International Journal of Innovative Computing, Information and Control
Volume **9**, Number **9**, September **2013**
ICIC International ©2013 ISSN 1349-4198
pp. 3839–3849

GENDER CLASSIFICATION BASED ON EVOLUTIONARY EXTREME LEARNING MACHINE

Xiaodong Li, Wei Jiang, Weijie Mao, Jian Li and Kai Chen

State Key Laboratory of Industrial Control Technology
Institute of Cyber-Systems and Control
Zhejiang University
No. 38, Zheda Road, Yuquan Campus, Hangzhou 310027, P. R. China
hzxiaodong22@163.com; { jiangwei; wjmao; jianli; kaichen }@iipc.zju.edu.cn

Received August 2012; revised January 2013

Abstract. *We introduce an evolutionary extreme learning machine (E-ELM) and incremental bilateral two-dimensional principal component analysis (IB2DPCA) for gender classification, a novel algorithm for face feature extraction, which directly extracts the proper features from image matrices. There are both ways of contributions in this paper. First, we proposed the IB2DPCA-EELM, the IB2DPCA being our new idea. Second, we showed that it can be another feasible tool for E-ELM of gender classification. The proposed method is based on curvelet image decomposition of human faces and a subband that exhibits a maximum standard deviation is dimensionally reduced using an improved dimensionality reduction technique. Using IB2DPCA learns from a small pool of labeled data and then iteratively selects the most informative samples from the unlabeled set to increasingly improve the classifier accuracy. Other notable contributions of the proposed algorithm include significant improvements in gender classifier, extreme reduction in training time and minimal dependence on the number of prototypes. Extensive experiments are performed using challenging databases and results produce better performance compared with other algorithms.*
Keywords: Evolutionary extreme learning machine, Gender classification, PCA

1. **Introduction.** Humans recognize faces with natural eyes, nose, etc.; however, gender identification is very challenging since gender belongs to a class of natural objects. Despite very complex, the human optical system efficiently discriminates and identifies gender through facial features. It is advantageous for computer-aided gender recognition to deal with large number of facial information ability, but wetware has limited memory and easily loses some sources. Computer-aided are influenced by some factors such as aging, viewpoint variations, illumination, skin color, hair style and noised ambient.

Gender classification has two main approaches. The first approach is based on the geometrical feature approach. Burton et al. extracted point-to-point distances from 73 points on face images and used discriminant method as a classifier [1]. Brunelli and Poggio [2] extracted 16 geometric features such as eyebrow thickness and pupilto-eyebrow distance and used HyperBF networks as a classifier. The second approach is the appearance-based approach which uses a whole face image. Cottrell and Metcalfe [3] reduced the dimension of whole face images by auto-encoder network and classified gender based on the reduced input features. Tamura et al. [4] used a neural network and showed that even very low resolution face image can be utilized for gender classification. Gutta et al. [5] used the mixture of experts with entireness of radial basis functions (RBF) networks and a decision tree as a gating network. Moghaddam and Yang [6] showed that support vector machines (SVMs) did better than other classifiers such as RBF networks, Fisher linear discriminant, nearest neighbor. Jain and Huang [7] extracted whole features by independent component analysis (ICA) and classified it with linear discriminant analysis (LDA). Chen et al. [8], a row-preserved 2DPCA is first applied to reduce the column dimension of a 2D image and a column-preserved 2DPCA is applied to reduce the row dimension. Jabid et al. [9,10], different edge responses around each pixel are encoded to distinguish all kinds of textures available in facial images. Kim et al. [11] used Gaussian process classifiers (GPCs) which are Bayesian kernel classifiers. Hadid and Pietikäinen [12] propose and study an approach for spatiotemporal face and gender recognition from videos using an extended set of volume LBP features and a boosting scheme. As analyzed by Balci and Atalay [13], the paper focuses on classification of face images according to their gender using Principal Component Analysis (PCA) and pruning Multi-Layer Perceptrons (MLP).

A gender category system improving robust speed and accuracy through dimensionality reduction techniques have been developed. At first, it is very important to obtain facial feature information. Kirby and Sirovich [14] represented human faces as a linear combination of weighted eigenvectors using principal component analysis (PCA). PCA based gender recognition systems suffer from poor discriminatory power and high computational load, therefore, to eliminate these limitations Bartlett et al. [17] proposed the use of independent component analysis (ICA). In [19], Linear Discriminant Analysis (LDA)-based classification engines have proved to provide comparable accuracy to other algorithms, while being significantly faster. An eigenspace based adaptive approach that utilizes a specific kind of genetic algorithm called evolutionary pursuit (EP) [20]. Bach and Jordan [22] proposed the use of kernel Hilbert space for ICA to adaptively generate nonlinear functions and to obtain a robust algorithm with regards to variations in source density, degree of non-Gaussianity, and presence of outliers. A kernel machine-based discriminant analysis method [23] that deals with nonlinear distribution of the face pattern has also been used for improved representation of faces under variations in perspective, lighting and facial expression. The use of support vector machines (SVM) [26] improves the discriminative power of extracted face features.

