An Empirical Study using Combination of SVM with PSO based Scattering Ratio Optimization and K-means

Hamed Azami, *Student Member, IEEE*
Department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
hamed_azami@ieee.org

Behzad Bozorgtabar
Faculty of Information Science and Engineering
University of Canberra
Canberra, Australia
Behzad.Bozorgtabar@canberra.edu.au

***Abstract*— One of the most significant practical challenges for face recognition is a likeness of faces which leads to a big problem in classification of different classes. To tackle this problem, we present a novel method based on similarity of each face with other faces using the Pearson correlation coefficients. Besides, another problem is variability in lighting intensity which its physics are difficult for accurate model. In this paper, first, discrete wavelet transform (DWT) is used for feature extraction. Next, with respect to the correlation matrix, two algorithms are employed, namely, K-means clustering and particle swarm optimization (PSO) based scattering ratio matrix of correlation features. Then for each cluster, the process of classification is continued by normalization of the each subset firstly and then the decision making for each subset is performed by support vector machine (SVM). The experiments are performed on the ORL and Yale databases and the results show that there are a significant improvement in 45 features based weighted recognition rate.**

Keywords- face recognition; Pearson correlation coefficients; discrete wavelet transform; particle swarm optimization; K-means; support vector machine

I. INTRODUCTION

A method that can verify or indentify a person from a digital image is named face recognition. As a special study of pattern recognition, face recognition has proved to be very useful in daily life such as for security access control systems, content-based indexing, and bank teller machines [1-5]. Generally, there are two kinds of approaches to face recognition, namely, feature-based and template matching (holistic approaches). The feature-based approach is based on the shapes and relationships between the individual facial features containing mouse, nose, eye and chin. On the other hand, the holistic approach handles the face images globally and extracts important facial features based on high-dimensional intensity values of face images automatically [6]. Although the feature-based approaches are more robust against rotation, scale, and illumination variations, they significantly depend on the accuracy of facial feature detection methods. It has been argued that the existing feature-based methods are not reliable enough for extracting individual facial features [6]. On the other hand, due to the well-known statistical methods in holistic face recognition, it has attracted more attention compared with the feature-based approach [6].

Illumination is considered one of the most difficult tasks in face recognition. The illumination setup in which recognition is performed in most impractical cases to control, its physics difficult to accurately model and face appearance differs as variance of illumination is often larger than those differences between individuals. Reliable techniques for recognition under more extreme variations caused by noise or illumination are highly nonlinear and have proven elusive [7].

By implementing suitable feature extractor, discrete wavelet transform (DWT), we can tackle these problems. With considering of features in frequency domain, an image is represented as a weighted combination of main functions. High frequency domain devotes unimportant information to itself while the crucial information can be found in low frequency domain. Sellahewa and Jassim have proved that the low frequency approximation sub-band is suitable for face descriptor for recognition [8,9].

DWT coefficients are obtained by passing the image through the series of filter bank stages. The procedure of appropriate design of DWT and then selecting the low frequency approximation sub-band lead to improve the robustness of features space with respect to variation in illumination. After finding the robust features as face descriptors, our aim is to find the relations through the different faces to make decision about the face class assignment.

The next task is to perform classification of these lower-dimensional feature vectors. Various classifiers were used for discrimination among classes such as probabilistic neural networks (PNN), statistical models like hidden Markov models (HMMs), Gaussian mixture models (GMMs) and support vector machines (SVMs) [10-12].
SVM has been successfully applied to various pattern classification problems, such as handwritten digit recognition, text categorization and face detection, due to their powerful learning ability and good generalization ability. SVM maps the feature space to some new feature space where the classes are more separable, and then attempts to maximize the margin between the separating boundaries and support vectors. Standard Linear SVM

model tries to solve the two class classification problems which are linearly separable, but for solving nonlinear feature spaces, the SVM models use the nonlinear functions such as Radial Basis Function (RBF) to map the features space to some new linear separable features space and then find the appropriate decision boundary [13,14].

Then, the rest of this paper is organized as follows. Proposed methods are briefly introduced in Section 2. Section 3 determines our experimental results on both ORL and Yale datasets. Finally, conclusion is given in Section 4.

