma_0]^\top, \mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_N]^\top$, and $\mathbf{E} = [\epsilon_1, \ldots, \epsilon_N]^\top$. Here, $\mathbf{H}$ is called the hidden-layer output matrix of the neural network.

In the P-ELM, the hidden weights $\mathbf{a}_i$ and b_i are randomly assigned motivated by [33]. That is, they need not be tuned during the process of learning and may be simply assigned with random values. Then Eq. (2.25) becomes a linear system and the output weight β can be estimated as

$$\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}, \tag{2.27}$$

where $\mathbf{H}^\dagger$ is the Moore–Penrose pseudo inverse of the hidden-layer output matrix $\mathbf{H}$ (see [43]). Since the hidden weights are randomly assigned, which may result in variations in different trials of simulations. In order to improve the performance of P-ELM, we will introduce the concept of ensemble in the P-ELM, called Ensemble of Polyharmonic ELM (EP-ELM). Now we propose EP-ELM classifier as follows:

Proposed EP-ELM Classifier 2.2. For the given set of training data $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N \subset \mathbb{R}^n \times \mathbb{R}^m$, let $f^{(r)}$ denote the r-th network to be learned by P-ELM as

$$f^{(r)}(\mathbf{x}) = \sum_{i=1}^{L} \beta_i^{(r)} G(\mathbf{a}_i^{(r)}, b_i^{(r)}, \mathbf{x}) + \gamma_1^{(r)} x_1 + \cdots + \gamma_n^{(r)} x_n + \gamma_0^{(r)}, \tag{2.28}$$

where $G(\mathbf{a}, b, \mathbf{x})$ and the number L of hidden nodes are given in prior, $\gamma_i^{(r)} = [\gamma_{i1}^{(r)}, \gamma_{i2}^{(r)}, \ldots, \gamma_{im}^{(r)}]^\top$, $i = 0, 1, \ldots, n$, $\mathbf{x} = [x_1, x_2, \ldots, x_n]^\top$, and the index r is used to denote which network the parameters belong to. Suppose that EP-ELM is composed of p P-ELMs. Set $r = 1$.

1. Randomly assign hidden-node parameters $(\mathbf{a}_i^{(r)*}, b_i^{(r)*})$, $i = 1, 2, \ldots, L$.
2. Compute the hidden-layer output matrix
$$\mathbf{H} = \begin{bmatrix} G(\mathbf{a}_1^{(r)*}, b_1^{(r)*}, \mathbf{x}_1) & \cdots & G(\mathbf{a}_L^{(r)*}, b_L^{(r)*}, \mathbf{x}_1) & x_{11} \cdots & x_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ G(\mathbf{a}_1^{(r)*}, b_1^{(r)*}, \mathbf{x}_N) & \cdots & G(\mathbf{a}_L^{(r)*}, b_L^{(r)*}, \mathbf{x}_N) & x_{N1} \cdots & x_{Nn} & 1 \end{bmatrix}, \tag{2.29}$$
where $\mathbf{x}_i = [x_{i1}, \ldots, x_{in}]^\top, i = 1, 2, \ldots, N$.
3. Calculate the output weight $\hat{\beta}^{(r)} : \hat{\beta}^{(r)} = \mathbf{H}^\dagger \mathbf{T}$, where $\hat{\beta}^{(r)} = [\beta_1^{(r)*}, \beta_2^{(r)*}, \ldots, \beta_L^{(r)*}, \gamma_1^{(r)*}, \ldots, \gamma_n^{(r)*}, \gamma_0^{(r)*}]^\top$ and $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_N]^\top$.
4. The function $f^{(r)}$ for the P-ELM classifier is
$$f^{(r)}(\mathbf{x}) = \sum_{i=1}^{L} \beta_i^{(r)*} G(\mathbf{a}_i^{(r)*}, b_i^{(r)*}, \mathbf{x}) + \gamma_1^{(r)*} x_1 + \cdots + \gamma_n^{(r)*} x_n + \gamma_0^{(r)*}. \tag{2.30}$$
5. Set $r = r + 1$. If $r \leq p$, then go to Step 1.
6. The function f for the EP-ELM classifier is
$$f(\mathbf{x}) = \frac{1}{p} \sum_{r=1}^{p} f^{(r)}(\mathbf{x}). \tag{2.31}$$

Remark 2.3 In the proposed EP-ELM classifier, the parameter p is set to denote the number of P-ELM in the EP-ELM. The purpose of ensemble of P-ELM is to obtain a relatively stable classifier. Because the hidden weights in P-ELM are randomly assigned, the recognition rate of the learned P-ELM will change in different trials. While from the law of large number in probability, when the number p of trials tends to infinity, the recognition rate of the learned EP-ELM will tend to a constant.

