Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2013) 21: 1479 – 1494

doi:10.3906/elk-1203-51

Optimized features selection for gender classification using optimization algorithms

Sajid Ali KHAN,[1] **Muhammad NAZIR,**[2] **Naveed RIAZ**[1,*]
[1]Department of Computer Science, Szabist University, Islamabad, Pakistan
[2]National University of Computer & Emerging Sciences,FAST, Islamabad, Pakistan

Received: 12.03.2012 • **Accepted:** 29.05.2012 • **Published Online:** 12.08.2013 • **Printed:** 06.09.2013

Abstract: Optimized feature selection is an important task in gender classification. The optimized features not only reduce the dimensions, but also reduce the error rate. In this paper, we have proposed a technique for the extraction of facial features using both appearance-based and geometric-based feature extraction methods. The extracted features are then optimized using particle swarm optimization (PSO) and the bee algorithm. The geometric-based features are optimized by PSO with ensemble classifier optimization by the genetic algorithm. Using this approach, we have obtained promising results in terms of the classification error rate and computation time minimization. Moreover, our optimized feature-based method is robust to illumination, noise, and occlusion changes.

Key words: Gender classification, facial features, particle swarm optimization, genetic algorithm, bee algorithm

1. Introduction

Gender classification using facial images has a wide scope of applications, such as in customer oriented advertising, human-computer interface, and demographics. Feature extraction is an important subtask of gender classification. Gender classification approaches are categorized into 2 classes based on feature extraction. These are appearance-based feature extraction (also known as global features) and geometric-based feature extraction (also known as local features).

In the geometric-based method, features are extracted from some facial points like the face, nose, and eyes. Burton et al. [1] reported an 85% accuracy rate after locating 73 facial points. The facial points' extracted features are then passed to a discriminant analysis classifier to classify the gender. Fellous et al. [2] reported a 90% accuracy rate after finding 22 normalized distances using a face database containing 109 images. Li et al. [3] classified the gender by utilizing not only the 5 facial features (nose, eyes, mouth, forehead, and eyebrows) but also external information like clothes and hair features. They performed experiments on FERET, BCMI, and AR face datasets. The problem with their approach is that their feature extraction method was affected by a complex background.

In an appearance-based method, features are extracted from the whole face instead of extracting features from facial points. Colomb et al. [4] reported a 91.9% accuracy after performing experiments on a face database containing 90 images. They used the SexNet network to classify gender. Nazir et al. [5] used discrete cosine transform (DCT) to extract the facial features. The K-nearest neighbor (KNN) classifier was trained and tested

*Correspondence: n.r.ansari@szabist-isb.edu.pk

by these features. Experiments were performed on the Stanford University Medical Students (SUMS) face database. The problem with their proposed method is that it is not robust to occlusion changes. Mousavi et al. [6] reported an 87.5% accuracy rate after they applied their fuzzy inference system to the Informatics and Mathematical Modelling (IMM) face image database, which contains 240 face images. They extracted geometric-based facial features by calculating the distance between different face points. The problem with this method is the proper adjustment of the threshold value. Rai and Khanna [7] proposed a new technique to classify gender from face images. They combined wavelet and Radon transform to extract the important facial features. They performed experiments on the SUMS face database and reported a 90% accuracy rate using 35 features. Their system consumed more time when locating the face portion and also did not support any pose changes. Ravi and Wilson [8] presented a novel gender classification strategy. The face portion was located using skin color and then after extracting the geometric-based facial features, they applied a support vector machine (SVM) to classify the gender. The drawback of this method is in choosing the correct threshold value for the facial feature extraction. Shobeirinejad and Gao [9] extracted discriminative facial features using an interlaced derivative pattern (IDP). Experiments were performed on the Face Recognition Grand Challenge (FRGC) face database and yielded a 91.2% classification accuracy rate. Xu et al. [10] fused the appearance- and geometric-based features to classify gender. The appearance-based features were extracting using Haar wavelets, and for geometric-based feature extraction, they used an active appearance model. Their method is robust to illumination, pose, and expression changes. To extract the geometric-based features, they have to locate 83 facial points for each face, which makes their system more time-consuming.