II. Proposed Methods

DWT, a powerful feature extraction method, is done as follows: In the first level of decomposition, the image is split into four sub-bands, namely HH1, HL1, LH1, and LL1, as illustrated in Fig. 1. The HH1 sub-band gives the diagonal details of the image; meanwhile HL1 and LH1give horizontal and vertical features. The LL1 sub-band is the low resolution residual consisting of low frequency components which is further split at higher levels of decomposition [15]. Fig. 2 shows an image from the ORL face database with images obtained after one-level wavelet and three-level wavelet transform, respectively.

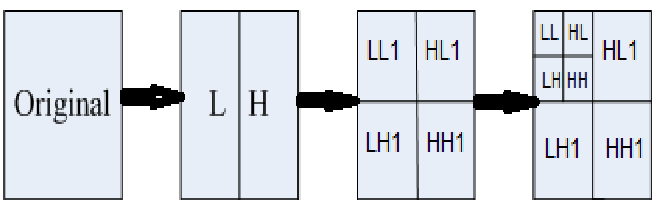


Figure 1. The process of decomposing an image

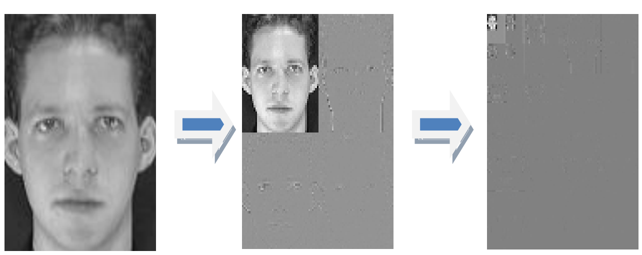


Figure 2. The process of DWT feature extractor

To evaluate the similarity of each face comparison with other classes, the Pearson correlation coefficients are evaluated for each pair of faces through their DWT coefficients and then an array of these face pairs correlations is computed. The Pearson correlation coefficient for two array of *X* and *Y* and can be obtained by:

$$r = \frac{SP}{\sqrt{SS_x \times SS_y}} \tag{1}$$

where

$$SP = \sum XY - \frac{\sum X \times \sum Y}{N} \tag{2}$$

$$SS_y = \sum Y^2 - \frac{\left(\sum Y\right)^2}{N} \tag{3}$$

As the same, SS_x can be evaluated for the *X*, and N is the dimension of each array. By evaluating the Pearson correlation coefficient of each face pair, an $M \times M$ matrix is obtained which M is the number of indexed faces and each row of the matrix contains $F_i : r_{i1}, r_{i2}, \ldots, r_{iM}$.

The within-class-scatter S_W and between-class- scatter S_B can be measured by knowing the covariance matrix and the prior probability of each class. $trace\{S_W\}$ is a measure of the variance of the features in all classes and $trace\{S_B\}$ is a measure of the mean in each class from the whole features. With these definitions, the total scatter matrix S_T can be defined by a measure of variance of all training samples from the global mean. The aim is that, the ratio of $trace\{S_T\}$ to $trace\{S_W\}$ should be as large as possible.

K-means algorithm is used for partition the classes into some new clusters which the classes of each cluster are similar to each other. K-means clustering is a method commonly used to automatically partition a data set into *k* groups [16]. It proceeds by selecting *k* initial cluster centers and then iteratively refining them as follows:

1. Each instance d_i is assigned to its closest cluster center.
2. Each cluster center C_i is updated to be the mean of its constituent instances.

Considering clusters, K-means algorithm is applied to the correlation matrix and the indexed mean feature vector for each cluster is evaluated after some iteration. This step of face recognition is so vital because if the similarity analysis makes the wrong decision, then the future steps also will face to problem, so it shall be robust with face image changes.

The better features for clustering can be selected by minimizing the reverse of this ratio through the PSO. The idea of PSO was first raised by J. Kennedy and R. Eberhart in 1995. PSO is an evolutionary computing algorithm inspired by nature and is based on repetition. The social behavioral of animals like birds and fish when they are together has been the inspiration source for this algorithm [17]. PSO, the same as other evolutionary algorithms, begins with a random matrix as an initial population. Unlike genetic algorithms (GA), normal PSO doesn't have evolutionary operators like mutation and breeding. Each member of the population is called a particle. In fact, in the PSO algorithm a certain number of particles that are formed randomly make the initial values.

