Optimized Features Selection using Hybrid PSO-GA for Multi-View Gender Classification

Sajid Khan[1], Muhammad Nazir[2], Naveed Riaz[3] and Muhammad Khan[4]
[1]Department of Software Engineering, Foundation University Rawalpindi Campus, Pakistan
[2]Department of Information Technology, University of Central Punjab, Pakistan
[3]Department of Computer Science, University of Dammam, Saudi Arabia
[4]Department of Computer Science, SZABIST, Pakistan

Abstract: *Gender classification is a fundamental face analysis task. In literature, the focus of most researchers has been on the face images acquired under controlled conditions. Real-world face images contain different illumination effects and variations in facial expressions and poses that make gender classification more challenging task. In this paper, we have proposed an efficient gender classification technique for real world face images (Labeled faces in the Wild). After extracting both global and local features using Discrete Cosine Transform (DCT) and Local Binary Pattern (LBP), we have fused these features. Proposed algorithm provides support for variations in expressions and poses. To reduce the data dimensions, fused features are passed to hybrid PSO-GA that eliminates irrelevant features and results in optimized features. Support Vector Machine (SVM) is trained and tested by using optimized features. Using this approach we have received a 98% accuracy rate. We are utilizing the minimum number of features so our technique is faster as compared to other state-of-the-art gender classification techniques.*

Keywords: *Gender classification, feature extraction, pattern recognition, active shape model, real world face images.*

Received February 6, 2013; accepted April 4, 2013; published online April 23, 2014

1. Introduction

In today's technical globe, gender classification plays a vital role. It is extensively utilized in applications such as customer-oriented publicizing, discernible surveillance, intelligent user interfaces and demographics.

With the evolution of Human-Computer Interaction (HCI), to meet with the people growing demands for secure, reliable and convenient services, computer vision approaches such as face identification [1], gesture recognition and gender classification will play an important role in our lives.

Features are classified into two categories: Appearance-based (global) and geometric-based (local) features extraction. In Appearance-based strategy, an image is considered as a high dimensional vector and features are extracted from its statistical information without depending on knowledge about the object of interest. This technique is simple and fast, but unreliable when local appearance variations occur. In geometric-based approach, geometric features like nose, mouth and eyes are extracted from the face portion. This approach has the advantage of rotations and variation's invariability but misses a lot of helpful information.

Golomb *et al.* [6] trained two-layer neural network called SEX-NET and attained 91.9% classification accuracy rate. They used a total of 90 frontal face images. Brunelli and Poggio [3] extracted geometric features and utilized them to train the networks. They claimed 79% classification accuracy rate. Sun *et al.* [18], asserted that Genetic Algorithm (GA) works considerably well for crucial feature's selection task. They utilized Principal Component Analysis (PCA) to create feature's vector and GA to select the vital features. They attained 95.3% accuracy rate after training Support Vector Machine (SVM) classifier by employing those vital features. Jain *et al.* [8] extracted facial features using Independent Component Analysis (ICA) and then they categorized gender using Liner Discriminant Analysis (LDA). They provided results using normalized FERET database and attained 99.3% classification accuracy rate. Sun *et al.* [17] utilized Local Binary Pattern (LBP) to create feature's vectors in the input of adaboost and attained 95.75% classification accuracy rate. Baluja and Rowley [2] attained 93% classification accuracy rate. They utilized pixel comparison operators with Adaboost classifier. Nazir *et al.* [14] utilized Discrete Cosine Transform (DCT) change to extract the vital facial features and utilized K-Nearest Neighbor (K-NN) classifier to categorize gender. Khan *et al.* [10] utilized Discrete Wavelet Transform (DWT) Change to extract facial features. They claimed that classifier performance can be improved by assembling different classifiers using weight majority technique. They performed experiments on Stanford University

Medical Students (SUMs) face database and received 95.63% classification accuracy rate.

