IJICTR International Journal of Information & Communication Technology Research ITRC

Volume 7- Number 3- Summer 2015 (31-42)

Using Hybrid of Eigenface and Imperialist Competition Algorithm for Feature Selection in Face Recognition

Mohammad Reza Yousefi Darestani
Electrical Engineering Department
Islamic Azad University-South Tehran Branch
Tehran, Iran
m.r.darestani@gmail.com

Mansour Sheikhan
Electrical Engineering Department
Islamic Azad University-South Tehran Branch
Tehran, Iran
msheikhn@azad.ac.ir

M. Khademi
Applied Mathematics Department
Islamic Azad University-South Tehran Branch
Tehran, Iran
khademi@azad.ac.ir

Received: August 27, 2014-Accepted: April 14, 2015

***Abstract*— Design of a robust Face Recognition (FR) system is greatly affected due to varying illumination and pose conditions. The accuracy of FR system can be increased by normalizing and compensating the illumination variations in the pre-processing stage. To improve the robustness of FR systems against illumination variations, a method is proposed in this study which is based on Contourlet Transform (CT), hybrid of Principal Component Analysis (PCA) and Imperialist Competition Algorithm (ICA) techniques for feature reduction, and an ICA-optimized Multi-Layer Perceptron (MLP) classifier. First, each face is decomposed using the CT. So, the contourlet coefficients of low and high frequency in different scales and various angles are obtained. The frequency coefficients are employed as a feature vector for further processing. The PCA is then used to reduce the dimensionality of the feature vector. The ICA is also exploited to search the feature space for an optimal feature subset. Then, the reduced-size feature vector is applied to the face classifier that is based on MLP neural network whose structure and learning rate are optimized by ICA. The proposed method is robust to variation of imaging conditions and pose variations. The proposed technique provides better results when tested on ORL and Extended Yale-B databases as compared with other existing techniques such as hybrid model based on discrete wavelet transform and PCA (in terms of precision, sensitivity, and accuracy) and different state-of-the-art methods.**

Keywords- face recognition; contourlet transform; principal component analysis; imperialist competition algorithm; multi-layer perceptron.

I. INTRODUCTION

Biometric recognition refers to the automatic recognition of individuals based on the features derived from their anatomical or behavioral characteristics [1]. The anatomical characteristics of the human beings are achieved by measuring their anatomical attributes such as iris, face, finger-print, and hand shape, while the behavioral characteristics are extracted from their behaviors such as voice and keystroke. A biometric-based recognition system is more reliable than traditional recognition systems (e.g., systems that are based on identification cards and passwords) [2]. These qualities made biometric recognition techniques much more attractive than the traditional methods of identification [3].

The biometric technology has been employed for more reliable identification and verification in passenger access control in the restricted areas, control procedures in airports, database access, border control, and financial services [4]. There is a subtle difference between recognition and verification concepts. Recognition (or identification) is defined as matching biometric data of an unknown subject with an identity (1:*N* problem), while verification (or authentication) checks whether certain biometric data corresponds to a given identity (1:1 problem) [5].

Basically, a biometric recognition system is a pattern recognition system that operates by getting biometric data from an individual, extracting a feature set from the acquired data, and comparing this feature set against the template set in the database [6]. In recent years, Face Recognition (FR) has attracted more attention in the field of biometric authentication [7]. Because of its wide applications in human computer interaction, security, law enforcement and entertainment [8], it has become a very active research field in pattern recognition and computer vision areas [9]. Because of the potential unique, stable and non-invasive characteristics of an FR system, it is known as the most promising for high security environments among various biometric techniques such as recognition of fingerprint, palm vein, signature and palm print [10].

The primary task in an FR system is the extraction of features [11]. The features are usually local or global structural descriptors of an image [12]. The classification stage then works on the feature space, where different kinds of classifiers may be employed such as k-nearest neighbor [13, 14], collaborative representation-based [15, 16], sparse representation-based [17-20], joint representation and pattern learning [21], semi-supervised local ridge regression [22], graph regularized sparsity discriminant analysis [23], and neural networks [24, 25]. In most cases, the extracted features restrict the performance of these image classification systems; so, they should be selected with great care. For example, since the images or objects are often shifted, scaled and rotated, it is desirable to define (or design) the features so that they are invariant or robust to these changes [26]. Furthermore, selecting the features that can effectively classify patterns is often a nontrivial task in many applications [12]. Feature selection can be used for feature reduction as a next stage after the feature extraction. It is a process of selecting a subset of the best discriminant features.

The most challenging aspects of FR systems are variations in head, pose, expression and illumination conditions [27]. In contrast to the head pose and expression, the illumination changes lead to a greater challenge in an FR system [28]. It has been proved that illumination variations are more significant than the inherent differences between the individuals for FR systems [9].

Most traditional methods for FR such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA) are sensitive to illumination variations [29]. Hence, face illumination normalization is a critical task in FR systems and many well-known algorithms have been developed to tackle this problem [30].

Existing methods for illumination normalization can be divided into two categories: preprocessing-based and model-based. In model-based approaches, attempts are made to model the variations caused by varying lighting conditions [31]. Theoretically, the model-based approaches are ideal; however, achieving a model which contains all the possible changes is very difficult. In addition, applying model-based approaches to real-world applications needs some additional constraints or assumptions [32]. For example, Chen et al. [26] proposed that illumination variation can be reduced by truncating the low frequencies Discrete Cosine Transform (DCT) coefficients in the logarithmic domain. In light of the development of time-frequency analysis, many researchers proposed extraction of the illumination-insensitive features from the frequency domain. In most cases, the high-frequency sub-bands are used for recognition because they are insensitive to illumination variations. The well-known works in this field include: (a) using wavelet transform in Waveletface, (b) extraction of Gabor features, (c) using DCT, and (d) using Discrete Fourier Transform (DFT) in Spectroface.

To cope with the illumination variations in FR systems, a new FR system is proposed in this study. This method is based on the Contourlet Transform (CT) features, using hybrid of PCA and Imperialist Competition Algorithm (ICA) for feature selection, and a Multi-Layer Perceptron (MLP) neural network classifier that its learning rate and number of hidden nodes are optimized by ICA, as well (Fig. 1). As shown in Fig. 1, the CT is applied for feature extraction in the first stage due to the faces' excellent properties of time-frequency localization and adaptive multi-scale decomposition. So, the contourlet coefficients of low and high frequency in different scales and various angles are achieved. Then, the PCA is employed to reduce the dimensionality of the feature vector and the ICA is used as the second stage of feature selection to determine the number of important features. In the next stage, the learning rate and number of hidden-layer nodes of MLP-based neural classifier are optimized by the ICA.