This paper combines incremental bidirectional two dimensional principal component analysis (IB2DPCA) and evolutionary extreme learning machine (E-ELM) to enhance the gender identification performance. Discriminative feature sets are generated using IB2DPCA to ascertain classification accuracy. We do some experiments to prove superiority of our proposed algorithm outperforming existing state-of-the-art methods. A few sample images of one of the subject from ORL [29] and YALE [30] database are shown in Figures 1 and 2, respectively.

FIGURE 1. Sample images in the ORL database

FIGURE 2. Sample images in the YALE database

2. **B2DPCA.** Karhunen-Loeve expansion, also known as principal component analysis (PCA), is a data representation technique widely used in pattern recognition and compression schemes. Wiskott et al. [25] employed a bunch graph matching technique to overcome limitations and flaws of linear PCA. Yang et al. [28] proposed two dimensional PCA for image representation. Pioneering work by Kirby and Sirovich [14] used PCA for enhanced representation of facial images; however, PCA fails to capture minor variance unless they are explicitly accounted in the training data. As opposed to PCA, 2DPCA is based on 2D matrices rather than 1D. Therefore, image matrix does not need to quantify prior to feature extraction. Instead an image covariance matrix is directly computed using original image matrices.

3. Proposed Gender Classification Algorithm.

3.1. Curvelet transform based gender recognition.

Curvelet transform [27] tends to represent smooth objects with discontinuity along a general curve. Because of its improved directional elements and highly anisotropic scaling property, it can provide an equally good (or superior) basis for face gender recognition compared with wavelets. To validate this empirically, different experiments have been performed. The gender recognition system has two stages: training stage and classification stage. During training, curvelet transform is applied to decompose the images into curvelet subbands. It has been stated repeatedly that curvelet transform is very good at representing objects with 'curve-punctuated smoothness' [31], i.e., objects which display smoothness except for discontinuity along a general curve. Facial images with edges are good examples of this kind of objects. Say, the images are represented by eight bits, i.e., 256 gray levels. In such an image two very close regions can have differing pixel values. Such a gray scale image will then have a lot of "curved edges" (which are typically curved in facial images) and consequently the curvelet transform will capture this crucial edge information. Extracted curvelet coefficients can then be directly fed to a classifier to perform the identification task. An image of size, say 64×64, when decomposed using curvelet transform scale 3 (coarse, fine, finest) and angle 8 will produce approximate (coarse) subband of size 21×21 and 24 detailed coefficients of slightly larger size. Working with such large number of feature vectors is extremely cost. Hence, it is of utmost importance to determine a representative set of features. Selection of subbands has been done on the basis of the amount of variance they account for. The coarse coefficients, (as they account for the maximum variance and contain maximum energy of the image-data) and a set of fine coefficients with next highest amount of total variance have been chosen. Thus selecting the appropriate curvelet faces reduces the dimension of the original image (represented as a vector) from 1×4096 to 1×1258. Yet a feature vector of length 1×1258 needs some kind of dimensionality reduction. Traditional dimensionality reduction methods like PCA/LDA can be applied on the selected subbands thereafter and an efficient and representative dataset can be produced. A simple Nearest Neighbor (NN) classifier is used to carry out the recognition task. Unless mentioned otherwise, a scale value of 3 has been used throughout this work, in order to weigh a balance between the speed and performance of the system. It has been observed that higher scale (4 or 5) has shown only marginal improvement or even no improvement in terms of recognition accuracy.

Figure 3 shows the curvelet subbands for a sample face taken from YAL dataset. These curvelet decomposed images are named "Curvelet Faces". Digital curvelet transform (scale = 2, angle = 8) is applied on the original image of size 64×64. Eight detailed curvelet coefficients (four of them are of size 34×88 and rest four are of size 68×36) are