There are two parameters for each particle, namely, position and velocity of the particle, which are defined by a space vector and a velocity vector, respectively. These particles form a pattern in an n-dimensional space and move to the desired value. The best position of each particle in the past and the best position among all particles are stored separately. According to the experience from the previous moves, the particles decide how to make the next move. In every iteration, all particles in the *n*-dimensional problem space move to an optimum point. In each iteration, the position and velocity of each particle can be modified according to the following equations:

$$v_i(t+1)=wv_i(t)+C_1r_1\left(p_{best_i}(t)-x_i(t)\right)+C_2r_2\left(g_{best_i}(t)-x_i(t)\right) \quad (4)$$

$$x_i(t+1)=x_i(t)+v_i(t+1) \quad (5)$$

where *n* represents the dimension ($1 \le n \le N$), C_1 and C_2 are positive constants, generally considered 2.0. r_1 and r_2 are random numbers uniformly between 0 and 1; *w* is inertia weight that can be constant or defined by equation [18].

Equation (6) expresses that the velocity vector of each particle is updated ($v_i(t+1)$) and the new and previous values of the vector position ($x_i(t)$) create the new position vector ($x_i(t+1)$). In fact, the updated velocity vector affects both local and global values. The best response of the local positions is the best solution of the particle until current execution time (p_{best}) and the best global solution is the best solution of the entire particles until current execution time (g_{best}). We set PSO parameters with trial and error as follows: population size=15; C_1=C_2=2; iteration=50; *w*=1.

Then, the process of classification is continued by first normalization of cluster's feature vectors and then applying the SVM classifier for that cluster.

The preliminary SVM was a binary classification method that tries to find the optimal linear/nonlinear decision surface based on the concept of structural risk minimization. The decision surface is a weighted representation of the elements of training set [13]. The elements on the decision surface are defined by a set of support vectors which characterizes the boundary between two (or more) classes. Generally, the problem of multi-class is solved by combining multiple two class SVMs. The input to a SVM algorithm is a training data set $\{(x_1,y_1),(x_2,y_2),\ldots\}$. x_i represents the data and y_h=1 or -1 is the corresponding label considering two-class problem. The outputs of a SVM algorithm are a set $\{(s_1,\alpha_1,y_1),(s_2,\alpha_2,y_2),\ldots\}$ where s_i denotes a support vector, α_i is the weight of s_i, y_i is the class label of s_i. By supposing a constant term *b*, the linear decision surface can be rewritten as [14]:

$$w.z+b=0 \quad (6)$$

$$w=\sum_i \alpha_i y_i s_i \quad (7)$$

With assumption of face recognition as a multi-class classification, the task is to generalize the standard SVM model. We select the one-versus-the-rest approach that constructs SVMs which the train k^{th} model chooses the k^{th} class as the positive examples and the remaining $(k-1)$ classes as the negative examples. Comparison with one-versus-one, it significantly needs less training time. At last, indexed SVM classifier is learned through quadratic programming in order to find the class's boundaries with maximum margins. This technique helps to find the accurate relations between nearest similarity faces. Fig. 3 illustrates the flowchart of the proposed approach.

Being easy to recognize different images rather than similar images, the computational modeling of this concept can be adapted that the familiar subsets can be selected with some strategies and then according to each subset, independent accurate classification approaches can be performed according to face image features.