3 Experiments

In order to evaluate the performances of our proposed method for human face recognition, we carry out some face recognition experiments on five famous face databases, including Faces94, JAFFE, Sheffield, ORL, and Yale Databases. These databases are widely used to evaluate the performances of various face recognition algorithms. Before carrying on our experiments, we briefly introduce these human face databases firstly.

Faces94 Database was built by the University of Essex. It contains images of 152 distinctive individuals of various racial origins, mainly first year undergraduate students between 18 and 20 years old, and each individual has a series of 20 images. There are also some older staff members and students.

A Japanese female facial expression (JAFFE) contains 220 images of varying facial expressions sampled from 10 Japanese female models.

Sheffield face Database contains 564 images of 20 individuals with mixed race, gender, and appearance. Each person has successive pose images varying from left/right profiles to frontal views with small angular rotations. Furthermore, the size of those facial images is uniformly reduced to 112×92, and the background information is eliminated from input images so that only the central characteristics of the face are retained.

ORL face Database consists of 10 different images for each of the 40 distinct individuals. Each people is imaged in different facial expressions and facial details under varying lighting conditions at different times. All the pictures are captured with a dark background, and the individuals are in an upright and frontal position.

Yale face Database consists of 11 different gray-scale images for each of the 15 distinct persons. Each individual has different facial expression or configuration: with/without glasses, left/center/right light, happy/sad/normal, sleepy, surprised, and winking.

3.1 Setup

In all the experiments in this paper, all the features of training and testing image are produced by FDCT and 2DPCA. For each face database, we select randomly a part of images of each individual as training samples and the remaining as testing samples. That is, 45–50 % of each individual images are used as prototypes and the remaining as testing samples. Both the training and testing images are decomposed with FDCT at three scales and eight different angles. Then the approximate curvelet coefficients are dimensionality reduced with 2DPCA, trained and tested by EP-ELM classifier with $p = 50$.

The performance of the proposed method for human face recognition is carried out as follows:

1. For training human face images $\widetilde{A}_i(i = 1, \ldots, N_1)$, we firstly turn each image into gray level image and resize uniformly as $L_0 \times W_0$.
2. Compute the corresponding curvelet transform of each traning image and obtain the coarse subband coefficient matrix $A_i(i = 1, \ldots, N_1)$, whose size is $L_1 \times W_1$. The coarse coefficient matrices substitute the original images as training samples.
3. Calculate the covariance matrix D_l of training samples and obtain the orthonormal eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_d$ of D_l with respect to d largest eigenvalues.
4. Obtain d principal components for each sample image A_i by $\mathbf{y}_{ij} = A_i\mathbf{x}_j, j = 1, \ldots, d, i = 1, \ldots, N_1$.
5. Take $Y_i = [\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ldots, \mathbf{y}_{id}]^\top$ and convert it to be an input vector to train PFNN classifier, and denote the output $\mathbf{t}_i$ corresponding to the i-th people as $\mathbf{e}_i$, where $\mathbf{e}_i$ is an m-dimension unit vector.
6. Decompose the testing images $\widetilde{B}_j(j = 1, \ldots, N_2)$ by FDCT to obtain low frequency coarse matrices, then reduce the dimensionality by 2DPCA to get the corresponding principal components. Subsequently, these principal components are sent to the learned EP-ELM classifier for classification.

All the experiments are carried out in MATLAB R2010a environment running on a desktop with CPU AMD Athlon 2.7 GHZ and 1.75 GB RAM.

3.2 Experiments

Experiment 3.1. In this experiment, we compare the proposed method based on FDCT, 2DPCA, and EP-ELM classifier against various approaches on ORL Database and Yale Database. Table 1 shows the corresponding experimental results based on 40 principal components and 50 hidden nodes.