A major problem with all these proposed techniques is that these take into account only frontal face images (e.g., SUMS, FERET). The images in these databases consist of pictures that have a clean background are occlusions free, furnish merely frontal faces, encompass limited facial expressions and have consistent lighting effects. However, real-time images are usually captured in unconstrained settings and conditions. A real-time picture normally contains momentous emergence variations like illumination change, poor picture quality, makeup or occlusions and disparate facial expressions. Gender recognition in an unconstrained environment is a very challenging task compared to images captured in constrained settings like FERET and SUMS face databases. This problem has been highlighted by very few researchers. Shakharovich *et al.* [15] gathered 3500 face pictures from the web. They utilized Haar-like features and obtained 79% accuracy rate with adaboost and 75.5% with SVM. Gao *et al.* [5] claimed 95.5% classification accuracy rate by performing experiments on 10,100 real-time images. They also utilized Haar-like features but used probabilistic boosting tree. It is tough to use these results as benchmarks as their databases are not publicly available. Kumar *et al.* [11] performed experiments on real-time images. They received 81.22% classification accuracy rate by training countless binary "attributes" classifiers. Their emphasis was more on face verification and not on gender classification.

Following is the sequence of our paper. In section 2, we provide the details of the proposed methodology. Next, experimental setting and results are discussed and contrasted with some other state of the art techniques. Conclusion and future work are discussed in section 5.

2. Proposed Methodology

Our proposed method consists of the following steps. These steps are depicted in Figure 1.

- *Pre-Processing*: First we align the images by using commercial software [20] and then perform histogram equalization to normalize the face.
- *Face Extraction*: Facial portion is extracted and removal of the unwanted area using spatial co-ordinate system.
- *Features Extraction*: Local and global features are extracted using LBP and DCT.
- *Optimized Features Selection*: Only most important feature sets are selected using hybrid PSO-GA algorithm.
- *Classification*: In this step, SVM is used to classify gender.

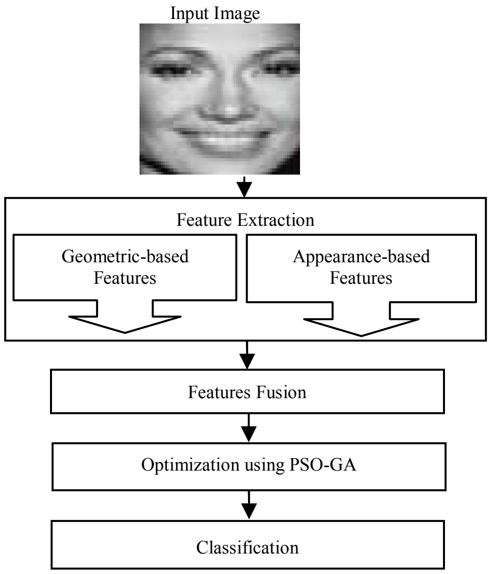


Figure 1. Proposed system architecture.

2.1. Facial Portion Extraction

A solitary tiny picture normally contains thousands of pixels. This huge data set seriously affects the computational time and makes the system slow that is why pixel-based approaches are very expensive. Thus, to remove the unwanted area and to extract merely facial portion, we use spatial co-ordinate system. By using hit and trial method, we set the X and Y coordinates values. Figure 2 depicts a sample of extracted images.

Figure 2. Sample of extracted faces.

2.2. Features Extraction

To extract the global features from facial image, we used DCT technique. DCT is a well-known transformation technique utilized in image compression applications, Majid *et al.* [13] used it for face recognition applications. DCT can be utilized for dimension reduction as well.