The performance of the proposed FR system that is based on CT features (selected by PCA+ICA method) and an ICA-optimized MLP is compared with similar systems such as Discrete Wavelet Transform (DWT)-based system and other modern techniques.

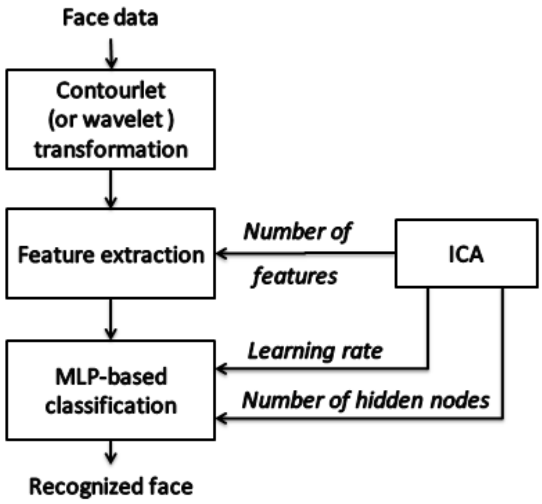

Fig. 1. Block diagram of the proposed method

This paper is organized as follows: Related work is presented in Section II. The preliminaries including contourlet transform, PCA, ICA, and MLP neural model are reviewed briefly in Section III. Simulation and experimental results are presented in Section IV. The paper is concluded in Section V.

II. Related Work

As face is the most common visual pattern in our environment, it attracted research community during recent two decades and significant research activities has been conducted in this area [2, 8, 34]. These implementations are classified into two groups named constituent-based and face-based recognition [35]. Constituent FR is based on association between human facial features such as eyes, nose, mouth, and facial margins [36]. This approach significantly relies on the accuracy of facial feature detection. As far as human faces have similar features with subtle changes in size and geometry, which make them different from one another, reliable extraction of facial features is a tremendously complicated task. Due to these complications, researchers proposed face-based recognition systems [34] wherein a human face is treated as a two-dimensional intensity pattern and recognition is achieved through detection and matching of its statistical properties.

To improve the speed and accuracy of an FR system, various dimensionality reduction techniques have been developed. Among face-based recognition approaches, statistical methods are powerful tools for feature extraction and data representation. Kirby and Sirovich represented human faces as a linear combination of weighted eigenvectors using PCA which extracted features that were most efficient for representation [37]. Similarly, the LDA "Fisher face" method was proposed to extract features that were most efficient for classification [38]. Because of some restrictions of the PCA and LDA, a variety of modifications have been proposed by other researches [18, 23, 39].

Transform-based approaches have been proposed to improve the performance of an FR system for images with high dimensionality. The privileges of these transforms make them an attractive class for feature extraction approaches. Discrete Fractional Fourier Transform (DFFT) [40], DWT [41], curvelet transform [42], contourlet transform [43], and DCT [44] are the well-known methods of this class.

Mixture of statistical and deterministic transforms creates a third kind of feature extraction methodologies with both advantages. In this type, the face images are transformed into a new domain followed by application of PCA or other dimensionality reduction techniques. Development of enhanced multi-resolution analysis techniques has encouraged research community to apply these tools to achieve a high level of class separability in pattern recognition applications. Various combinations of the DFT, DCT, PCA and LDA have been studied by researchers. For example, a combination of the DCT, PCA and the characteristics of the human visual system was used for recognition of faces [45]. Common wavelet-based FR architectures include wavelet-based PCA [46], wavelet-based LDA [47], and Gabor-based kernel association memory [48]. Emergence of curvelet that offers enhanced directional and edge representation has prompted researchers to apply them to several areas of image processing [49]. Curvelet-based PCA [50, 51] and curvelet-based LDA [52] are some curvelet-based FR approaches.

As similar recent related works, the following studies on using different features (and feature selection), classifiers, and obtaining robustness against illumination variations and pose changes can be mentioned: Vidya et al. [11] proposed a method called Selective Illumination Enhancement Technique (SIET). They employed threshold-based DWT feature extraction for enhancing the performance of an FR system. Also in their system, a Binary Particle Swarm Optimization (BPSO) feature selection algorithm was used to search the feature vector space for the optimal feature subset. Yang et al. [21] proposed a feature pattern dictionary using structured information and prior knowledge of image features to represent the unknown feature pattern weight of a query image in FR system. Zhang and Lin [18] proposed Coupled PCA (CPCA) that employed a feature-based representation for heterogeneous face images. In this way, the local binary patterns were employed to capture the local structure of face images and then CPCA was used to obtain the global face information. Liang et al. [14] proposed a Bayesian learning framework to extract discriminative features. They used the nearest neighbor classifier combined with the Mahalanobis distance for classification. Sharma and Patterh [24] used PCA and Adaptive Neuro-Fuzzy Inference System (ANFIS) for feature extraction and face recognition, respectively. Dai et al. [25] used Artificial Neural Networks (ANNs) with random weights to implement negative correlation learning for building ANN ensembles. Cament et al. [53] extracted Gabor features using a mesh to model face deformations produced by varying pose. The local normalization method was used for illumination compensation. Du et al. [17] proposed a sparse representation-based robust FR method called graph regularized low-rank sparse representation recovery. The training and test samples were corrupted in their study because of illumination variations, pose changes, and occlusion to evaluate the robustness of system. Zhang et al. [19] proposed an FR method using multiple view images to overcome the large pose variations.

III. PRELIMINARIES

The fundamental of techniques which employed in this study is reviewed briefly in this section.

A. Contourlet Transform

Contourlet can overcome the problem of intrinsic geometrical structure containing contours. CT is a multi-scale and directional image representation that uses first a wavelet-like structure for edge detection, and then a local directional transform for contour segment detection.

The primary goal of the contourlet construction is to obtain a sparse representation for a typical image that is piecewise smooth. Two-dimensional wavelets are only good for catching the point discontinuities, but do not capture the geometrical smoothness of the contours [54]. To get rid of the limitations of wavelets, the CT was proposed by employing a double filter bank arrangement in which the Laplacian Pyramid (LP) is used to catch the point discontinuities, and then a Directional Filter Bank (DFB) is used to connect point discontinuities into a linear structure [55].

In this research, CT is applied to images in 2 levels and ‘pkva8’ and ‘pkva’ filters were employed as directional filter and pyramid filter, respectively. For example, contourlet image decompositions of a face sample in 2-level with ‘pkva8’ filter is shown in Fig. 2.