As observed from Table 1, the recognition rates of our proposed method outperform existing other face recognition methods.

Table 1 Comparative accuracy (%) for YALE and ORL face database

Methods	YALE	ORL
Stand eigenface [6]	76	92.2
Waveletface [8]	83.3	92.5
Waveletface + PCA [8]	84	94.5
Waveletface + LDA [19]	83.5	95.6
Waveletface + weighted modular PCA [22]	83.6	95
Curveletface + LDA [27]	83.5	95.5
Curveletface + PCA [27]	83.9	96.6
Waveletface + KAM [21]	84	96.6
Curveletface + PCA + LDA [27]	92	97.7
Proposed FDCT + 2DPCA + EP-ELM	98.7	99.47

The main reason for the better performance of our proposed method is that our method is a perfect combination of FDCT, 2DPCA, and EP-ELM. As we all know, FDCT can provide a better representation of intermediate dimensional structures than wavelet transform by combining ideas from geometry with the ideas from traditional multiscale analysis. And 2DPCA preserves a good effect on dimensionality reduction. What is more, EP-ELM classifier has a better performance than traditional classifiers, such as BP and *k*-nearest neighbor classifier in other methods.

Since the curveletface + PCA + LDA [27] has better performance than some existing methods except our proposed method, we have only compared our method with it in the remaining experiments for simplicity.

Experiment 3.2. In this experiment, we compare the recognition accuracy of our proposed method with the curvelet-based PCA + LDA approach on Sheffield and JAFFE Databases using varying number of principal components with 50 hidden nodes (Fig. 2). Furthermore, we also carry out the similar experiment on ORL and Face94 Databases with 100 hidden nodes. The experimental results are shown in Tables 2 and 3 and Fig. 3, respectively.

Table 2 Average recognition rates for Sheffield and JAFFE databases with different numbers of principal components

Component	Sheffield		JAFFE	
	PCA + LDA	Proposed	PCA + LDA	Proposed
5	93.89	95	93.94	99.73
10	96.11	96.5	94.62	100
15	97.78	97.5	94.84	100
20	99.44	99.5	96.58	100
25	99.44	100	98.76	100
30	98.88	100	99.46	100

Table 3 Average recognition rates for ORL and Face94 databases with different numbers of principal components

Component	ORL		Face94	
	PCA + LDA	Proposed	PCA + LDA	Proposed
5	79.12	93.05	91.35	97.43
10	89.16	95.56	92.17	98.69
20	98.33	97.40	97.28	100
30	97.5	98.70	99.29	100
40	96.67	100	99.29	100
50	97.52	100	99.29	100