Given that a gray picture is expressed by f(x, y) of size N×N, DCT is described using Equation 1:

$$D(x, y) = \frac{2}{\sqrt{MN}} a(u)a(v)x \sum_{M=0}^{M-1} \sum_{N=0}^{N-1} I(m, n) \ldots \cos\left[\frac{(2m+1)u\pi}{2m}\right] \cos\left[\frac{(2n+1)v\pi}{2n}\right] \tag{1}$$

Where:

$$a(u) = \begin{Bmatrix} \sqrt{\frac{1}{M}}, u = 0 \\ \sqrt{\frac{2}{M}}, u = 1, 2, \ldots M-1 \end{Bmatrix} \text{ and } a(v) = \begin{Bmatrix} \sqrt{\frac{1}{N}}, u = 0 \\ \sqrt{\frac{2}{N}}, v = 1, 2, \ldots N-1 \end{Bmatrix}$$

The DCT coefficients with high variance are mainly located in the upper-left corner of the DCT matrix. So, we scan the DCT coefficient matrix in a zigzag manner starting from the upper-left corner and subsequently change it to a One-Dimensional (1D) vector. This connotes to sorting of coefficients with respect to importance. High importance coefficients are placed in the top-left corner of the block. At the end, 16 coefficients are selected from an image. When the number of selected coefficient's increases, so does the size of the feature vector. For 32 size feature vector first two coefficients from every single DCT block are selected.

- *Local Features Extraction*: Features extracted from the whole face portion are recognized as global features. Global features are not resilient to face occlusions and alignment, while those local features that are extracted from different points of a face are subjected to the variation of facial expressions, illumination and occlusions. Psychological experiments show that individual facial features like nose, mouth, eyes, eyebrows and chin hold extra information as compared to the whole face portion. We first locate five major facial components (nose, mouth, eyebrows, chin, eyes) using an active shape model, next these facial features are cropped and LBP is applied to extract features from these cropped components. Figure 3 provides the facial components.

Figure 3. Facial components.

LBP was utilized by Lian *et al*. [12] for gender classification. In LBP, the center pixel is selected and the neighborhood pixels are either modified to 0 if the gray levels are smaller than center or to 1 if the gray levels are larger than a center. The center pixel is next substituted by the binary code of its neighborhood like 00011011, as shown in Figure 4.

1	2	1
5	4	6
7	1	9

0	0	0
1		1
1	0	1

Figure 4. Binary code conversion.

2.3. Feature Fusion

The advantages of feature level fusion are obvious. By combining and fusing different features, we not only can preserve the discriminative information about face but can also eliminate redundant features. To take advantage of different feature's extraction techniques for better classification and to make a proposed technique more robust to variations in expressions, illumination and pose changes, we have fused features extracted by DCT and LBP. Steps of the feature fusion are provided below:

1. Extract the feature's using DCT (denoted as 'd'), LBP (denoted by 'l').
2. Fuse the features vectors (d, l) using concatenation (i. e., d+l) to form one features vector R.

2.4. Optimized Feature Selection

In machine learning, the feature's selection procedure is a process of selecting the optimal facial features and discarding redundant features that increase the classification accuracy rate. To select the optimal features, a combination of PSO and GA is used.

2.4.1. Genetic Algorithm

In 1970, Holland provided the idea of a GA. GA's are stochastic search algorithms modelled on the procedure of natural selection; that underlines biological evolution. GA's has successfully been applied in many search, optimization, and machine learning problems. GA's is used to simulate procedures in natural systems that are vital for evolution. GA searches for the best solution in search space intelligently and works in an iterative order, such that new generation is created from the old one. The strings in GA's are represented as binary codes. In GA's, a fitness function computes the fitness of every string. The standard operators used in GA's are mutation, crossover and selection.

2.4.2. Particle Swarm Optimization

PSO is a population-based stochastic optimization technique that was developed by Kennedy and Eberhart in [9]. A particle in search space can be considered as "an individual bird of a flock." To find the best solution PSO uses local and global information. Fittest solution is found by utilizing the fitness function and velocities of the particles. In PSO, the particle position gets updated by using local and global positions of every particle around its neighbor. The particles move across the problem space searching for the optimal particles (features). The procedure is then repeated for a fixed number of periods or until a minimum, error is attained [9].