Generally, gender classification consists of the following steps. Figure 1 depicts these steps.

- **Preprocessing:** Every face database needs some preprocessing, such as normalization of the illumination and face detection.
- **Feature extraction:** After performing the preprocessing, we need to extract the facial features. Generally, 2 types of features are extracted, namely geometric-based features and appearance-based features.
- **Classification:** Classification is the last step of gender classification, in which the face is successfully classified as that of a male or a female. For this purpose, different types of classifiers are used, e.g., KNN, neural network (NN), and SVM.

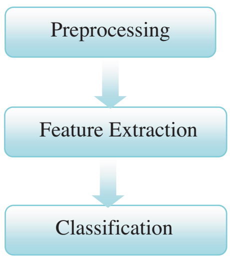


Figure 1. General steps of gender classification.

The practical applications of gender classification still suffer from problems like occlusion change, illumination effect, and computation time. To overcome the occlusion problem, we extract both appearance-based (using DCT) and geometric-based (using local binary pattern [LBP]) features. To minimize the illumination effect, histogram equalization is performed. The extracted features are then optimized to minimize the computation time and increase the accuracy rate. The facial features are evaluated using particle swarm optimization

(PSO) and the bee algorithm (BA). These provide us with the optimized features. Ensemble-based classifiers are trained and tested using the optimized features and the weight of the ensemble classifier is optimized by the genetic algorithm (GA). This provides us with a higher accuracy rate.

The main contributions of our work are:

- Formulation of the new feature selection algorithm. The algorithm is applied to appearance- and geometric-based features, namely DCT and LBP, and it selects the optimal features that increase the gender classification accuracy rate.
- Rather than using the whole face portion, only some of the important facial features are extracted, which reduces the computational complexity with improved accuracy (i.e. using only 15 features, we get a 97.5% accuracy). We use the geometric-based method to extract the facial features (i.e. locating 5 facial points), which makes our proposed method robust to occlusions and noise.

2. Proposed method

The main steps of the proposed method are given below.

The input is the FERET and SUMS face databases.

- First, histogram equalization is applied to the test the images, which results in the minimization of the illumination affects.
- In the second step, the face portion is detected using the Viola and Jones technique [11].
- Next, DCT is used to extract the appearance-based features and LBP is used to extract the geometric-based features.
- We use the BA and PSO to optimize our geometric- and appearance-based extracted features.
- The optimized features are then passed to the ensemble-based optimized classifier, which results in a higher accuracy rate.

Figure 2 shows our proposed system's architecture.

2.1. Face detection

In [11], Viola and Jones presented a new face detection technique known as cascade face detection. Their technique searches the face portion; starting from the top left corner, it goes down to the bottom right corner. Their technique has 3 main steps: first, to make the computation process faster, the image is represented as an 'integral image'. Next, to select efficient features, they use the Adaboost algorithm. In the last step, the background region is eliminated using a cascade of the Adaboost classifiers. Figure 3 shows the extracted faces after applying the Viola and Jones technique.

2.2. Feature extraction

The first step in gender classification is the extraction of the features. In this paper, the DCT and LBP techniques are used for the appearance- and geometric-based feature extractions.

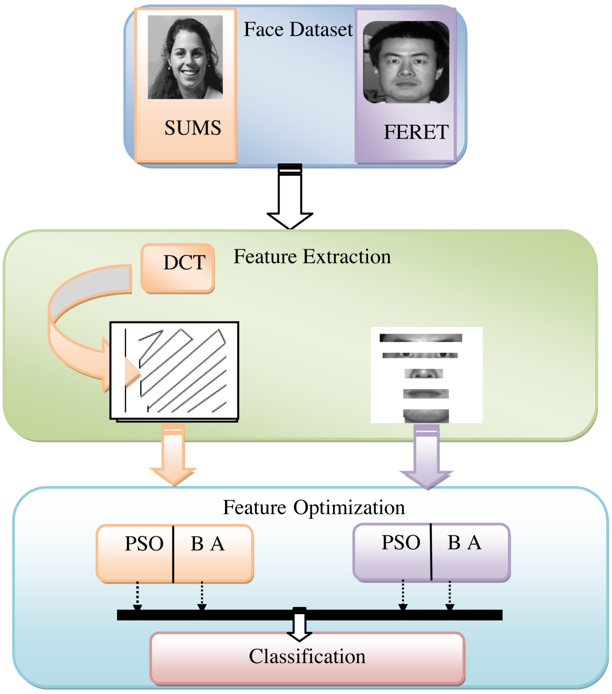


Figure 2. Proposed system's architecture.