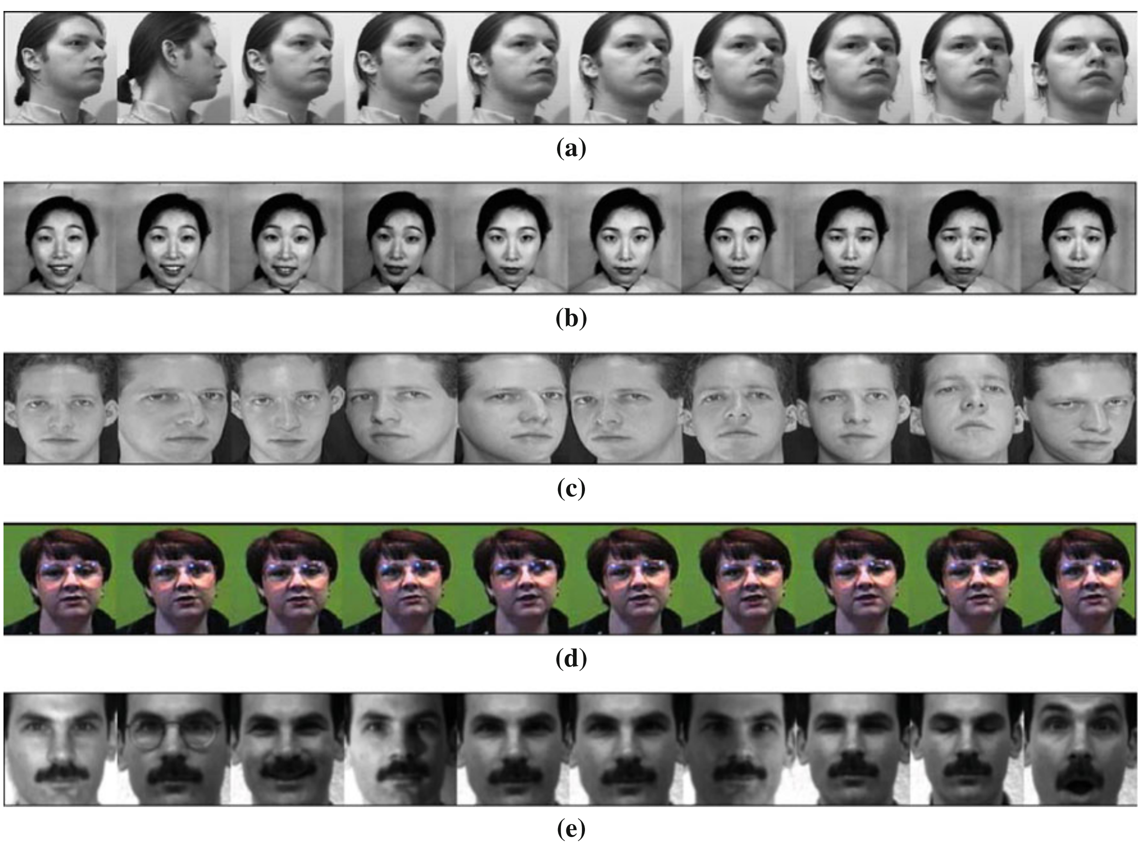


Fig. 2 Images from five datebases. **a** Sheffield face data. **b** JAFFE face data. **c** ORL face data. **d** Face94 face data. **e** YALE face data

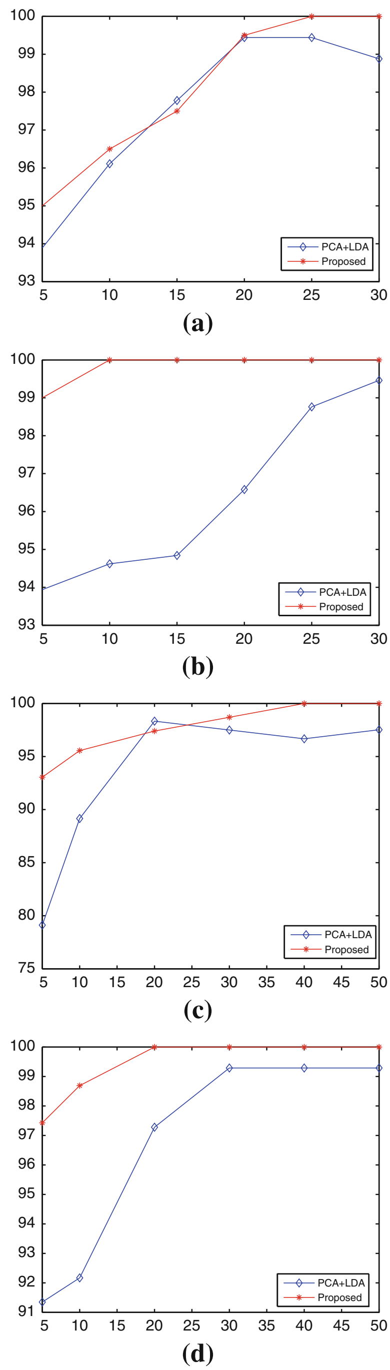


Fig. 3 Average recognition rate (*y*-axis) with number of principal components (*x*-axis). **a** Sheffield datebase. **b** JAFFE datebase. **c** ORL datebase. **d** Face94 database

Tables 2 and 3 show the comparison of the recognition rate of our proposed method with the curvelet-based PCA + LDA approach on the other four databases (Sheffield, JAFFE, ORL, and Face94 Databases) with varying number of principal components. In order to display those results of Tables 2 and 3 directly and clearly, we draw the corresponding curves in Fig. 3.