2.4.3. K-Nearest Neighbour

The K-NN acquaintance method was first introduced by Fix and Hodges in 1951 and is one of the most accepted nonparametric methods [4]. K-NN classifies the new data on the basis of the knowledge of training data. Random procedures are used to solve the lazy problems. Working procedure of K-NN is that it calculates the distance from the query instance and the training data. In our work, we used leave-one-out cross validation strategy of 1-NN and distance between neighbours is calculated using Euclidean distance. 1-NN strategy is easy to implement as it is independent of user specified parameters. The particles are coded as binary strings i.e., $S=F1, F2, \ldots Fn, n=1, 2, \ldots m$, where bit value '1' represents selected features and '0' represent unselected features. The initial population is generated randomly and Leave-one-out cross validation strategy is used to evaluate the fitness value. To select the best solution from population, mutation and cross-over operators are applied without modification. We used rand-based roulette-wheel scheme. In this work the two cutting points are chosen using the 2 point crossover operator. First the check is performed for mutation and if mutation exists then the offsprings are mutated and the code conversion is performed from 1 to 0 and similarly from 0 to 1. Next if the mutated chromosomes have more worth than parents then replacement is performed with the worst parent and in another case the replacement is performed with the low quality one. The performance of GA is enhanced by applying the PSO for each new generation. In this work, the adaptive values are used to renew the particle. pbest and gbest are the best adaptive values in pbest group. The particle position and speed are tracked after the values of pbest and gbest are obtained. The below equations are used to perform the updating process.

$$v_{pd}^{new} = wxv_{pd}^{new} + c_1 x \; rand_1 x(pbest_{pd} - x)_{pd}^{old} + c_2 x \; rand_2 x(pbest_{pd} - x)_{pd}^{old} \tag{2}$$

$$s(v_{pd}^{new} = \frac{1}{1+e^{-v_{pd}^{new}}}) \tag{3}$$

$$if(rand < S(v_{pd}^{new}))then \; x_{pd}^{new}=1; else \; x_{pd}^{new}=0 \tag{4}$$

Before updating the particle, the worth of gbest is checked to prevent particle from getting trapped in local optimum. The gbest value is updated if it is found that its value remains a number of times. Equation 2 is used to calculate the features after the update. Here represents the velocity.

The features are selected and represented by "1" if value of greater than random produced disorder {0.0~1.0} and the features are unselected and represent by "0" if value of is less than the random produced disorder {0.0~1.0}. The GA was configured to encompass 20 chromosomes and was run for 100 generations in each trial. The crossover rate and mutation rate was 1.0 and 0.1 respectively. The number of particles utilized is 20. The two factors $rand_1$ and $rand_2$ are random numbers between (0, 1), whereas C_1 and C_2 are acceleration (learning) factors, with $C_1=C_2=2$. The inertia weight 'w' was set as 0.9. The maximum number of iterations utilized in our PSO was 100.

2.5. Classification using SVM

The gender classification problem can be thought of as a two class problem and the goal of this problem is to separate the two classes by mean of a function. SVM is a useful technique for data classification and is easier to use than neural networks. SVM takes data (features) as input and predicts on the basis of training data set that in which class these features belong. The goal of SVM is to find the optimal hyper plan such that the error rate is minimized for an unseen test sample. According to the structural risk minimization principals and VC dimension minimization principle [16] a linear SVM uses a systematic approach to find a linear function with the lowest capacity.