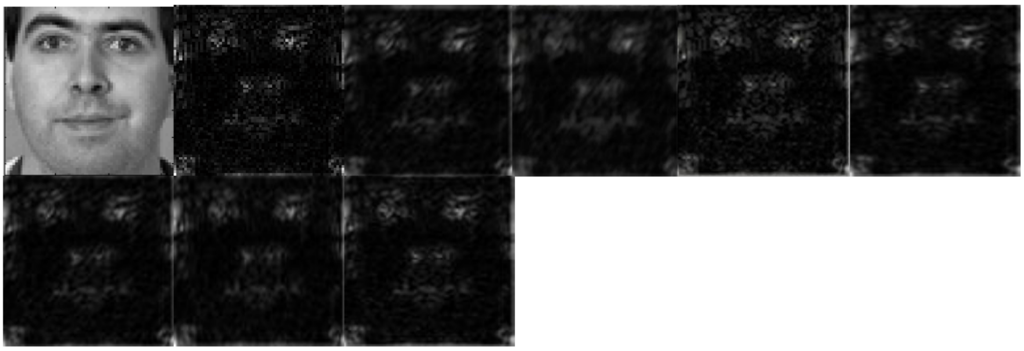

FIGURE 3. Curvelet transform of faces: first image in the first row is the original image and others are detailed coefficients at eight angles; all the images are resized to the same dimension for the purpose of illustration only.

for eight different angles. Figure 3 shows these eight curvelet coefficients. The images are of different size, but have been resized to the same size only for the sake of good show.

3.2. **Incremental bidirectional two-dimensional principal component analysis (IB2DPCA).** Karhunen-Loeve expansion, also known as principal component analysis (PCA), is a data representation technique widely used in pattern recognition and compression schemes. However, PCA fails to capture minor variance unless they are explicitly accounted in the training data. As opposed to PCA, 2DPCA is based on 2D matrices rather than 1D vectors. Therefore, image matrix does not need to quantify prior to feature extraction. Instead an image covariance matrix is directly computed using original image matrices.

Although B2DPCA reduces the dimension of image feature matrix, it ignores the problem that different feature vectors influence the recognition accuracy differently. B2DPCA-based gender recognition systems are difficult to reduce computational cost and memory requirement load. To improve this drawback, an incremental approach is usually adopted. In this paper, the left-multiplying unilateral two-dimension principal component analysis (LU2DPCA) and right-multiplying unilateral two-dimension principal component analysis (RU2DPCA) are proposed to extract face feature by directly projecting the image matrix. Here the proposed algorithms summarize the method described as follows.

First, it will improve the mean value:

$$\overline{A}' = \frac{1}{n+1}(n\overline{A} + A_{n+1}) \tag{1}$$

Then update the set of eigenvectors by adding a new vector and applying a rotational transformation. In order to do this, we first compute the orthogonal residual vector $J_{n+1} = (Ua_{n+1} + \overline{A}) - A_{n+1}$ and normalize it to obtain $\hat{J}_{n+1} = h_{n+1}/||h_{n+1}||_2$ for $\|h_{n+1}\|_2 > 0$ and $\widehat{h}_{n+1} = 0$. Matrix of eigenvectors $U' \in \mathbb{R}^{m\times(k+1)}$ is computed by

$$U' = \left[U\widehat{h}_{n+1}\right] R \tag{2}$$

where $R \in \mathbb{R}^{(k+1)\times(k+1)}$ is a rotation matrix. U is a set of eigenvector $U = [u_j]$ $j = 1, \ldots, p$. R is the solution of the eigen-problem of the following form:

$$DR = R\Lambda' \tag{3}$$

We construct $D \in \mathbb{R}^{(k+1)\times(k+1)}$ as

$$\frac{n}{n+1}\begin{bmatrix} \Lambda & 0 \\ 0^T & 0 \end{bmatrix} + \frac{1}{n}\begin{bmatrix} aa^T & \gamma a \\ \gamma a^T & \gamma^2 \end{bmatrix} \tag{4}$$

where $\gamma = \widehat{h}_{n+1}^T(A_{n+1} - \overline{A})$ and $a = U^T(A_{n+1} - \overline{A})$. To build D have many different ways. However, effectively updating mean is only allowed for this way.

Let X denote q dimensional unitary column vector. To project a $p \times q$ image matrix A to X, linear transformation $Y = AX$ is used which results into a p dimensional projected vector Y. The total scatter of the projected samples is determined to measure the discriminatory power of the projection vector X. The total scatter is characterized by the trace of S_x, i.e., covariance matrix of the projected feature vectors; $J(X) = tr(S_x)$ where $tr()$ represents the trace of S_x:

$$S_x = E[Y - E(Y)][Y - E(Y)]^T = E[(A - EA)X][(A - EA)X]^T, \tag{5}$$

$$tr(S_x) = X^T[E(A - EA)^T(A - EA)]X. \tag{6}$$

$G_t = E[(A - EA)^T(A - EA)]$ is a non-negative $q \times q$ image covariance matrix. If there are n training samples, the αth image sample is denoted by $p \times q$ matrix A_α:

$$G_t = {}^1\!/_n \sum_{\alpha=1}^{n} (A_\alpha - \hat{A})^T(A_\alpha - \hat{A}), \tag{7}$$

$$J(X) = X^T G_t X \tag{8}$$

maximizing the generalized total scatter criterion $J(X)$ is called the optimal projection axes. X_{opt} denotes a collection of n orthonormal eigenvectors $X_1, \ldots, X_n$ of G_t corresponding to n largest eigenvalues. Hence, dimensionality of every image A_α is reduced by right-multiplying and left-multiplying the image with optimal projection axes as $X_{opt}^T A_\alpha X_{opt}$.