Table 4 Average recognition rates and training time (s) for Face94 databases

Component	Accuracy		Training time (s)	
	PCA + LDA	Proposed	PCA + LDA	Proposed
5	91.35	97.43	30.51	0.1341
10	92.17	98.69	30.74	0.1352
20	97.28	100	30.67	0.1361
30	99.29	100	31.55	0.1367
40	99.29	100	33.17	0.1335
50	99.29	100	33.36	0.1372

Table 5 Average recognition rates (%) of the proposed method with different numbers of hidden nodes and principal components on JAFFE database

Components	Number of hidden nodes						
	50	60	70	80	90	100	110
5	93.95	95.73	97.81	97.88	99.32	99.71	100
10	96.57	98.69	99.32	99.08	100	100	100
15	95.78	98.77	99.46	99.76	100	100	100
20	96.84	99.56	99.83	99.92	100	100	100
25	98.86	100	100	100	100	100	100

As observed from Tables 2 and 3 or Fig. 3, our proposed method has higher recognition rate than curvelet-based PCA + LDA approach except some points on Sheffield and ORL databases. Especially, the better performance of our proposed method is obvious on JAFFE Database and Face94 Database. The main reason is that the combination of FDCT and 2DPCA can provide more discriminable features than PCA + LDA, and EP-ELM has stronger ability of classification. The main reason for the different performances of human recognition methods on distinct databases is that there are different degrees of similarity of people in distinct databases. It is also worth mentioning that increasing the number of principal components does not necessarily increase the accuracy.

Experiment 3.3. In this experiment, we compare the recognition accuracy and training time of our proposed method with the curvelet-based PCA + LDA approach on Face94 Data base using varying number of principal components with 100 hidden nodes. The experimental results are shown in Table 4.

Table 4 clearly validates our claim that the proposed method has superior recognition rate and faster speed than PCA + LDA based approach. Especially, the training time of our proposed method is largely decreased than that of the curvelet-based PCA + LDA approach. The main reasons are that 2DPCA takes less time than PCA since the image matrix in 2DPCA does not need to be vectorized prior to feature extraction, and EP-ELM classifier has an extremely faster speed than *k*-nearest neighbor classifier. Therefore, our proposed method is more suitable for real-time applications.

Experiment 3.4. In this experiment, we give the performance of our proposed method with different principal components and numbers of hidden nodes varying from 50 to 110 on JAFFE Database. The experimental results are shown in Table 5.

From Table 5, we find that the recognition rate of our proposed method becomes better with the increase of hidden nodes under the same principal components. The main reason is that the performance of EP-ELM classifier becomes better with the increase of hidden nodes, and it tends to stable generally. It is also worth mentioning that increasing the number of principal components does not necessarily increase the accuracy.

Experiment 3.5. In this experiment, we compare the recognition rate of our proposed method with the popular classification method ELM and PELM on Sheffield Database to show the advantage of combination of FDCT, 2DPCA, and EP-ELM. For ELM and PELM methods, all images are directly classified, not pretreated in prior. The numbers of hidden nodes of ELM, PELM, and EP-ELM are all 50. The experimental results are shown in Table 6.

As observed from Table 6, we find that the best method is our proposed method, then P-ELM ,and the last is ELM. The main reasons are that P-ELM has better accuracy than ELM by adding a low-degree polynomial in the single-hidden-layer feedforward networks, while our proposed method improves separability of images by FDCT and 2DPCA, and EP-ELM has better stability with ensemble of several P-ELM classifiers. Therefore, our proposed method is a perfect combination of feature extraction and classifier.

Experiment 3.6. In this experiment, we want to further show the performance of EP-ELM by comparing our proposed method with ELM and P-ELM plus FDCT and 2DPCA. This experiment is carried out on Sheffield database with 50 hidden nodes. The experimental results are shown in Table 7.