Figure 3. A sample of the detected faces by Viola and Jones.

2.2.1. DCT (global feature extraction)

DCT is a popular transformation technique used in signal and image processing. DCT was found to be a highly effective method for gender classification that yields a high accuracy rate with a low computational cost. It separates the image into different parts with respect to their importance and transforms the input into a linear combination of weighted basis functions. The general equation of DCT for a f (x, y) image is defined as:

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

$$\cos\left[\frac{(2m+1)u\pi}{2M}\right]\cos\left[\frac{(2n+1)v\pi}{2n}\right], \tag{1}$$

where $a(u) = \begin{Bmatrix} \sqrt{\frac{1}{M}} & \text{for} & u = 0 \\ \sqrt{\frac{2}{M}} & \text{for} & u = 1, 2, \ldots \quad M-1 \end{Bmatrix}$

and $a(v) = \begin{Bmatrix} \sqrt{\frac{1}{N}} & \text{for} & v = 0 \\ \sqrt{\frac{2}{N}} & \text{for} & v = 1, 2, \ldots \quad N-1 \end{Bmatrix}.$

First, the image was resized to 32 × 32 pixels, and then blocks of 8 × 8 pixels were generated. We applied DCT to the 8 × 8 pixel blocks. After applying DCT, the coefficients with a high variance were selected in a zigzag manner. As these contribute more to representing the gender, the dimensions are reduced. Hence, the coefficients with a higher variance get placed at the top left corner. We scanned the DCT coefficients in a zigzag manner, starting from the top left corner, and converted it to a one-dimensional vector. We selected the first coefficient from each of the 16 blocks, which means that if our feature vector size is 32, then we get the first 2 coefficients from every block. Feature vectors in the size of 100, 150, and 200 were created from these important coefficients. Figure 4 depicts the DCT-based feature extraction process.

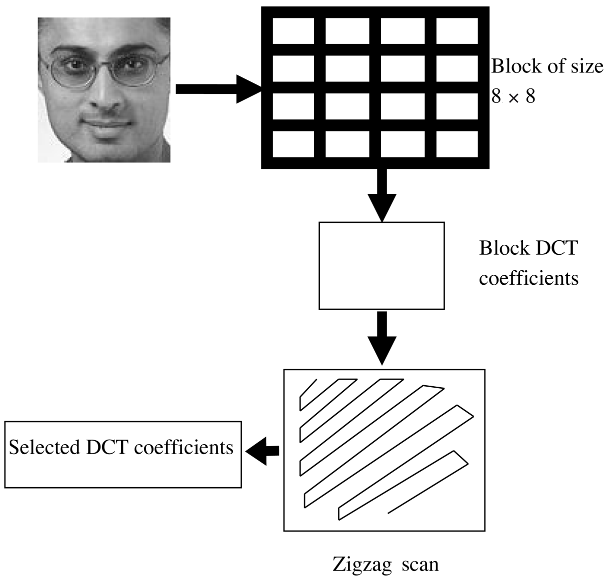


Figure 4. DCT-based feature extraction process.

2.2.2. LBP (local features extraction)

The features extracted from the whole face are known as global features. Global features are dependent on facial occlusions and alignment, while those local features that are extracted from the different points of a face are subjected to the variation of facial expressions, illumination, and occlusions [12]. Psychological experiments show that individual facial features like the nose, mouth, eyes, eyebrows, and chin carry more information when compared to the whole face [13]. We first locate 5 facial components (nose, mouth, eyebrows, chin, and eyes) using an active shape model [14], and then these facial features are cropped and LBP is applied to extract the features from these cropped components. Figure 5 depicts the facial components of a face.

Figure 5. The 5 extracted facial components.