The proposed technique closely follows the work of [28] and generates an image covariance matrix $G_{t\alpha}$ and further optimizes it exploiting optimal project axes. The dimensionality reduction algorithm operates independently along row and column directions in order to better preserve the neighborhood relationship and to generate distinctive feature sets. Once optimal projection axes $X_{opt\alpha}$ is calculated, the dimensionality of every image A_α is reduced along its columns to generate new image sets A_β using Equation (12). The newly generated image sets are subsequently treated as a new database and a latest image covariance matrix $G_{t\beta}$ and optimal projection axes $X_{opt\alpha}$ are evaluated. Finally, every new image A_β is pre-multiplied by $X_{opt\beta}$. Hence, unlike traditional B2DPCA, a two incremental approach is adopted in our proposed IB2DPCA algorithm and have a higher image recognition. The proposed algorithm presents the major implementation steps for computation of IB2DPCA in Algorithm 1 for clarity:

$$A_\beta = A_\alpha X_{opt\alpha}, \quad A_\theta = X_{opt\beta}^T A_\beta. \tag{9}$$

Algorithm 1: The proposed IB2DPCA algorithm.

Input: Input images A_n, $n \in \mathbb{Z}^+$, new input images A_{n+1}, $n \in \mathbb{Z}^+$

Output: $n \times n$ output matrix A_θ

1: Compute non-negative image covariance matrix by $G_{t\alpha} = \frac{1}{n}\sum_{\alpha=1}^{n}\left(A_\alpha - \hat{A}_\alpha\right)^T(A_\alpha - \hat{A}_\alpha)$ where $\hat{A}_\alpha = \frac{1}{n}\sum_{i=1}^{n} A_i$, update the mean $\hat{A}'_\alpha = \frac{1}{n+1}\sum_{i=1}^{n}(i\hat{A}_\alpha + A_{i+1})$, $\hat{A}_\alpha = \hat{A}'_\alpha$.

2: The trace of the covariance matrix characterizes total scatter $J(X) = X^T G_{t\alpha} X$.

3: $X_{opt\alpha} = \{X_1, X_2, \ldots, X_n\}$ where X_i represents a principal orthogonal residual vector.

4: Reduce dimensionality along columns: $\hat{A}_\beta = \hat{A}_\alpha X_{opt\alpha}$.

5: The image covariance matrix for A_β is determined as $G_{t\beta} = \frac{1}{n}\sum_{\beta=1}^{n}\left(A_\beta - \hat{A}_\beta\right)^T(A_\beta - \hat{A}_\beta)$ where $\hat{A}_\beta = \frac{1}{n}\sum_{i=1}^{n} A_i$, update the mean $\hat{A}'_\beta = \frac{1}{n+1}\sum_{i=1}^{n}(i\hat{A}_\beta + A_{i+1})$, $\hat{A}_\beta = \hat{A}'_\beta$.

6: Using total scatter criterion (step-2), compute $X_{opt\beta}$ comprising of n largest eigenvectors.

7: Row-wise dimension reduction by $A_\theta = X_{opt\beta}^T \hat{A}_\beta$.

3.3. **Evolutionary extreme learning machine.** Extreme learning machine (ELM) was proposed in Huang et al. [15,16]. In this section, a hybrid approach named E-ELM using DE and MP generalized in verse is proposed. Firstly, we randomly generate the population. Each individual in the population is composed of a set of input weights and hidden biases [21]:

$$\theta = [W_{11}, W_{12}, \ldots, W_{1k}, W_{21}, W_{22}, \ldots, W_{2k}, \ldots, W_{n1}, W_{n2}, \ldots, W_{nk}, \ldots, b_1, b_2, \ldots, b_k]$$

All W_{ij} and b_j are randomly initialized within the range of $[-1, 1]$.

Secondly, for each individual (a set of weights and biases), the corresponding output weights are analytically computed by using the MP generalized in verse as done in ELM instead of any iterative tuning [21].