Table 6 Recognition rates (%) of the proposed method with ELM and PELM on Sheffield database

Methods	Recognition rate
ELM	80.5
PELM	98.1
FDCT + 2DPCA + EP-ELM	99.5

Table 7 Recognition rates (%) of the proposed method with ELM and PELM plus FDCT and 2DPCA on Sheffield database

Methods	Recognition rate
FDCT + 2DPCA + ELM	82.3
FDCT + 2DPCA + PELM	98.7
FDCT + 2DPCA + EP-ELM	99.5

From Table 7, we find that the best method is still our proposed method, then FDCT + 2DPCA + PELM, and the last is FDCT + 2DPCA + ELM. The main reason is that P-ELM has better accuracy than ELM by adding a low-degree polynomial in the single-hidden-layer feedforward networks, and EP-ELM has better stability with ensemble of several P-ELM classifiers. While compared with Experiment 3.5, the performances of ELM and PELM have been improved by dealing with images using FDCT and 2DPCA in prior.

4 Conclusion

We have proposed a classifier, called EP-ELM, the main function of which is to reconcile the balance between the FNN and polynomial. Since the proposed EP-ELM classifier improved the accuracy of recognition by adding a polynomial with low degree in the FNNs, a method for human face recognition by combining FDCT, 2DPCA, and EP-ELM has been proposed. The FDCT is first used to compute features, and then, these features are dimensionality reduced by using 2DPCA to produce distinctive feature sets. Finally, these features are vectorized to be input to the EP-ELM classifier to analytically learn an optimal model. Experimental results corroborate our claim that the proposed method achieves improved recognition rate.

Acknowledgments The research was supported by the National Nature Science Foundation of China (No. 61101240, 61272023), the Zhejiang Provincial Natural Science Foundation of China (No. Y6110117), and the Science Foundation of Zhejiang Education Office (No. Y201122002).

References

1. Zhao W, Chellappa R, Rosenfeld A, Phillips PJ (2003) Face recognition: a literature survey. ACM Comput Surv 35(4):399–458
2. Lewandowski M, Makris D, Nebel JC (2010) Automatic configuration of spectral dimensionality reduction methods. Pattern Recogn Lett 31:1720–1727
3. Zhou H, Sadka AH (2011) Combining perceptual features with diffusion distance for face recognition. IEEE Trans Syst Man Cybern-Part C Appl Rev 41(5):577–588