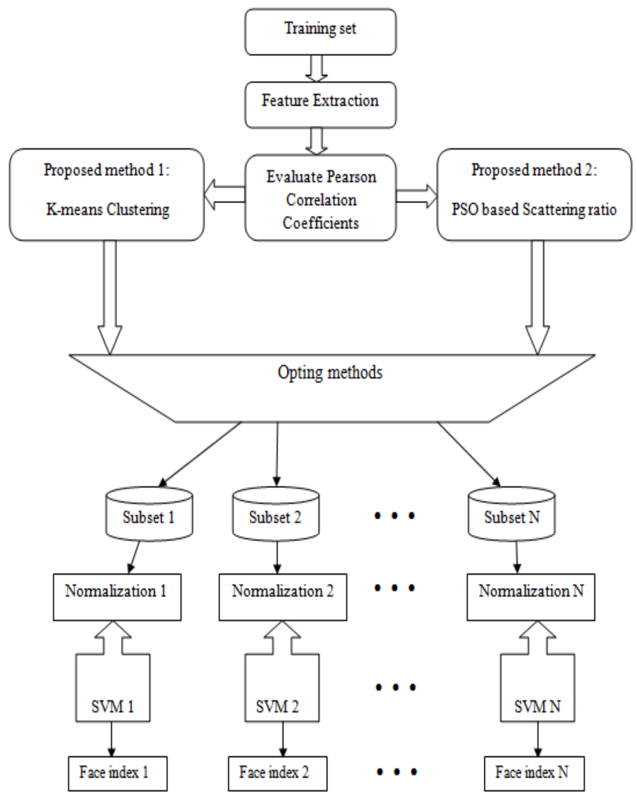


Figure 3. The flowchart of proposed methods

III. SIMULATION RESULTS

To test the proposed methods, ORL and Yale face image databases were selected. So that compare among various methods, weighted recognition rate is used with respect to different numbers of feature. The advantage of this criterion is that the effect of the number of the training samples on various approaches could be studied through the simulations.

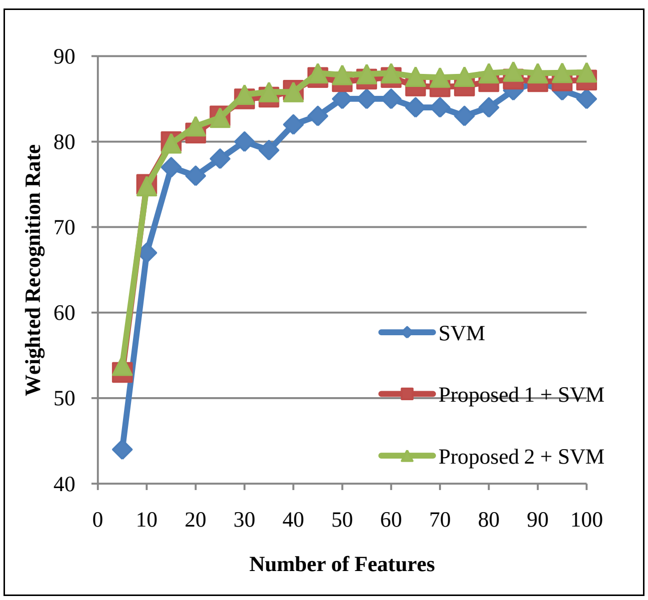
A. ORL Face Database

The ORL database consists of 40 groups, each one contains ten 112×92 gray scale images of a single subject. Each subject's images differ in lighting, facial expression, details (i.e. glasses/no glasses) and even sliding.

Then, the K-means algorithm with different number of clusters is used to partition the dataset into familiar clusters. We reached the best result with three clusters. In order to improve the performance of classifier, the features vectors are normalized with the zero means and unit variances. Indexed SVM classifier with the procedure of quadratic programming then is learned to identify each face's class.

One important point for recognition is suitable initialization of parameters because this stage helps us to precisely classifying the familiar subsets. But the procedure of similarity analysis is so important and the errors resulted from this part can significantly decrease the recognition performance. For achieving clusters which contain similar classes, we proposed PSO based scattering ratio optimization for clustering.

These algorithms are run with different number of training samples from 1 to 5 iterations and the weighted mean recognition rate then is evaluated. Fig. 4 illustrates the weighted mean recognition rates of proposed approaches and conventional classification method in terms of the number of features. Generally speaking that our proposed methods outweigh the conventional SVM method is obvious. As far as weighted recognition rate is concerned, if we increase the number of features (DWT coefficients), the weighted recognition rate will rise until it leveled off in 40 features, and the figures for all methods will be same after this point.



Fgure 4. Comparison of the weighted mean recognition rates of proposed approaches and conventional classification method in terms of the number of features for ORL face database.

45 features of DWT coefficients are assumed as face descriptors. Fig. 5 shows the effect of number of training samples in face recognition that the more training samples are used, the better recognition rate will be. The results demonstrate that compare with linear SVM the proposed approaches lead in improvements, in 45 features from 5% for PSO to 4.5% for K-means algorithm.

Fig. 6 gives information about the recognition rate for one sample of each class with respect to the number of features. In compare with weighted recognition rate, the recognition rate is decreased for 1 training sample. As it's shown, for lower number of features, both proposed methods 1 and 2 are same and outweigh other methods but by increasing in the number of features, the second proposed method (PSO + SVM) has a better recognition rate.