B. Eigenface Algorithm

As mentioned before, the PCAs are used in FR application with two main purposes: (a) reducing the dimensions of data, (b) extracting the most representative features out of the input data. So, although the size is reduced, the main features are kept and represent the original data.

In other words, the PCA transforms the original p-dimensional feature vector into an L-dimensional linear subspace that is spanned by the leading eigenvectors of the covariance matrix of feature vector in each cluster ($L < p$). PCA is theoretically the optimum transformation for a given data in the least square sense [56]. However, the problem of choosing the optimum number of principal components remains as an important question.

The basic idea of eigenfaces is that all face images are similar in all configurations and they can be described in its basic face images. The steps of eigenface procedure are as follows [57]:

1. Obtain N training images $I_1, I_2, I_3, \dots, I_N$.

2. Represent each image I_i as a vector. Each image is of size $p.q \times 1$, Let $p.q = n$.

3. Find the mean face vector Ψ for N images:

$$\Psi = \frac{1}{N}\sum_{i=1}^{N} I_i \quad (1)$$

4. Subtract the mean face from each face vector I_i to get a set of vectors Φ_i. This results in obtaining distinguishing features from each face and omitting information that is common:

$$\Phi_i = I_i - \Psi \quad (2)$$

5. Find the covariance matrix C: $C = AA^T = A^T A$

where $A = [\Phi_1, \Phi_2, \dots, \Phi_N]$.

The covariance matrix is simply constructed by placing one modified image vector in each column.

6. Calculate the eigenvectors u and the eigenvalues d of C.

7. Multiply the mean subtracted images by the corresponding eigenvectors.

8. Select the top Leigenvectors having the largest eigenvalues.

It is noted that the eigenvalues corresponding to the eigenvectors are called eigenfaces.

C. Imperialist Competition Algorithm

ICA is a socio-politically motivated global search strategy introduced for dealing with different optimization tasks [58]. This evolutionary optimization strategy has shown acceptable performance in both convergence rate and better global optima achievement [59, 60]. Figure 3 depicts the flowchart of the ICA [58].

Similar to other evolutionary algorithms, this algorithm starts with an initial population. Each individual of the population is called a country. Some of the best countries (in the optimization terminology, the countries with the least cost) are selected to be the imperialist states and the rest form the colonies of these imperialists.

After forming initial empires, the colonies in each of them start moving toward their relevant imperialist country. Figure 4 shows the movement of colonies toward their relevant imperialist where γ is the assimilation angle coefficient and β is the assimilation coefficient. The total power of an empire depends on both the power of the imperialist country and the power of its colonies. This fact is modelled by defining the total power of an empire as the power of imperialist country plus a percentage of mean power of its colonies. While moving toward the imperialist, a colony might reach to a position with lower cost than the imperialist. In this case, the imperialist and the colony change their positions. Thus, the algorithm will continue by the imperialist in the new position and the colonies will be assimilated by the imperialist in its new position.

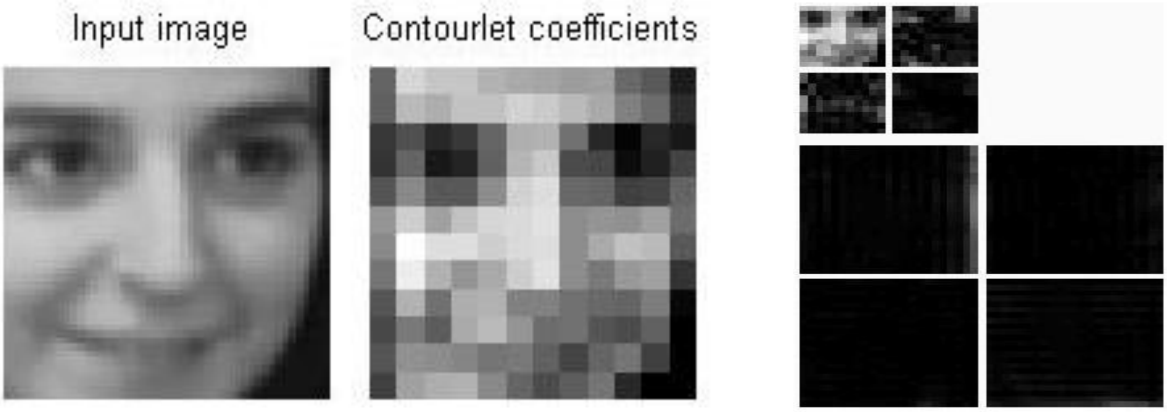

Fig. 2. Decomposition of an image using contourlet transform (2-level and ‘pkva8’ filter as directional filter)