LBP was used by Lian and Lu [15] for gender classification. In LBP, the center pixel is selected and the neighborhood pixels are converted to 0 if the gray levels are smaller than the center or to 1 if the gray levels are greater than the center. The center pixel is then replaced by the binary code of its neighborhood, like 00011011, as shown in Figure 6.

1	2	1
5	4	6
7	1	9

0	0	0
1		1
1	0	1

Figure 6. LBP descriptor.

2.3. Feature optimization

The global and local features are optimized by PSO and the BA.

2.3.1. Particle swarm optimization

The global and local features are optimized by PSO and the BA.

PSO is an evolutionary optimization technique that was proposed by Kennedy [16] in 1995. We have implemented a binary PSO (BPSO) to evaluate the optimized features. BPSO was developed by Kennedy and Eberhart [17]. The particles in the BPSO travel in the search space and flip different bits. The velocity can be described by the number of bits changed at every iteration. The particle position is denoted by either 0 or 1.

$$X_i\,(t+1) = \left\{ \begin{array}{c} 1 \; if \;\; N < \dfrac{1}{1+e^{-V_i}\;\,(t+1)} \\ Otherwise \qquad 0 \end{array} \right\} \tag{2}$$

Eq. (2) is a binary activation function for the X_i position and it identifies whether the particle position should be updated or not.

Each particle is represented as a binary value, P = F1 F2 ... n, n = 1, 2 ... m, where m represents the length of the feature vectors. If our data dimension size is 10, i.e. n = 10, then the binary analyzes the P =

F1 F2 ... F10 features. PSO randomly selects the optimized features (F1 F4 F6 F8 F9) by setting the 1, 4, 6, 8, 9 bits to ON.

We take 100 particles and associate a feature vector with each particle. Features carrying less discriminative information about the gender are eliminated (OFF) by the BPSO during each iteration and the remaining features are the required optimized features. This step is repeated up to the number of feature vectors.

2.3.2. Bee algorithm

The global and local features are optimized by PSO and the BA.

Pham et al. [18] proposed the BA in 2006. This algorithm is based on the act of honey bees searching for food. This algorithm has found many applications in the areas of engineering, data clustering, multiobjective optimization, and job scheduling for a production machine.

The algorithm starts when a spy bee examines the region. The bees travel randomly and gather information about the food. When the spy bees return to the hive, they perform a waggle dance to inform the other bees about the flower patches. The colony evaluates the flower patches' relative merits because of the waggle dance and sends other bees to more promising regions. This act of searching for food is mapped as exploitation. When spy bees search for food randomly in the whole search space, this is known as exploration. For the algorithm to work, we need to set the following parameters: the number of spy bees (n), number of bees deployed in the elite sites (nep), and number of elite sites selected based on the fitness (m). First, the spy bees give the initial results after exploring the search space. Next, 'm' out of 'n' sites are selected for exploitation after the computation of the fitness of the results by the bees. The number of bees sent to exploit the region is proportional to the fitness of the region. The fittest bee from each patch is chosen. At the end, the solutions are ranked based on their fitness. The algorithm continues until a specified criterion has been found.

The steps of our algorithm, which are inspired by [18], are the following.

- First, select different feature vectors.
- Fitness (accuracy) is computed for each feature set.
- If the fitness criteria are not met, then new feature vectors are selected and the fitness is evaluated again.
- Only those feature vectors that meet the fitness criteria are selected from each patch.
- This process is repeated until the stopping criteria is met.

2.4. Classification

In this step, we ensemble different classifiers (more than one) to get a high accuracy rate.

2.4.1. Ensemble classifiers

Combining the outcome of different classifiers is known as the classifier ensemble [19]. It has been observed by researchers that the performance of the ensemble classifiers is better than that of a single classifier. Some famous ensemble techniques are the mean, median, majority voting, and product-based techniques [20].

We trained and evaluated 3 different classifiers on the optimized features. These classifiers are SVM, back propagation NNs (BPNNs), and KNN. These classifiers are then optimized through the GA.