The SVM classifier correctly separates the training data of labeled set of M training samples (X_i, Y_i), where $X_i \varepsilon RN$ and Y_i is the associated label i.e., $(Y_i \varepsilon \{-1,1\})$. The hyper plan is defined as:

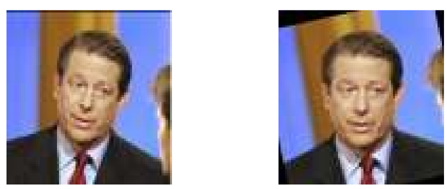
$$f(x) = \sum_{i-1}^{M} y_i \infty_i k(x, x_i) + b \tag{5}$$

3. Results and Discussions

We used MATLAB 2009a environment for our experiments. Label Faces in the Wild (LFW) [7] face database is used for experiments. This database contains 13,233 images collected from a web. To remove unwanted area, face portion is extracted using spatial co-ordinate system and then as shown in Figure 5, histogram equalization is applied to normalize the face image. We selected 400 face images, 200 male and 200 female for our experimental setting. As shown in Figure 6, all the faces are aligned using commercial aligned software Wolf *et al.* [20].

a) Face image before normalization. b) Face image after normalization.

Figure 5. a and b visually represents the image before and after normalization process.



a) Orignal face image. b) Align face image.

Figure 6. a and b visually represents original and align image.

We did avoid those images for which it was difficult to establish the ground truth as shown in Figure 7.

Figure 7. Images avoided to select.

5-fold cross validation is used in all experiments. First, the image is resized to 32×32, and then it is divided into blocks of size 8×8. DCT is applied to each block. From each block, four coefficients of highest importance are selected. For geometric-based feature's extraction, first the five facial points are located using active shape model and then LPB is used to extract the local features from these points. Figure 8 shows the classification accuracy of DCT-SVM and LBP-SVM based features; extracted using different features set (i.e., 100, 200, 300 and 400). Figure 8 also shows that by using FS-400 we get the accuracy of 78%, which is higher as compared to other feature's set (like FS-100, FS-200 and FS-300) accuracy rate.

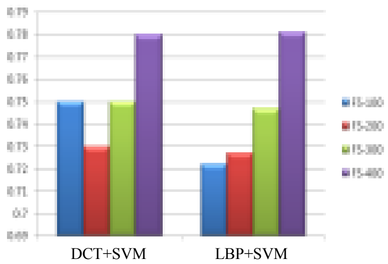


Figure 8. DCT-SVM and LBP-SVM based extracted features classification accuracy.

After extraction of global and local features, fusion of features is performed using concatenation method and features vectors of size 500, 600 and 700 are generated. Figure 9 shows the classification accuracy rate after feature's fusion. Accuracy increases with this feature fusion method, but the drawback is that the data dimension size also increases. So, to reduce the data dimension size, optimized features are selected in the next step.

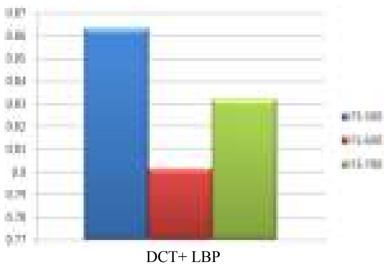


Figure 9. Classification accuracy after features fusion.

In the next step GA and PSO are combined using leave-one-out cross validation strategy of 1-NN and distance between neighbours is calculated using Euclidean distance. The features set (i.e., 500) are passed to GA-PSO, which eliminates the redundant features and GA-PSO get back the result as optimized features (i.e., 40). After optimization step the SVM is trained and tested by optimized features. 1:3 and 3:1 testing to a training ratio is used for SVM. Figure 10 shows the accuracy rate after feature's optimization. The results depict that the hybrid GA-PSO is very useful for features optimization. By using this technique, the enhancement in the classification accuracy rate (i.e., 98%) is noteworthy and it also reduced the data dimension size (i.e., from 500 features set to 40 features) that is shown in Figure 10.

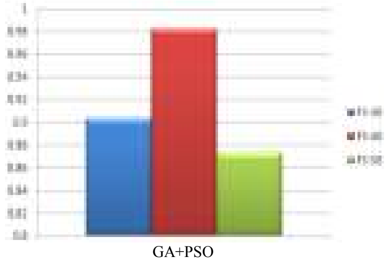


Figure 10. Optimized features classification accuracy.