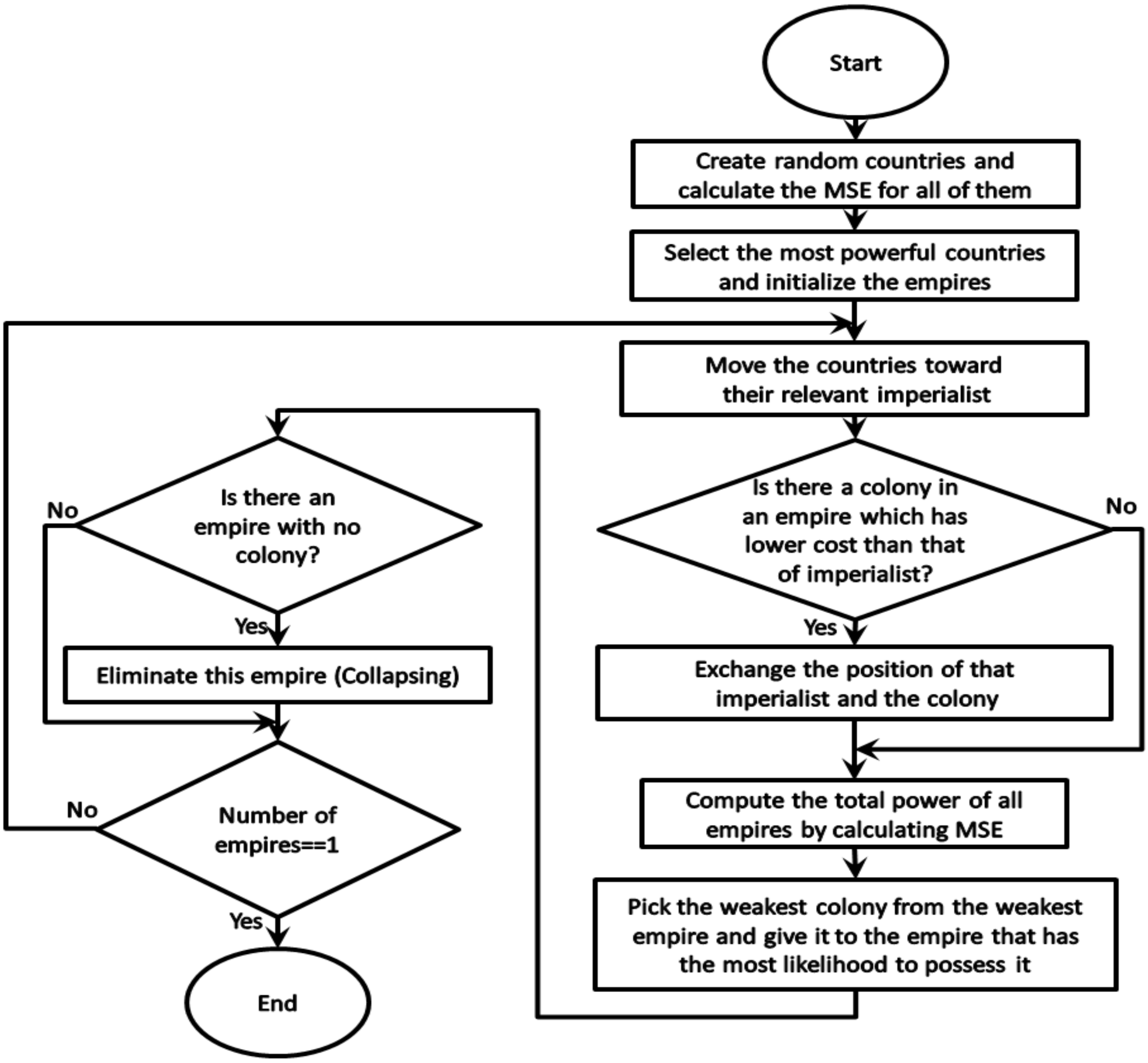

Fig. 3. Flowchart of imperialist competition algorithm

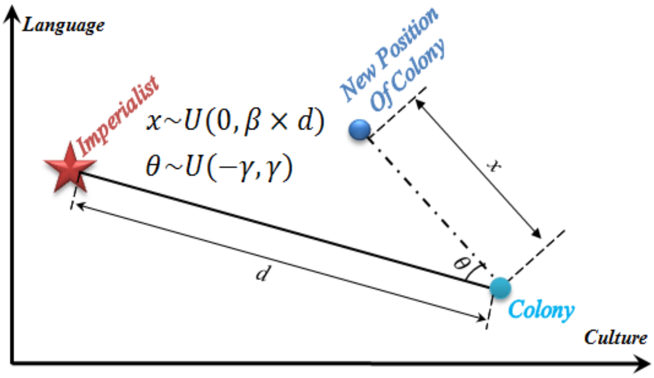

Fig.4. Movement of a colony toward its relevant imperialist in a randomly deviated direction [58]

The total power of an empire is mainly affected by the power of imperialist country. However the power of the colonies of an empire has an effect, albeit negligible, on the total power of that empire. This fact is modelled by defining the total cost of an empire as follows:

$$TC\_n = \mathrm{Cost}(\mathrm{imperialist\_n}) + \xi \times \mathrm{mean}\{\mathrm{Cost}(\mathrm{colonies\ of\ empire\_n})\} \quad (3)$$

where $TC_n$is the total cost of the n^{th} empire and ξ is a positive small number. A little value for ξcauses the total power of the empire to be determined by just the imperialist and increasing it will increase the role of the colonies in determining the total power of an empire.

The imperialistic competition gradually brings about a decrease in the power of weaker empires and an increase in the power of more powerful ones. Different criteria can be used to stop the algorithm. One idea is to exert a number of maximum iteration of the algorithm, called maximum decades, to stop the algorithm or the end of imperialistic competition (when there is only one empire) can be considered as the stop criterion of the ICA. On the other hand, the algorithm can be stopped when its best solution in different decades cannot be improved for some consecutive decades.

D. MLP Neural Network

An MLP consists of two or more layers of nonlinear nodes. The number of input and output nodes is uniquely determined by the number of input features and output classes. The important question is how to appropriately set the number of hidden nodes. Too many hidden nodes can deteriorate the prediction capabilities of the network, while too few hidden nodes can obstruct the learning process. There is no specific guideline to determine the optimum number of hidden nodes, except based on one's experience. It is generally understood only that setting too few or too many hidden nodes causes lack-of-fit or over-fitting in the network. Although, the trial-and-error method is usually used to set the network parameters, ICA is employed in this paper to set the number of hidden nodes and learning rate of the MLP. In this case, we search the range of [20, 50] and [0, 0.7] for the number of hidden nodes and the learning rate, respectively. Furthermore, as mentioned before, ICA was used to select the best feature subset in the training phase. The MLP was trained by the scaled conjugate gradient algorithm [61].

IV. Simulation and Experimental Results

A comprehensive, systematically annotated database is needed to compare the performance of FR algorithms. A database contains face images that have been captured at different pose angles, with a wide variety of illumination conditions and under a diverse encountered illumination color temperatures.

The publicly available standard Olivetti-Oracle Research Lab (ORL) and Extended Yale B databases [62] have been used as the face dataset in this study. The ORL database contains 10 tightly-cropped images of 40 individuals. All images are of grey scale with a 48×48 pixels resolution. These 8-bit grey images were taken at different times, slightly varying lighting, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (with glasses/without glasses) [63]. The subjects were in an upright frontal position. There was a rotation tolerance of $\pm 20^{\circ}$. Figure 5 represents sample images from ORL database.

The Extended Yale B database contains images of 38 individuals, in which each person has different images with the size of 192 × 168 pixels under 9 poses

Fig. 5. Sample images from ORL database

and 64 illumination conditions. In our experiments, we

Fig. 6. Sample images from Extended Yale B database

chose the former 40 images per person. The sample images of this database are shown in Fig 6. So, the robustness of the proposed method against varying illumination and pose conditions will be evaluated.

Although the lighting condition of each image of person of Extended Yale-B database is well defined, but to know the quality of the image in respect to luminance, we calculated the global luminance quality and accordingly normalization process was employed.

The poor luminance quality index of the image reflects the poor illumination condition and hence it requires illumination normalization [64]. The size of pictures in these two databases is not same, so the resize process was applied to the pictures (i.e., the images in the Extended Yale-B database were resized into 48×48 pixels).