2.4.2. Weighted majority voting

Normally, different classifiers have different accuracy rates on different dataset attributes. Classifiers with less accuracy are not efficient for the dataset compared to classifiers with a high accuracy. Each classifier is weighted in such a way that the contribution of a classifier in the final decision is proportional to its accuracy [21].

The classifier Si support for class Cj of the given sample x is Ri,j:

$$Rij = \left\{ \begin{array}{ll} 1\, if\ Si\ labels\ x\ in\ Cj \\ o, \quad otherwise \end{array} \right\}.$$

For class Cj, the majority vote support is Mj(x):

$$Mj\,(x) = \sum_{i=1}^{l} b_i\ d_j = c\ \sum_{i=1}^{l} b_i\ d_j,$$

where the b_is are the coefficients for the classifiers Si.

The value of the function is the sum of the coefficients of these numbers of ensembles whose output for x is Cj [22].

2.4.3. Ensemble classifier optimization using the GA

We use the GA to optimize the weights of the ensemble classifiers. First, the GA normalizes the weights of the classifiers between [0, 1]. The length of classifier L is equal to chromosome m. The ensemble classifiers' weights are obtained by applying mutation and crossover. When the generation reaches G_{max} or the population converges to a satisfactory solution [23], the GA is then terminated. Figure 7 depicts these steps.

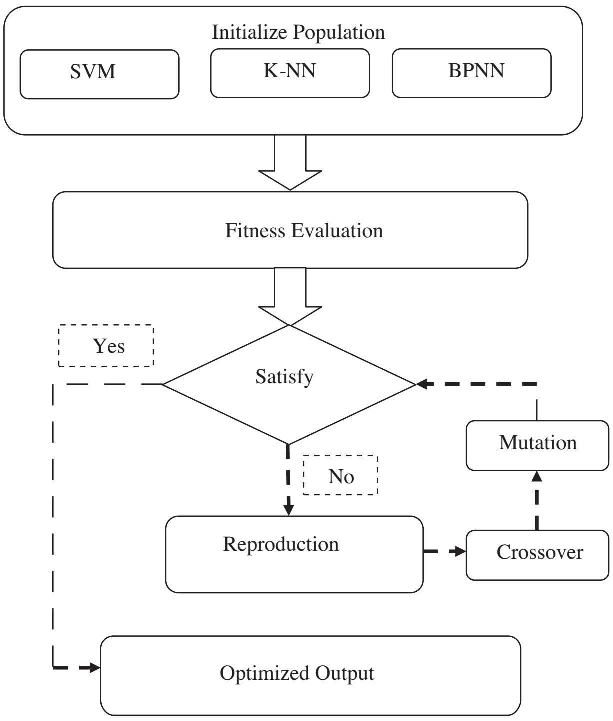


Figure 7. Classifier ensemble and optimization.

3. Experimental results and discussions

We used a MATLAB environment to implement our proposed system. Two face databases, FERET [24] and SUMS, were used to evaluate our system. The FERET database contains thousands of images, out of which we selected 200 images having pose and illumination variations. The SUMS database contains 400 images, 200 male and 200 female. Figure 8 shows sample images from the face datasets. We performed experiments on these databases using a 1:9 and 9:1 training-to-testing ratio. First, we detected the face portion using the Viola and Jones algorithm [11], and then histogram equalization was performed to normalize the illumination effect, as shown in Figure 9. The face image was subdivided into 8 $\times$ 8 blocks. We then picked each block and computed the important coefficients using DCT in a zigzag manner. These selected coefficients were then passed to the BA and PSO to select more optimized features in order to increase the accuracy rate and minimize the computational cost.

To optimize the weights of the ensemble classifier, the attributes of the GA are set as provided below.

- **Chromosome length:** m = 3 (we ensemble 3 classifiers).
- **Population size:** N = 30 (set small to reduce the computational complexity).
- **Maximum generation:** Gmax = 1000.
- **Cross over rate:** cr = 0.7.
- **Mutation rate:** mu = 0.01.

On average, our GA converges at a value of 53.

Figure 8. FERET and SUMS face database samples.

Figure 9. a) Face before normalization and b) face after normalization.

3.1. Experiments using the global features

The DCT-extracted features are passed to the BA, which eliminates the less important features and provides the optimized features, resulting in a reduction in the dimensions.