Table 1 presents a comparison of our proposed technique with other gender classification techniques. In our proposed technique, we overcome the problems like high dimensions (i.e., by utilizing the minimum number of features), variation in pose or occlusions (i.e., by combining local and global features). We used real-life images that have a large amount of variations in facial expression and pose. The database that we have used is also publicly available, which helps to benchmark for future. In Table 1, proposed technique using real-life face images and it is also publicly available as compared to other techniques face databases. The high recognition rate of 98% is obtained.

Table 1. Propsed method comparison with state-of-the-art gender classification techniques.

Methods	Data Dimensions	Real Life	Publically Available	Recognition Rate
Proposed	**30**	**Yes**	**Yes**	**98%**
Sun *et al.* [18]	150	No	No	95.3%
Jain and Hung [8]	200	No	Yes	96%
Bluja *et al.* [2]	2409	No	Yes	94.3%
Nazir *et al* [14]	256	No	Yes	99.3%
Shakhnarovich *et al.* [15]	35,00	Yes	No	79%
Gao *et al.* [5]	10,100	Yes	No	95.5%
Shan. [16]	2891	Yes	Yes	95%
Khan *et al.* [10]	20	No	Yes	95.63%

4. Conclusions and Future Work

Gender classification is believed to be one of the most active research areas of pattern recognition. Currently, most of the acclaimed work in this area revolves around a frontal facial picture established classification. However, we have performed experiments on real-world face images. In this work, both geometric and appearance-based features are extracted and fused, which makes the systems supportive of variations in expressions and poses. Furthermore, we also concentrated on reducing the data dimensions and endeavoured to produce optimal features set that extra precisely embodies a gender face. If inadequate features are used, even the best classifiers normally fail to accomplish higher accuracy. Therefore, in this paper, we present a method for optimizing the features by employing hybrid GA-PSO algorithms. After an optimization process, we received 98.3% classification accuracy rate using LFW images database. Furthermore, we could reduce a smaller set of features. This additionally reduces the time complexity and makes the system fast. Furthermore, we aim to discover Swarm-based Optimization Algorithms (SOAs) like ant dominion optimization and make our arrangement more precise and stable.

References

[1] Azeem A., Sharif M., Raza M., and Murtaza M., "A Survey: Face Recognition Techniques under Partial Occlusion," *the International Arab Journal of Information Technology*, vol. 11, no. 1, pp. 1-10, 2014.

[2] Bluja S. and Rowley H., "Boosting Sex Identification Performance," *the International Journal of Computer Vision*, vol. 71, no. 1, pp. 111-119, 2007.

[3] Brunelli R. and Poggio T., "HyberBF Networks for Gender Classification," *in Proceedings of DARPA Image Understanding Workshop*, Washington, USA, pp. 311-314, 1995.

[4] Cover T. and Hart P., "Nearest Neighbor Pattern Classification," *in Proceedings of IEEE Transaction Information Theory*, NJ, USA, pp. 21-27 , 1967.

[5] Gao W., Ai H., and Moghaddam B., "Face Gender Classification on Consumer Images in a Multiethnic Environment," *in Proceedings of the International Conference on Biometrics*, Madrid, Spain, pp. 168-174, 2009.

[6] Golomb B., Lawrence D., and Sejnowski T., "Sexnet: A Neural Network Identifies Sex from Human Faces," *in Proceedings of the Advance in Neural Information Processing Systems*, CA, USA, pp. 572-577, 1990.

[7] Huang J., Ramesh M., and Berg T., "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," *Technical Report*, University of Massachusetts, Ohaio, USA, 2007.

[8] Jain A., Huang J., and Fang S., "Gender Classification using Frontal Facial Images," *in Proccedings of IEEE International Conference on Pattern Recognition*, Florida, USA, pp. 1-4, 2008.