4. Lei Z, Liao SC, Pietikäinen M, Li SZ (2011) Face recognition by exploring information jointly in space, scale and orientation. IEEE Trans Image Process 20(1):247–256
5. Al-Osaimi FR, Bennamoun M, Mian A (2012) Spatially optimized data-level fusion of texture and shape for face recognition. IEEE Trans Image Process 21(2):859–872
6. Kirby M, Sirovich L (1990) Application of the Karhunen–Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12(1):103–108
7. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In: Proceedings CVPR, 1991, pp 586–591
8. Feng GC, Yuen PC, Dai DQ (2000) Human face recognition using PCA on wavelet subband. J Electron Imaging 9(2):226–233
9. Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neur Netw 13(6):1450–1464
10. Lu J, Plataniotis KN, Venetsanopoulos AN (2003) Face recognition using LDA-based algorithms. IEEE Trans Neur Netw 14(1):195–200
11. Schoelkopf B, Smola AJ, Mueller KR (1997) Kernel principal component analysis. Lect Notes Comput Sci 1327:583–588
12. Yang J, Frangi AF, Yang J, Zhang D, Jin Z (2005) KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. IEEE Trans Pattern Anal Mach Intell 27(2):230–244
13. Yang J, Zhang D, Frangi AF, Yang JY (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Mach Intell 26(1):131–137
14. Zhang D, Zhou ZH (2005) 2DTPCA: two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing 69:224–231
15. Ghazali KH, Mansor MF, Mustafa MM, Hussain A (2007) Feature extraction technique using discrete wavelet transform for image classification. In: Proccding of 5th student conference on research and development, Dec. 2007, pp 1–4
16. Murtagh F, Starck JL (2008) Wavelet and curvelet moments for image classification: application to aggregate mixture grading. Pattern Recogn Lett 29:1557–1564
17. Jemai O, Zaied M, Amar CB, Alimi AM (2010) FBWN: an architecture of fast Beta wavelet networks for image classification. In: Proccding of the 2010 international joint conference on neural networks (IJCNN), Barcelona, July, 2010, pp 1–8
18. Hu HF (2011) Variable lighting face recognition using discrete wavelet transform. Pattern Recogn Lett 32:1526–1534
19. Chen JT, Wu CC (2002) Discriminant waveletfaces and nearest feature classifers for face recognition. IEEE Trans Pattern Anal Mach Intell 24(2):1644–1649
20. Huang K, Aviyente S (2008) Wavelet feature selection for image classification. IEEE Trans Image Process 17(9):1709–1719
21. Zhang BL, Zhang H, Ge JS (2004) Face recognition by applying wavelet subband representation and kernel associative memory. IEEE Trans Neur Netw 15(1):166–177
22. Zhao M, Li P, Liu Z (2008) Face recognition based on wavelet transform weighted modular PCA. In: Proceedings of the congress in image and signal processing, pp 589–593
23. Donoho DL, Duncan MR (2000) Digital curvelet transform: strategy, implementation and experiments. In: Proceedings of the SPIE, vol 4056, pp 12–30
24. Candès EJ, Demanet L, Donoho DL, Ying L (2006) Fast discrete curvelet transforms. Multiscale Model Simul 5(3):861–899
25. Cheng SS, Yang YB, Li YW (2007) Study on classification based on image fusion with curvelet transform. In: Proceedings of the SPIE, vol 6790, p 67900G. doi:10.1117/12.742231
26. Sun AR, Tan YH (2007) Hyperspectral data classification using image fusion based on curvelet transform. In: Proceedings of the SPIE, vol 6787, p 678710. doi:10.1117/12.750049
27. Mandal T, Wu QMJ, Yuan Y (2009) Curvelet based face recognition via dimension reduction. Signal Process 89(3):2345–2353
28. Rowley HA, Baluja S, Kanade T (1998) Neural network-based face detection. IEEE Trans Pattern Anal Mach Intell 20(1):23–38
29. Meng JE, Wu SQ, Lu JW, Toh HL (2002) Face recognition with radial basis function (RBF) neural networks. IEEE Trans Neur Netw 13(3):697–710
30. Ma L, Khorasani K (2004) Facial expression recognition using constructive feedforward neural networks. IEEE Trans Syst Man Cybern Part B Cybern 34(3):1588–1595
31. Diego IM, Serrano A, Conde C, Cabello E (2010) Face verification with a kernel fusion method. Pattern Recogn Lett 31:837–844
32. Sing JK, Thakur S, Basu DK, Nasipuri M, Kundu M (2009) High-speed face recognition using self-adaptive radial basis function neural networks. Neural Comput Appl 18:979–990
33. Huang GB, Zhu Q, Siew C (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
34. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. Int J Mach Learn Cybern 2(2):261–271
35. Wang XZ, Dong CR (2009) Improving generalization of fuzzy if-then rules by maximizing fuzzy entropy. IEEE Trans Fuzzy Syst 17(3):556–567
36. Wang XZ, Dong LC, Yan JH (2011) Maximum ambiguity based sample selection in fuzzy decision tree induction. IEEE Trans Knowl Data Eng. doi:10.1109/TKDE.2011.67
37. Pan C, Park DS, Yang Y, Yoo HM (2012) Leukocyte image segmentation by visual attention and extreme learning machine. Neural Comput Appl 21(6):1217–1227
38. Pan C, Park DS, Lu HJ, Wu XP (2012) Color image segmentation by fixation-based active learning with ELM. Soft Comput 16(9):1569–1584
39. Zhou ZH, Zhao JW, Cao FL (2012) Surface reconstruction based on extreme learning machine. Neural Comput Appl. doi:10.1007/s00521-012-0891-8
40. Wang XZ, Chen AX, Feng HM (2011) Upper integral network with extreme learning mechanism. Neurocomputing 74(16):2520–2525
41. Wu J, Wang ST, Chung FL (2011) Positive and negative fuzzy rule system, extreme learning machine and image classification. Int J Mach Learn Cybern 2(4):261–271
42. Duchon J (1977) Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: Schempp W, Zeller K (ed) Constructive theory of functions of several variables, No. 571 in Lecture Notes in mathematics. Springer, Berlin, pp 85–100
43. Rao CR, Mitra SK (1971) Generalized inverse of matrices and its applications, vol 15. Wiley, New York