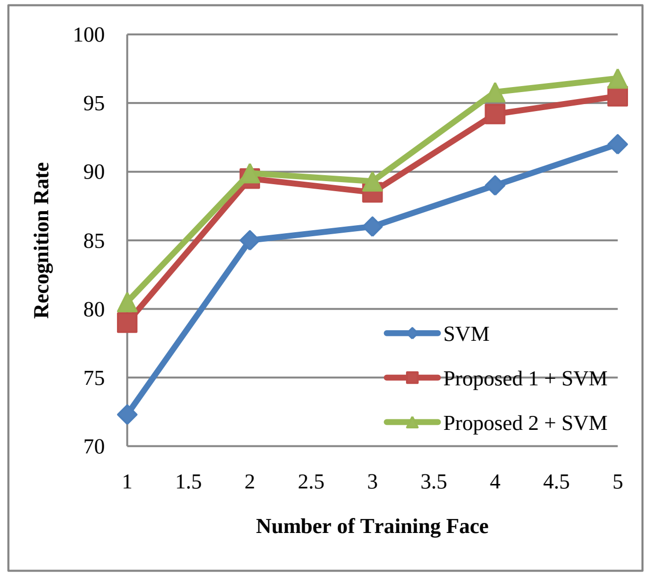


Figure 5. The performance of face recognition methods in terms of the number of training samples of ORL face database

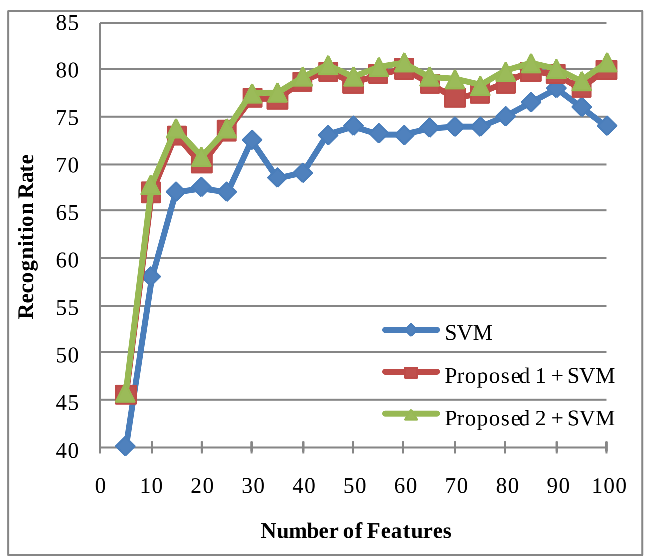


Figure 6. Evaluation of the recognition rates with assuming one learning sample of each class, in terms of the number of features for ORL face database

B. *Yale Face Database*

The Yale face database contains 165 images of 15 individuals (each person provides 11 different images). After feature extraction, to obtain good clusters we implemented k-means and PSO. In testing module each face subset is evaluated by *n* mean feature vector which obtained after employing proposed methods. Afterwards the class is recognized by first features vectors normalization and then classification with the indexed SVM classifier.

Fig. 7 shows weighted recognition rates of proposed methods for different number of features. Overall trend for all methods is an upward until they reached saturation. A comparison study of proposed methods shows that behavior of trend can be divided in a two stage. At the first stage for number of features lower than 65, K-means algorithm outweighs other methods whereas weighted recognition rate is above 80%.

In the second stage, for the number of features above 65, proposed method 1, PSO works better than others. This fact must not be ignored that all proposed methods outranks linear SVM, for example with 45 features, proposed algorithms have a better performance based weighted recognition rate for both PSO and K-means. Besides, Fig. 8 illustrates recognition rate for different number of training samples.