In the simulation steps, the images were decomposed using two distinctive transformation methods (wavelet and contourlet). Then, the decomposed images were settled alongside to extract the features. Because of the small size of the images, 1-level decomposition with db2 filter was exploited in wavelet transform. The 2-level decomposition with "pkva" filter as pyramid filter and "pkva8" filter as directional filter were used in CT, as well. So, the48×48 pixel images were decomposed to 24×24 and 12×12 pixel images by employing wavelet and contourlet transformation, respectively.

Then, the test and training data were separated randomly and non-returnable. In this way, 50% of images (i.e., 200 images) in ORL database were used for training and the remaining images were used for testing (similar to [63]). For the Extended Yale-B databases, we randomly chose 20 images per person for training (i.e., 50 percent of selected images) and the remaining samples were used for testing (as performed in [65]).

Using PCA, the eigenvector and eigenvalue of training images were calculated and 50 of the best features in each image were extracted. Afterward, the features which selected by ICA were fed into the neural network to train it. So, the size of input matrix of neural network was $280 \times N$ (in which N was optimized by ICA). The columns of this matrix were the best features and the rows were the images which were dedicated to be trained.

As far as the ORL and the Extended Yale-B databases contain 40 and 38 individuals, respectively, we should employ different structure for the neural network. In other words, the output neurons of the network should cover the number of individuals that each of them indicates a particular person. Using ICA, the number of hidden-layer neurons and the required learning rate of the neural network were optimized. In the next step, the obtained optimized value was employed to train the network. In addition, in order to inhibit the saturation of neurons before delivering the data to the input of network, the data should be normalized (in [0, 1] range). Therefore, the network was trained by the optimized number of hidden layer neurons and learning rate to obtain its suitable weights and bias values. The Mean Squared Error (MSE) in the training phase is depicted in Fig. 7. In the test phase, using the feature vector and the mean value of training image, eigenvalues were extracted from the test images.

To obtain the class of each image, this matrix was fed to the input layer of neural network. Because of the sigmoid activation function in the output layer, the output value of neurons was between [0, 1]. As far as the maximum output shows the output class, it was replaced by 1 and the remaining outputs were replaced by 0.

The definitions of accuracy and precision metrics (used in the following) in terms of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are given in Eqs. (4) and (5).

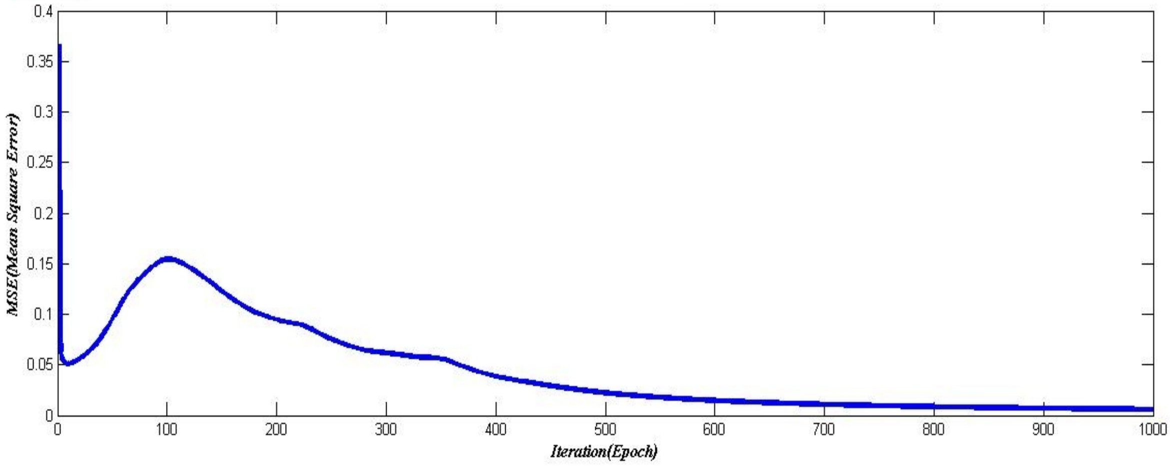

Fig. 7. MSE in each epoch for CT-PCA optimized-MLP using ICA method

$$Accuracy(ACC) = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Positive\ Predictive\ Value\ (PPV) or\ Precision = \frac{TP}{TP+FP} \quad (5)$$

It is noted that an ROC curve is defined by FP and TP rates as x and y axes, respectively. The Receiver Operating Characteristic (ROC) curves for the test data in three simulated methods (i.e., PCA+optimized-MLP, DWT+PCA+optimized-MLP, and CT+PCA+optimized-MLP using ICA optimization method) are shown in Fig. 8.

In this paper, the simulations were run on a PC powered by an Intel core-i5, 2.4 GHz CPU, and 4 GB of RAM. The parameters used for simulations are shown in Table 1. Table 2 shows the performance results of the proposed simulated models. As shown in Table 2, CT+PCA results in superior performance when tested on two different datasets. Also, because of more reduction of hidden-layer neurons, the training time in CT+PCA is shorter than the DWT+PCA and PCA. The performance of the proposed simulated models without using of the optimization algorithm is given in Table 3.

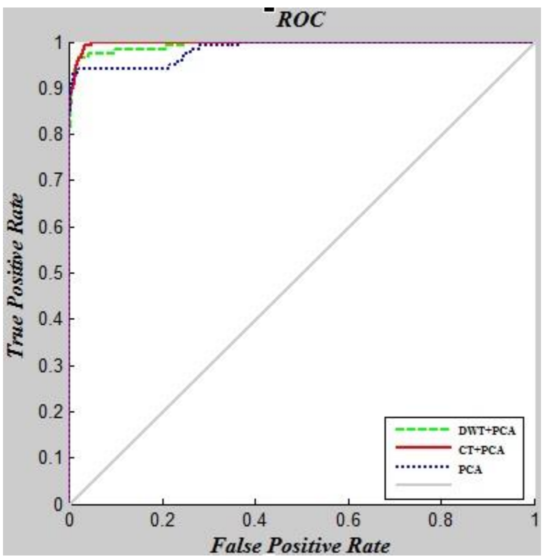

Fig. 8. ROC curves of ICA-optimized simulated methods in this study

The performance of proposed model is compared with five famous different methods when tested on ORL dataset reported in [63] (Table 4). In Table 4, FDS stands for Fisher Discriminant with Schur decomposition.