Then the fitness of each individual is evaluated. Suppose that the cost function is root mean squared error (RMSE):

$$\sqrt{\frac{\sum_{j=1}^{n}\left\|\sum_{i=1}^{k}\beta_i g(W_i \cdot X_j + b_i - t_j)\right\|_2^2}{m \times n}} \tag{10}$$

In the normal case, the RMSE on the whole training dataset is used as the fitness. However, it may cause the networks to over-fit it. In BP, a validation dataset, which normally has no overlap with the training dataset, is used to avoid the over-fitting. In addition, as the β is the minimum norm least square solution to the training dataset already, the training error with respect to different individual should be very similar. Therefore, in order to save time, we set the fitness to the RMSE on the validation set only instead of the whole training set as used in [24].

After the fitness of all individuals in the population is calculated, we apply the three steps of DE: mutation, crossover and selection. As analyzed by Bartlett [18], networks tend to have better generalization performance with smaller weights. Therefore, to further improve the generalization performance, we add one more criteria into the selection: the norm of output weights $\|\beta\|$. According to our new selection strategy, when the difference of the fitness between different individuals is small, the one resulting in smaller $\|\beta\|$ is

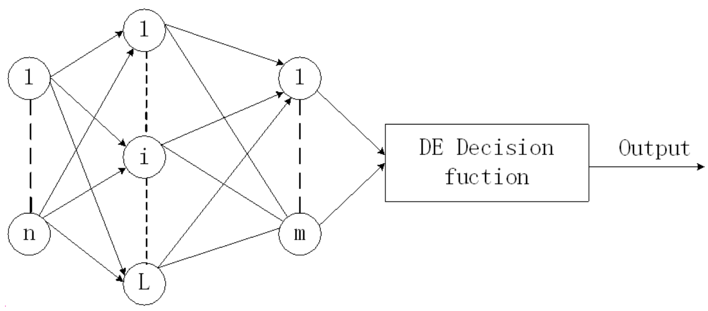

FIGURE 4. An evolutionary extreme learning machine network

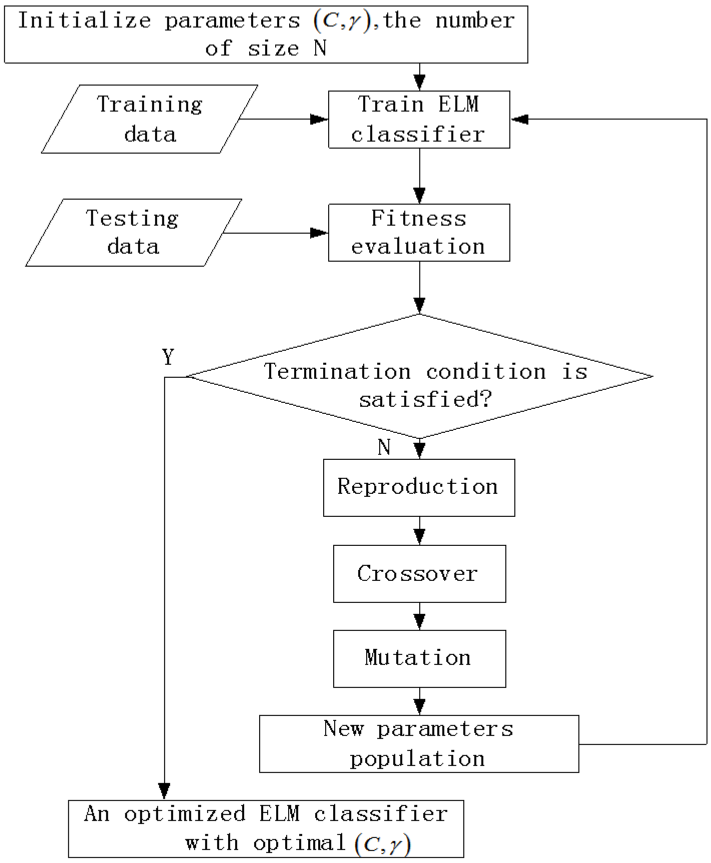

FIGURE 5. Architecture of E-ELM with Gaussian RBF kernel

selected. The determination of new population $\theta_{i,G+1}$ can be described as follows:

$$\theta_{i,G+1} = \begin{cases} \mu_{i,G}, & \text{if } f(\theta_{i,G}) - f(\mu_{i,G}) > \varepsilon f(\theta_{i,G}), \\ \mu_{i,G}, & f(\theta_{i,G}) - f(\mu_{i,G}) < \varepsilon f(\theta_{i,G}) \text{ and } \left\|\beta_{\mu_{i,G}}\right\| < \left\|\beta_{\theta_{i,G}}\right\|, \\ \theta_{i,G}, & \text{else,} \end{cases} \tag{11}$$

where $f(\cdot)$ is a fitness function (validation error) and ε is a preset tolerance rate. Once the new population is generated, repeat the same DE process until the goal is met or a preset maximum learning epochs is finished. The construction of E-ELM is shown in Figure 4.