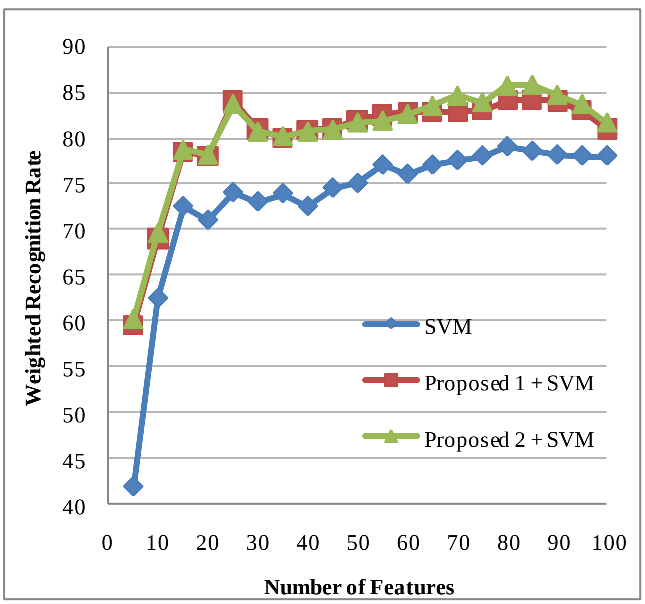


Figure 7. Comparison of the weighted mean recognition rates of proposed approaches and conventional classification method based on the number of features for Yale face database

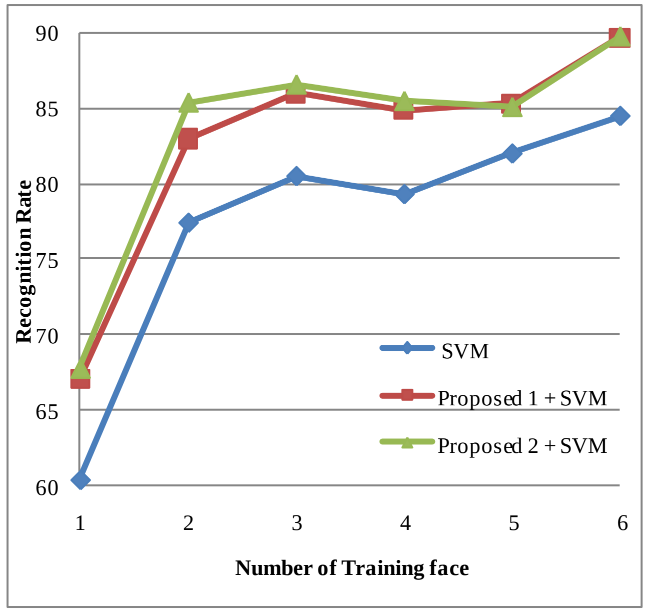


Figure 8. The performance of face recognition methods in terms of the number of training samples of Yale face database

Having no access to many images, we considered recognition rate of proposed techniques for one training sample as it shown in Fig. 9.

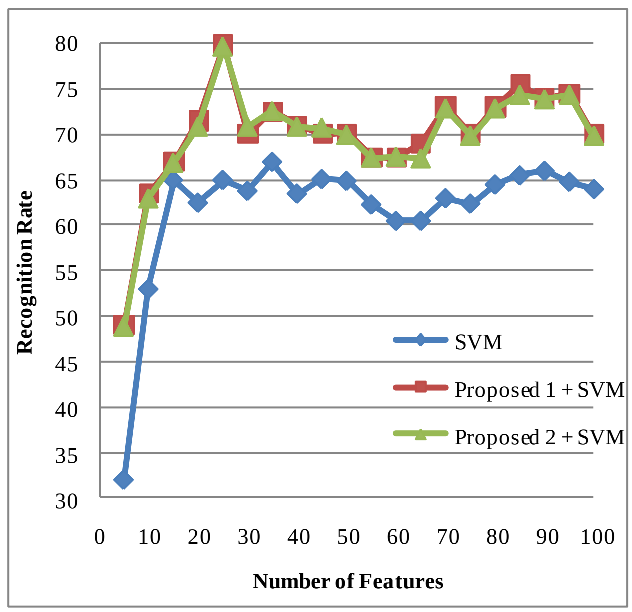


Figure 9. Evaluation of the recognition rates with assumption of one learning sample of each class, in terms of the number of features for Yale face database

As it's shown, the performance of our approaches is significantly better than conventional SVM. Also compared to other methods, PSO seems a bit more efficient. Another important criterion that must not be ignored is processing time. Although in training phase, search techniques like

PSO consume more time compared with conventional SVM, but in detection module, results can be computed quickly and with lower memory requirements. This might prove lucrative for embedded systems programmers, which have storage and processing constraints.