The performance of proposed model is compared with other modern different methods when tested on Extended Yale B dataset (Table 5). The first group of the state-of-the-art methods in Table 5 is patch-based methods including Patch-based Nearest Neighbor (PNN) [66], Patch-based Collaborative Representation Classification (PCRC) [67], Patch-based Sparse Representation Classification (PSRC) [68], and Modular IMage PCA (MIMPCA) [69]. The second one is Volterrafaces [70]. The third one is Top Level Wavelet Sub-Bands (TLWSB) [71] which is a combination of PCA and the Traditional DWT method (TDWT) [76] and an SVM classifier. As seen in Tables 4 and 5, the proposed method outperforms the state-of-the-art investigated methods when tested on ORL and Extended Yale B databases.

V. CONCLUSIONS

FR is a both challenging and important recognition technique. Among all the biometric techniques, FR approach has one great advantage, which is its user-friendliness (or non-intrusiveness). Feature extraction, feature selection, and classification are the three stages of typical FR systems. Two feature extraction methods (CT-PCA and DWT-PCA) were employed in this study. As compared to some common transforms, such as wavelet, contourlet can capture smooth contours in the images. Also, directionality and anisotropy are two other important advantages of CT. One-level decomposition and db2 filter were employed for DWT, and 2-level decomposition, “pkva8” directional filter, and “pkva” pyramid filter were used in CT. The feature vector was formed by applying ICA-based optimizer to determine the optimum number of features. The MLP neural network was used as the classifier in this study. The training and structure parameters (i.e., learning rate and number of hidden layer nodes) of MLP were optimized by ICA-based optimizer, as well.

Table 1. Parameters setting of each section in the proposed method

ICA parameters		Contourlet parameters		Wavelet parameters	
Number of countries	100	Level	2	Level	1
Number of initial imperialists	10	Directional filter	pkva8	Filter	db2
Number of decades	30	Pyramid filter	pkva		
Revolution rate	0.3				
Assimilation coefficient	2				
Assimilation angle coefficient	0.5				
Zeta	0.02				

Table 2. Performance comparison of ICA-optimized simulated models in this study

Preprocessing stage of ICA-optimized simulated model based on MLP classifier	Database	Picture size	Feature set size	Learning rate	Hidden layer size	Precision	Sensivity	Specifity	Accuracy
PCA	ORL	48×48	37	0.53	40	88.6	91.3	99.8	99.6
	Extended Yale B		32	0.52	42	89.2	90.7	99.3	99.3
DWT+PCA	ORL	24×24	33	0.44	38	90.1	89.7	99.7	99.5
	Extended Yale B		38	0.53	39	93.4	91.5	99.6	99.6
CT+PCA	ORL	12×12	34	0.58	35	92.5	99.3	99.8	99.6
	Extended Yale B		39	0.59	36	97.8	99.5	99.4	99.7

Table 3. Performance comparison of simulated models without using ICA optimization algorithm in this study

Preprocessing stage of simulated model based on MLP classifier (without optimization)	Database	Picture size	Feature set size	Learning rate	Hidden layer size	Precision	Sensivity	Specifity	Accuracy
PCA	ORL	48×48	35	0.6	37	85.2	91.1	99.3	99.3
	Extended Yale B		35	0.6	37	87.8	89.5	99.2	99.2
DWT+PCA	ORL	24×24	35	0.6	37	88.4	88.6	99.6	99.1
	Extended Yale B		35	0.6	37	92.1	90.2	99.5	99.3
CT+PCA	ORL	12×12	35	0.6	37	91.4	97.1	99.7	99.1
	Extended Yale B		35	0.6	37	95.2	98.4	99.2	99.2

Table 4. Performance comparison of proposed method with five different methods reported in [63] (tested on ORL database)

Method	**Precision (%)**
Eigenfaces	77
Laplacianfaces	81
Fisherfaces	77
FDS	82
Schurfaces	85
PCA (in this study)	88.6
DWT+PCA (in this study)	90.1
CT+PCA (in this study)	92.5

Table 5. Performance comparison of proposed method with seven different methods (tested on Extended Yale B database)

Method	**Precision (%)**
TDWT	89.78
TWSBF	40.12
PNN	95.11
PSRC	94.18
PCRC	97.43
MIMPCA	75.13
Volterrafaces	71.65
PCA (in this study)	89.2
DWT+PCA (in this study)	93.4
CT+PCA (in this study)	97.8

Experimental results showed that CT-PCA combined with ICA for optimizing feature selection and MLP's parameters has the highest precision or recognition rate (i.e., 97.8%) among the other simulated methods when tested on Extended Yale B database. Also, the proposed method is very fast because of efficient reduction of features and using an optimized structure and learning strategy for MLP. The method is suitable for real-time applications for visual surveillance and robotics systems and also can be used for similar classification tasks.

VI. References

[1] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," Security & Privacy, vol. 1, pp. 33-42, Apr. 2003.

[2] A. K. Jain, "Technology: Biometric recognition," Nature, vol. 449, pp. 38-40, Sep. 2007.

[3] Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin, "Infrared gait recognition based on wavelet transform and support vector machine," Pattern Recognition, vol. 43, pp. 2904-2910, Aug. 2010.

[4] R. M. Farouk, "Iris recognition based on elastic graph matching and Gabor wavelets," Computer Vision and Image Understanding, vol. 115, pp. 1239-1244, Aug. 2011.

[5] A. Serrano, I. M. de Diego, C. Conde, and E. Cabello, "Recent advances in face biometrics with Gabor wavelets: A review," Pattern Recognition Letters, vol. 31, pp. 372-381, Apr. 2010.

[6] A. Abhyankar and S. Schuckers, "A novel biorthogonal wavelet network system for off-angle iris recognition," Pattern Recognition, vol. 43, pp. 987-1007, Mar. 2010.

[7] K. Jaya Priya, and R. S. Rajesh, "Local fusion of complex dual-tree wavelet coefficients based face recognition for single sample problem," Procedia Computer Science, vol. 2, pp. 94-100, Dec. 2010.

[8] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," Proceedings of the IEEE, vol. 83, pp. 705-741, May 1995.

[9] Y. Cheng, Y. Hou, C. Zhao, Z. Li, Y. Hu, and C. Wang, "Robust face recognition based on illumination invariant in nonsubsampled contourlet transform domain," Neurocomputing, vol. 73, pp. 2217-2224, Jun. 2010.

[10] A. D. Rahulkar and R. S. Holambe, "Partial iris feature extraction and recognition based on a new combined directional and rotated directional wavelet filter banks," Neurocomputing, vol. 81, pp. 12-23, Apr. 2012.

[11] V. Vidya, N. Farheen, K. Manikantan, and S. Ramachandran, "Face recognition using threshold based DWT feature extraction and selective illumination enhancement technique," Procedia Technology, vol. 6, pp. 334-343, Nov. 2012.