Gender classification is a binary categorization problem. A hyper-plane is designed in such a way that it will leave more on either side, so that data on both sides can move a bit more freely with less risk of causing an error. A hyper-plane is selected with generalization performance of the classifier – the capability of the classifier, designed using the training data set, to operate satisfactory with data outside this set. An E-ELM uses an evolutionary algorithm to search ELM hyper-parameters with the input of the training data set. The evolutionary algorithm will be implemented until the E-ELM can give good generalized ability for testing data set. The procedure of E-ELM is shown in Figure 5.

4. **Experiment and Evaluation.** Extensive experiments are performed using our proposed method on distinguishing face databases: ORL [29] and YALE [30]. These are the most commonly used databases to evaluate and compare the performance of various face recognition algorithms.

4.1. **Comparative results.** These are the most commonly used databases to evaluate and compare the performance of various gender recognition algorithms. Gender classification experiments are achieved using our proposed method on two distinguishing face databases: ORL and YALE.

All images are zoomed in a 2 fold dimension reduction in gray image, but they can adapt to color image, just converted from RGB to gray level image. In all databases, a certain percentage of images of each subject are used as prototypes and the rest of images for testing purposes. Using curvelet transform analyzes both the testing and training image sets. Approximate curvelet coefficients are dimensionally reduced using IB2DPCA algorithm, however, trained and tested algorithm using E-ELM.

In Table 1, it is shown the experimental error rate with different method. The E-ELM classifier was tested with Gaussian RBF Kernel in order to explore the space of possibilities and performance. A Gaussian RBF kernel was found to perform the best (in terms of error rate). In the SVM experiment, a Gaussian RBF kernel was found to perform the second (in terms of error rate). The number of radial bases for classical RBF networks was heuristically set to 20 prior to actual training and testing number. Quadratic, Linear and Fisher classifiers were achieved with Gaussian distributions and in each case a likelihood ratio test was used for classification.

See Table 2 for detail on the high resolution and low resolution of error rate for gender identification. The human error rates obtained in our study are tabulated in Table 2. Comparing Tables 1 and 2, it is clear that E-ELM also perform better than humans with both low and high resolution faces. These results show that the classification of gender is more accurately modeled by E-ELM than any other classifier. It is not surprising that human face subjects achieve better with high resolution images than with low resolution images. E-ELM performance, however, was mostly unaffected by the change in resolution.

In Table 3, it is shown the average access time required by the different methods for the gender classification of a given face image. This time does not include the time required for the face detection/cropping, beard and eyes detection. It includes just the time required for the face analysis (feature extraction and gender classification). The

Table 1. Experimental error rate with different method

Classifier	Error Rate		
	Overall	Male	Female
E-ELM with Gaussian RBF kernel	2.86%	1.07%	3.98%
SVM with Gaussian RBF kernel	3.06%	2.01%	4.23%
Classical RBF	7.57%	6.48%	8.26%
Quadratic classifier	10.12%	9.02%	10.56%
Fisher linear discriminant	12.98%	11.87%	13.13%
Nearest neighbor	26.02%	25.23%	27.66%
Linear classifier	57.04%	55.67%	58.11%

Table 2. Human error rates

Gender of Human Subject	Error Rate	
	High Resolution	Low Resolution
Male	2.07%	20.33%
Female	1.22%	20.55%
Combined	1.66%	20.57%

TABLE 3. Processing times of different methods

Method		SVM	PCA	SVM+PCA	Ada-Rect	IB2DPCA+EELM
Time [mseg]	ORL	10.25	620	200	0.211	0.112
	YALE	9.88	610	185	0.288	0.119