IV. Conclusions

In this paper K-means clustering and PSO based scattering ratio optimization have been proposed for face recognition. The ORL and Yale database images have been used for conducting all the experiments. First, features have been extracted by DWT which greatly has decomposed image to different sub-bands as well as maintains the main facial features. Then, faces which have had similar to each other had to be placed in one cluster. Therefore, new subsets have been made which index SVM examines to find the appropriate decision boundaries for those similar faces. We have tested proposed methods from three aspects: weighted recognition rate, number of training samples and recognition rate for one training sample. In all examinations, proposed approaches have had a better performances rather than conventional SVM.

References

[1] H. Azami, S. Sanei and K. Mohammadi, "Improving the neural network training for face recognition using adaptive learning rate, resilient back propagation and conjugate gradient algorithm", Journal of Computer Applications, vol. 34, no. 2, pp. 22-26, 2011.

[2] E. Gumus, N. Kilic, A. Sertbas and O. N. Ucan, "Evaluation of face recognition techniques using PCA, wavelets and SVM" *Expert Systems with Applications*, vol. 37, pp. 6404-6408, 2008.

[3] M. R. M. Rizk and A. Taha, "Analysis of neural networks for face recognition systems with feature extraction to develop an eye localization based method", *International Conference on Electronics, Circuits and Systems*, pp. 847-850, 2002.

[4] J. Daugman, "Face and gesture recognition: overview", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 675-676, 1997.

[5] L. Wiskott J. Fellous, N. Kruger and C. Malsburg, "Face recognition by elastic bunch graph matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775-779, 1997.

[6] M. J. Er, W. Chen and S. Wu, "High-speed face recognition based on discrete cosine transform and RBF neural networks", IEEE Transactions on Neural Networks, vol. 16, no. 3, 2005.

[7] Y. A. Georghiades, P. Belhumeur and D. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose" *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001.

[8] H. Sellahewa and S. Jassim, "Wavelet-based face verification for constrained platforms*", in Proc. SPIE Biometric Technology. Human Identification II*, vol. 5779, pp. 173-183, 2005.

[9] H.Sellahewa,"Wavelet-based automatic face recognition for constrained devices*", Ph.D. dissertation*, Univ. Buckingham, Buckingham U.K, 2006.

[10] M. A. Aminan, "Empirical study on the characteristics of Gabor representations for face recognition*", International Journal of Pattern Recognition and Artificial Intelligence,* vol. 23, no. 3, 401-431, 2003.

[11] O. Déniz, M. Castrillón and M. Hernández, "Face recognition using independent component analysis and support vector machines", *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2153-2157, 2003.

[12] S. Lucey and T. Chen, "A GMM parts based face representation for improved verification through relevance adaptation", *in Proceeding International Conference of Computer Vision Pattern Recognition (CVPR)*, pp. 855-861, 2004.

[13] A. Mathur and G. M. Foody, "Multiclass an d Binary SVM Classification: Implications for Training and Classification Users", *IEEE Geoscience and Remote Sensing Letters*, Vol. 5, No. 2, 2008.

[14] J. Manikandan and B. Venkataramani, "Study and Evaluation of a Multi-Class SVM Classifier using Diminishing Learning Technique", *Neurocomputing*, Vol. 73, No. 10-12, pp. 1676-1685, 2010.

[15] H. Azami, S. Sanei and K. Mohammadi, "Improving the Neural Network Training for Face Recognition using Adaptive Learning Rate, Resilient Back Propagation and Conjugate Gradient Algorithm", *Journal of Computer Applications*, vol. 34, no. 2, pp. 22-26, 2011.

[16] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations" *Proceedings of the Fifth Symposium on Math, Statistics, and Probability,* pp. 281-297, Berkeley, CA: University of California Press, 1967.

[17] L. Mussi, S. Cagnoni, and F. Daolio, "GPU-based road sign detection using particle swarm optimization", *International Conference on Intelligent Systems Design and applications*, pp. 152-157, 2009.

[18] Z. Rafiee, S. Ganjefar and A. Fattahi, "A new PSS tuning technique using ICA and PSO methods with the fourier transform", *Iranian Conference on Electrical Engineering (ICEE)*, pp. 806-811, 2010.