[12] G. Yu and S. V. Kamarthi, "A cluster-based wavelet feature extraction method and its application," Engineering Applications of Artificial Intelligence, vol. 23(2): p. 196-202, Mar. 2010.

[13] N. Zhu, T. Xu, L. Wei, and T. Tang, "A nearest neighbor classifier based on virtual test samples for face recognition," Optik-International Journal for Light and Electron Optics, vol. 126, pp. 2799-2803, Nov. 2015.

[14] R. Liang, W. Shen, X. X. Li, and H. Wang, "Bayesian multi-distribution-based discriminative feature extraction for 3D face recognition," Information Sciences, vol. 320, pp. 406-417, Nov. 2015.

[15] A. Moeini, K. Faez, H. Moeini, "Unconstrained pose-invariant face recognition by a triplet collaborative dictionary matrix," Pattern Recognition Letters, vol. 68, pp. 83-89, Dec. 2015.

[16] Z. Liu, J. Pu, Q. Wu, and X. Zhao, "Using the original and symmetrical face training samples to perform collaborative reprentation for face recognition," Optik-International Journal for Light and Electron Optics, vol. 127, pp. 1900-1904, Feb. 2016.

[17] H. Du, X. Zhang, Q. Hu, Y. Hou, "Sparse representation-based robust face recognition by graph regularized low-rank sparse representation recovery," Neurocomputing, vol. 164, pp. 220-229, Sep. 2015.

[18] Z. Zhang and S. Liu, "Coupled principal component analysis based face recognition in heterogeneous sensor neworks," Signal Processing, Available online 3 Sep. 2015.

[19] X. Zhang, D. S. Pham, S. Venkatesh, W. Liu, and D. Phung, "Mixed-norm sparse representation for multi view face recognition," Pattern Recognition, vol. 48, pp. 2935-2946, Sep. 2015.
[20] S. Zhao and S. P. Hu, "A modular weighted sparse representation based on Fisher discriminant and sparse residual for face recognition with occlusion," Information Processing Letters, vol. 115, pp. 677-683, Sep. 2015.
[21] M. Yang, P. Zhu, F. Liu, and L. Shen, "Joint representation and pattern learning for robust face recognition," Neurocomputing, vol. 168, pp. 70-80, Nov. 2015.
[22] Y. Yi, C. Bi, X. Li, J. Wang, and J. Kong, "Semi-supervised local ridge regression for local matching based face recognition," Neurocomputing, vol. 167, pp. 132-146, Nov. 2015.
[23] S. Lou, X. Zhao, Y. Chuang, H. Yu, and S. Zhang, "Graph regularized sparsity discriminant analysis for face recognition," Neurocomputing, vol. 173, Part 2, pp. 290-297, Jan. 2016.
[24] R. Sharma and M. S. Patterh, "A new pose invariant face recognition system using PCA and ANFIS," Optik-International Journal for Light and Electron Optics, vol. 126, pp. 3483-3487, Dec. 2015.
[25] K. Dai, J. Zhao, and F. Cao, "A novel decorrelated neural network ensemble algorithm for face recognition," Knowledge-Based Systems, vol. 89, pp. 541-552, Nov. 2015.
[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, Nov. 2004.
[27] L. A. Cament, F. J. Galdames, K. W. Bowyer, and C. A. Perez, "Face recognition under pose variation with local Gabor features enhanced by active shape and statistical models," Pattern Recognition, vol. 48, pp. 3371-3384, Nov. 2015.
[28] X. Cao, W. Shen, L. G. Yu, Y. L. Wang, J. Y. Yang, Z. W. Zhang, "Illumination invariant extraction for face recognition using neighboring wavelet coefficients," Pattern Recognition, vol. 45, pp. 1299-1305, Apr. 2012.
[29] X. Xie and K. M. Lam, "An efficient illumination normalization method for face recognition," Pattern Recognition Letters, vol. 27, pp. 609-617, Apr. 2006.
[30] X. Xie, W. S. Zheng, J. Lai, P. C. Yuen, and C. Y. Suen, "Normalization of face illumination based on large-and small-scale features," IEEE Transactions on Image Processing, Biometrics Compendium, vol. 20, pp. 1807-1821, Jul. 2011.
[31] Z. Wenyi and R. Chellappa, "Illumination-insensitive face recognition using symmetric shape-from-shading," Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 286-293, Jun. 2000.
[32] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 218-233, Feb. 2003.
[33] W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 36, pp. 458-466, Apr. 2006.
[34] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Computing Surveys, vol. 35, pp. 399-458, Dec. 2003.
[35] B. S. Manjunath, R. Chellappa, and C. von der Malsburg, "A feature based approach to face recognition," Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 373-378, Jun. 1992.
[36] F. Goudail, E. Lange, T. Iwamoto, and K. Kyuma, "Face recognition system using local autocorrelations and multiscale integration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, pp. 1024-1028, Oct. 1996.
[37] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp. 103-108, Jan. 1990.
[38] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," Lecture Notes in Computer Science, vol. 1206, pp. 125-142, Jun. 2005.
[39] F. Wang, X. Liu, and C. Huang, "An improved kernel Fisher discriminant analysis for face recognition," Proceedings International Conference on Computational Intelligence and Security, pp. 353-357, Dec. 2009.
[40] B. Liu and F. Wang, "Face recognition approach based on 2D discrete fractional Fourier transform," Proceedings IEEE 3rd International Conference on Communication Software and Networks, pp. 656-660, May 2011.
[41] H. Hu, "Variable lighting face recognition using discrete wavelet transform," Pattern Recognition Letters, vol. 32, pp. 1526-1534, Jun. 2011.
[42] G. K. Verma, S. Prasad, and G. Bakul, "Robust face recognition using curvelet transform," Proceedings International Conference on Communication, Computing & Security, pp. 239-242, Feb. 2011.
[43] B. Wang, W. Li, and Q. Liao, "Face recognition based on nonsubsampled contourlet transform and block-based kernel Fisher linear discriminant, Proceedings International Conference on Acoustics, Speech and Signal Processing, pp. 1533-1536, Mar. 2012.
[44] G. Günlü and H. S. Bilge, "Face recognition by using 4D Discrete Cosine Transform," Proceedings IEEE 16th Signal Processing, Communication and Applications Conference, pp. 1-4, Apr. 2008.
[45] D. Ramasubramanian and Y. V. Venkatesh, "Encoding and recognition of faces based on the human visual model and DCT. Pattern Recognition," vol. 34, pp. 2447-2458, Aug. 2001.
[46] H. Wang, S. Yang, and W. Liao, "An improved PCA face recognition algorithm based on the discrete wavelet transform and the support vector machines," Proceedings International Conference on Computational Intelligence and Security Workshops, pp. 308-311, Dec. 2007.
[47] J. T. Chien and C. C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 1644-1649, Dec. 2002.
[48] B. L. Zhang, C. Leung, and Y. Gao, "Face recognition by combining kernel associative memory and Gabor transforms," Proceedings 18th International Conference on Pattern Recognition, pp. 465-468, Aug. 2006.
[49] D. L. Donoho and M. R. Duncan, "Digital curvelet transform: Strategy, implementation, and experiments," Proceedings SPIE vol. 4056, Wavelet Applications, pp. 12-30, Apr. 2000.
[50] H. Huo and E. Song, "Face recognition using curvelet and selective PCA," Proceedings International