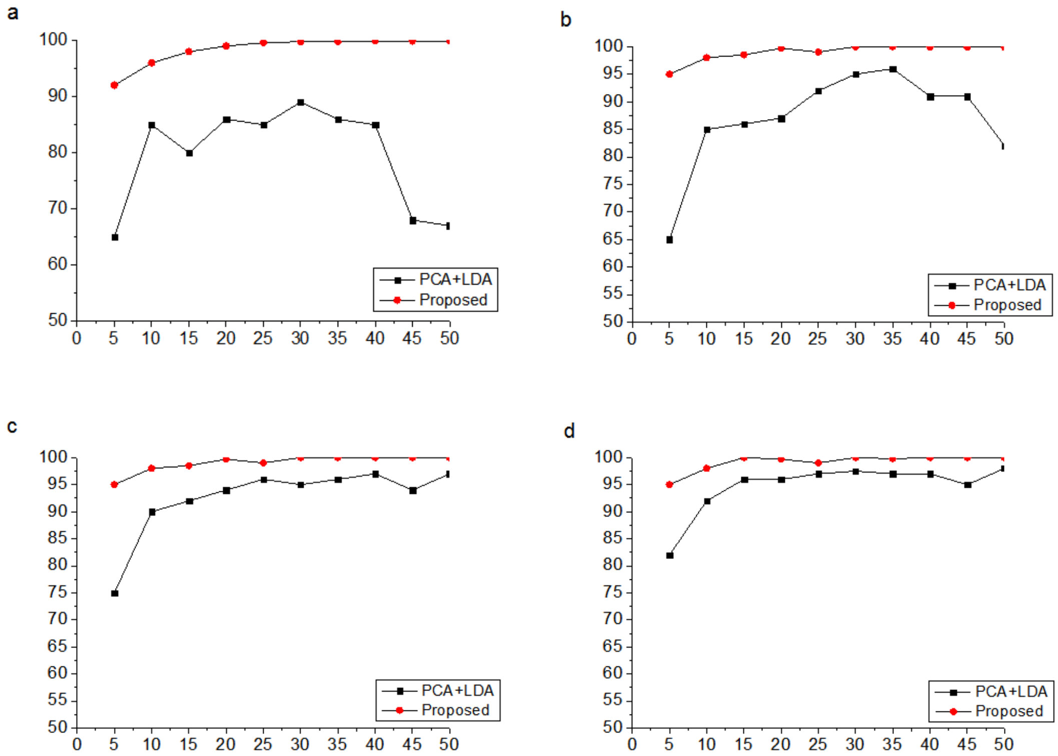

FIGURE 6. Average recognition rate (y-axis) vs. number of principal components (x-axis) for ORL database at varying prototypes. (a) 30% prototypes, (b) 40% prototypes, (c) 60% prototypes and (d) 70% prototypes.

experimental results reported here are based on the average of 120 independent trials. All simulations have been conducted within the MATLAB R2008 environment on an Intel(r) i3 2.13GHZ CPU and 2G RAM personal computer.

In addition to improved accuracy, the proposed algorithm is also independent of the number of prototypes. Recognition rates obtained for ORL database at 30%, 40%, 60% and 70% prototypes are plotted in Figure 6 (y-axis denotes the accuracy and x-axis denotes the number of principal components). In order to avoid within-scatter matrix singular cases, authors in [31,32] extracted curvelet coefficients at four scales. In contrast, our proposed method can reduce data dimension efficiently and is free of the singularity issues, i.e., independent of the scales of curvelet decomposition that radically degrade precision of the PCA+LDA based method.

5. **Conclusions.** In this paper, we propose an efficient human gender recognition method based on curvelet feature subspace of human face. The curvelet transform is used to compute sparse features with improved directionality in higher dimension. These sparse features reduce dimension using IB2DPCA to create unique characteristics sets. These

features are input to an E-ELM to analytically learn an optimal model. Experimental results demonstrate that our method achieves recognition at an essentially faster rate than existing methods. Additionally, our proposed method is independent of the number of prototypes used for training, scales of curvelet decomposition and the number of hidden units. Experimental results demonstrate that E-ELM generally achieves higher generalization performance than other algorithms such as SVM, BP, GALS, and the original ELM.

In future, we would like to do more research in generalized localized features extraction to improve better on robust recognition accuracy and classification speed, especially in intricate background, and extend this framework to multi-class problems (age, and race classification), and finding out a way of selecting different kinds of features together.

REFERENCES

[1] A. M. Burton, V. Bruce and N. Dench, What's the difference between men and women? Evidence from facial measurements, *Perception*, vol.22, pp.153-176, 1993.

[2] R. Brunelli and T. Poggio, HyperBF networks for gender classification, *DARPA Image Understanding Workshop*, 1992.

[3] G. Cottrell and J. Metcalfe, *EMPATH: Face, Emotion, and Gender Recognition Using Holons Advances in Neural Information Processing Systems*, MIT Press, 1991.

[4] S. Tamura, H. Kawai and H. Mitsumoto, Male/female identification from 8×6 very low resolution face images by neural networks, *Pattern Recognition*, vol.29, no.2, pp.331-335, 1996.

[5] S. Gutta, J. R. J. Huang, P. Jonathon and H. Wechsler, Mixture of experts for classification of gender, ethnic origin, and pose of human faces, *IEEE Trans. on Neural Network*, vol.11, no.4, pp.948-960, 2000.

[6] B. Moghaddam and M. Yang, Learning gender with support faces, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.5, pp.707-711, 2002.