[9] Kennedy J. and Eberhart R., "Particle Swarm Optimization," *in Proceeding of the IEEE International Conference on Neural Networks*, Perth, Western Australia, pp. 1942-1948, 1995.

[10] Khan A., Nazir M., Naveed N., and Riaz N., "Efficient Gender Classification Methodology using DWT and PCA," *in Proceedings of the 14th IEEE International Multitopic Conference*, Karachi, Pakistan, pp. 155-158, 2011.

[11] Kumar N., Belhumeur N., and Nayar K., "A Search Engine for Large Collections of Images with Faces," *in Proceeding of the European Conference on Computer Vision*, FL, USA, pp. 340-353, 2009.

[12] Lian X., Lu B., Cooper D., and Graham J., "Gender Classification by Combining Facial and Hair Information," *in Proceedings of Advances in Neuro-Information Processing*, Bangkok, Thailand, pp. 647-654, 2009.

[13] Majid A., Khan A., Mirza M., and Moghaddam B., "Gender Classification using Discrete Cosine Transformation: A Comparison of Different Classifiers," *in Proceeding of the International Multi Topic Conference*, Islamabad, Pakistan, pp. 59-64, 2003.

[14] Nazir M., Ishtiaq M., Batool A., Jaffar A., and Mirza M., "Feature Selection for Efficient Gender Classification," *in Proceedings of the 11th WSEAS International Conference*, Iasi, Romania, pp. 70-75, 2010.

[15] Shakhnarovich G., Viola P., and Moghaddam B., "A Unified Learning Framework for Real Time Face Detection and Classification," *in Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, USA, pp. 14-21, 2002.

[16] Shan C., "Learning Local Binary Pattern for Gender Classification on Real-World Face Images," *Pattern Recognition Letter*, vol. 33, no. 4, pp. 431-437, 2012.

[17] Sun N., Zheng W., Sun C., Zou C., and Zhao L., "Gender Classification Based on Boosting Local Binary Pattern," *in Proceedings of the 3rd International Conference on Advances In Neural Networks*, Chengdu, China, pp. 194-201, 2006.

[18] Sun Z., Bebis G., Yuan X., and Louis S., "Genetic Feature Subset Selection for Gender Classification: A Comparison Study," *in Proceeding of the 6th IEEE workshop on Applications of Computer Vision*, Florida, USA, pp. 165-170, 2002.

[19] Vapnik V., *The Nature of Statistical Learning Theory*, Springer-Verlag, Berlin, German, 1995.

[20] Wolf L., Hassner T., Taigman Y., "Similarity Scores Based on Background Samples," *in Proceeding of the Asian Conference on Computer Vision*, Xi'an, China, pp. 1-10, 2009.

Sajid Khan is a Lecturer in Software Engineering Department at Foundation University, Pakistan. His research areas are pattern recognition, face analysis and digital image processing. He received his BS computer science degree from the PMAS University of Arid Agriculture, Pakistan. He completed his MS degree in computer science from SZABIST University Islamabad, Pakistan.

Muhammad Nazir received his MSc degree in computer science from PMAS-University of Arid Agriculture Rawalpindi, Pakistan and his MS degree in computer science from National University of computer and emerging sciences, NU-FAST, Pakistan. Currently, he is pursuing his PhD degree from the same University. He is serving as Assistant Professor at University of Central Punjab, Lahore. His research interests include digital image processing and computational intelligence.

Naveed Riaz is an Assistant Professor in Computer Science Department at Dammam University, Saudi Arabia. His research areas are pattern recognition, model-based diagnosis and theoretical computer sciences. He received his PhD degree from Graz University of Technology Austria and an MS degree from National University of Science and Technology, Pakistan.

Muhammad Khan obtained PhD degree in computer system engineering. His research interests are in the fields of software engineering, cloud computing, cyber administration, digital forensic analysis and machine learning techniques.