Conference on Intelligent Control and Information Processing, pp. 348-351, Aug. 2010.
[51] T. Mandal, Q. M. J. Wu, and Y. Yuan, “Curvelet based face recognition via dimension reduction,” Signal Processing, vol. 89, pp. 2345-2353, Dec. 2009.
[52] M. El Aroussi, S. Ghouzali, M. El Hassouni, M. Rziza, and D. Aboutajdine, “Curvelet-based feature extraction with B-LDA for face recognition,” Proceedings IEEE/ACS International Conference on Computer Systems and Applications, pp. 444-448, May 2009.
[53] L. A. Cament, F. J. Galdames, K. W. Bowyer, and C. A. Perez, “Face recognition under pose variation with local Gabor features enhanced by active shape and statistical models,” Pattern Recognition, vol. 48, pp. 3371-3384, Nov. 2015.
[54] D. D. Y. Po and M. N. Do, “Directional multiscale modeling of images using the contourlet transform,” IEEE Transactions on Image Processing, vol. 15, pp. 1610-1620, Jun. 2006.
[55] M. N. Do and M. Vetterli, “The contourlet transform: An efficient directional multiresolution image representation,” IEEE Transactions on Image Processing, vol. 14, pp. 2091-2106, Dec. 2005.
[56] H. Imtiaz and S. A. Fattah, “A wavelet-based dominant feature extraction algorithm for palm-print recognition,” Digital Signal Processing, vol. 23, pp. 244-258, Jan. 2013.
[57] B. J. Devi, N. Veeranjaneyulu, and K. V. K. Kishore, “A novel face recognition system based on combining eigenfaces with fisher faces using wavelets,” Procedia Computer Science, vol. 2, pp. 44-51, Dec. 2010.
[58] E. Atashpaz-Gargari and C. Lucas, “Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition,” Proceedings IEEE Congress on Evolutionary Computation, pp. 4461-4467, Sep. 2007.
[59] E. Atashpaz-Gargari, F. Hashemzadeh, R. Rajabioun, and C. Lucas, “Colonial competitive algorithm, a novel approach for PID controller design in MIMO distillation column process,” International Journal of Intelligent Computing and Cybernetics, vol. 1, pp. 337-355, Jan. 2008.
[60] M. Abdechiri, K. Faez, and H. Bahrami, “Neural network learning based on chaotic imperialist competitive algorithm,” Proceedings 2nd International Workshop on Intelligent Systems and Applications, pp. 1-5, May 2010.
[61] M. F. Møller, “A scaled conjugate gradient algorithm for fast supervised learning,” Neural Networks, vol. 6, pp. 525-533, Apr. 1993.
[62] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and pose,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 643-660, Jun 2001.
[63] G. Ghinea, R. Kannan, and S. Kannaiyan, “Gradient-orientation-based PCA subspace for novel face recognition,” IEEE Access, vol. 2, pp. 914-920, Sep. 2014.
[64] H. Sellahewa and S. A. Jassim, “Image-quality-based adaptive face recognition,” IEEE Transactions on Instrumentation and Measurement, vol. 59, pp. 805-813, Apr. 2010.
[65] Z. H. Huang, W. J. Li, J. Shang, J. Wang, and T. Zhang, “Non-uniform patch based face recognition via 2D-DWT,” Image and Vision Computing, vol. 37, pp. 12-19, May 2015.
[66] R. Kumar, A. Banerjee, B. C. Vemuri, and H. Pfister, “Maximizing all margins: Pushing face recognition with kernel plurality,” Proceedings IEEE International Conference on Computer Vision, pp. 2375-2382, 2011.
[67] P. Zhu, L. Zhang, Q. Hu, and S. C. K. Shiu, “Multi-scale patch based collaborative representation for face recognition with margin distribution optimization,” Lecture Notes in Computer Science, vol. 7572, pp. 822-835, 2012.
[68] J. Wright, A. Y. Yang, A. Ganesh, and S. S. Sastry, “Robust face recognition via sparse representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 210-227, Feb. 2009.
[69] G. D. C. Cavalcanti, T. I. Ren, and J. F. Pereira, “Weighted modular image principal component analysis for face recognition,” Expert Systems with Applications, vol. 40, pp. 4971-4977, Sep. 2013.
[70] R. Kumar, A. Banerjee, and B. C. Vemuri, “Volterrafaces: Discriminant analysis using Volterra kernels,” Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 150-155, Jun. 2009.
[71] Z. H. Huang, W. J. Li, J. Wang, T. Zhang, “Face recognition based on pixel-level and feature-level fusion of the top-level’s wavelet sub-bands,” Information Fusion, vol. 22, pp. 95-104, Mar. 2015.
[72] B. Li and Y. Liu, “When eigenfaces are combined with wavelets,” Knowledge-Based Systems, vol. 15, pp. 343-347, Jul. 2002.

Mohammad Reza Yousefi Darestani was born in Tehran, Iran, in 1987. He received the B.Sc. and the M.Sc. degrees in electrical engineering from the Islamic Azad University-South Tehran Branch, Iran, in 2010 and 2013, respectively. His current research interests include image processing, face recognition, intelligent system, human-computer interaction, computational intelligence, sensor network, routing algorithm and optimization problems.

Mansour Sheikhan is currently an Associate Professor in Electrical Engineering Department of Islamic Azad University-South Tehran Branch. His research interests include speech signal processing, neural networks, and intelligent systems. He has published more than 80 journal papers, about 70 conference papers, four books in Farsi, and six book chapters for IET and Taylor & Francis.

Maryam Khademi received her B.Sc. in Mathematics from Kharazmi University, Tehran in 1989 and her M.Sc. and Ph.D. in Applied Mathematics from Iran University of Science and Technology and University of Tehran in 1994 and 2004, respectively. She has been working as Assistant Professor at Islamic Azad University-South Tehran Branch since 1995.