[7] A. Jain and J. Huang, Integrating independent components and linear discriminant analysis, *Proc. of the 6th International Conference on Automatic Face and Gesture Recognition*, 2004.

[8] Y.-W. Chen, R. Xu and A. Ushikome, Serially-connected dual 2D PCA for efficient face representation and face recognition, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4367-4372, 2009.

[9] T. Jabid, M. H. Kabir and O. Chae, Local directional pattern (LDP) for face recognition, *International Journal of Innovative Computing, Information and Control*, vol.8, no.4, pp.2423-2437, 2012.

[10] M. I. Razzak, B. A. Elmogy, M. K. Khan and K. Alghathbar, Efficient distributed face recognition in wireless sensor network, *International Journal of Innovative Computing, Information and Control*, vol.8, no.4, pp.2811-2822, 2012.

[11] H.-C. Kim, D. Kim and Z. Ghahramani, Appearance-based gender classification with Gaussian processes, *Pattern Recognition Letters*, vol.27, pp.618-626, 2006.

[12] A. Hadid and M. Pietikäinen, Combining appearance and motion for face and gender recognition from videos, *Pattern Recognition*, vol.42, no.11, pp.2818-2827, 2009.

[13] K. Balci and V. Atalay, PCA for gender estimation: Which eigenvectors contribute? *The 16th International Conference on Pattern Recognition*, Quebec City, Canada, pp.363-366, 2002.

[14] M. Kirby and L. Sirovich, Application of the Karhunen-Loeve procedure for the characterization of human faces, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.12, no.1, pp.103-108, 1990.

[15] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks, *Proc. of the International Joint Conference on Neural Networks*, Budapest, Hungary, pp.25-29, 2004.

[16] G. B. Huang, Q. Zhu and C. Siew, Extreme learning machine: Theory and applications, *Elsevier Neurocomputing*, vol.70, no.1, pp.489-501, 2006.

[17] M. S. Bartlett, J. R. Movellan and T. J. Sejnowski, Face recognition by independent component analysis, *IEEE Trans. on Neural Networks*, vol.13, no.6, pp.1450-1464, 2002.

[18] P. L. Bartlett, The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network, *IEEE Trans. on Inform. Theory*, vol.44, no.2, pp.525-536, 1998.

[19] J. Bekios-Calfa, J. Buenaposada and L. Baumela, Revisiting linear discriminant techniques in gender recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp.858-864, 2011.

[20] C. Liu and H. Wechsler, Evolutionary pursuit and its application to face recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.22, no.6, pp.570-582, 2000.

[21] Q.-Y. Zhu, A. K. Qin, P. N. Suganthan and G.-B. Huang, Rapid and brief communication: Evolutionary extreme learning machine, *Journal of the Pattern Recognition*, pp.1759-1763, 2005.

[22] F. R. Bach and M. I. Jordan, Kernel independent component analysis, *Journal of Machine Learning Research*, vol.3, pp.1-48, 2002.

[23] J. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms, *IEEE Trans. on Neural Networks*, vol.14, no.1, pp.117-126, 2003.

[24] R. Ghosh and B. Verma, A hierarchical method for finding optimal architecture and weights using evolutionary least square based learning, *Int. J. Neural Syst.*, vol.12, no.1, pp.13-24, 2003.

[25] L. Wiskott, J. M. Fellus, N. Kruger and C. Von Der Malsburg, Face recognition by elastic bunch graph matching, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.775-779, 1997.

[26] B. Moghaddam and M.-H. Yang, Learning gender with support faces, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.24, no.5, 2002.

[27] E. J. Candes, L. Demanet, D. L. Donoho and L. Ying, Fast discrete curvelet transform, *SIAM Multi-scale Modeling and Simulations*, 2007.

[28] J. Yang, D. Zhang, A. F. Frangi and J. Yang, Two-dimensional PCA: A new approach to appearance based face representation and recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.26, no.1, pp.131-137, 2004.

[29] *ORL Face Database*, AT&T Laboratories Cambridge, http://www.camorl.co.uk/facedatabase.html, 2005.

[30] *Yale Face Database*, Columbia Univ., http://www.cs.columbia.edu/belhumeur/pub/images/yalefa ces/, 2005.

[31] T. Mandal, Q. M. J. Wu and Y. Yuan, Curvelet based face recognition via dimension reduction, *Elsevier Signal Processing*, vol.89, no.3, pp.2345-2353, 2009.

[32] A. A. Mohammed, R. Minhas, Q. M. Jonathan Wu and M. A. Sid-Ahmed, Human face recognition based on multidimensional PCA and extreme learning machine, *Pattern Recognition*, vol.12, no.3, pp.2588-2597, 2011.