3.1.1. Experiments on the BA

The attributes of the BA are set as provided below.

- **Scout bees:** n = 80.
- **Fitness criteria:** maximum accuracy.
- **Elite regions:** 10.
- **Elite bees:** nep = 3.
- **Termination criteria:** max accuracy > 0.98 or iteration > 100.

First, 100 DCT-extracted features are selected randomly, and then the fitness (accuracy) for each feature set is found. Table 1 contains the DCT-based extracted features with their individual feature accuracies. These features are processed through the BA and their accuracy is evaluated using the fitness function (i.e. max accuracy).

Table 1. Accuracy of 100 features.

1	2	3	4	5	......	......	......	99	100
30%	20%	60%	80%	20%				55%	90%

Second, 30 features with high accuracy are selected and are sorted in descending order, as shown in Table 2.

Table 2. Sorted accuracy rate in descending order.

1	2	3	4	5	......	......	......	29	30
90%	85%	83.6%	83%	81.5%				80.1%	80%

The best features, having high accuracy, are then selected out of these 30 features, as shown in Table 3.

Table 3. The 4 best selected features.

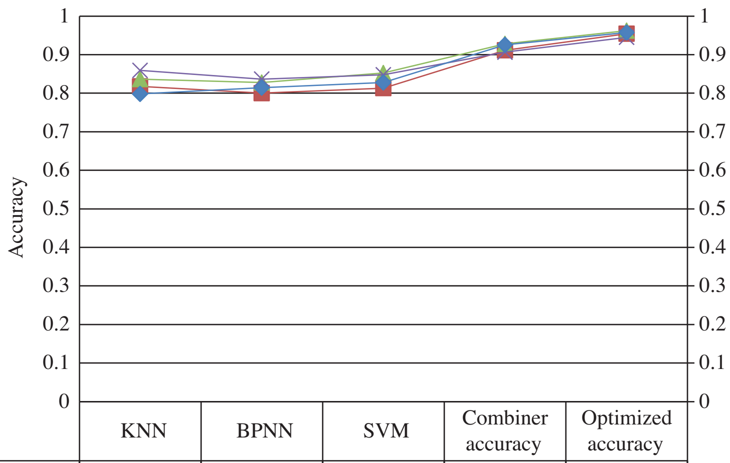
1	2	3	4
90%	85%	83.6%	83%

The best optimized features are passed to the ensemble classifier and the weights of the ensemble classifiers are optimized through the GA, resulting in an increase in the accuracy of the features.

Figure 10 shows the results of the different classifiers using different sets of features. We have calculated the classifier's accuracy and the ensemble classifier's accuracy separately. The graph depicts that the accuracy of the system is noticeably improved after the ensemble classifiers' weights are optimized by the GA.

We may conclude from Figure 10 that after passing the BA-based optimized features to a single classifier (i.e. KNN, BPNN, or SVM), the results suggest a lower accuracy rate than expected. For example, the first

BA-based optimized feature set has a 90% accuracy rate, but when this feature set is tested and trained by the KNN classifier, we then receive a 79.8% accuracy rate. When we use the same feature set to train and test our GA-based optimized ensemble classifier, we get a 95.7% accuracy rate.



	KNN	BPNN	SVM	Combiner accuracy	Optimized accuracy
FS2	0.8177	0.8	0.8132	0.9113	0.9543
FS3	0.8365	0.8276	0.8523	0.9279	0.9622
FS4	0.859	0.8365	0.8476	0.9064	0.9451
FS1	0.798	0.8143	0.8276	0.9254	0.9574

Figure 10. BA-based optimized feature accuracy comparison using different classifiers.

3.1.2. Experiments on PSO

In these experiments, we optimize the DCT-based extracted features using the BPSO technique. The DCT features (100, 150, 200, and 250) are passed to the BPSO and the BPSO eliminates those redundant features that are less important, and the resultant optimized features are 40, 60, 75, and 90. These optimized features are used to train and test our ensemble classifier, and then the GA is used to optimize the weights of the ensemble classifier, which results in a higher accuracy rate. The graph in Figure 11 depicts our proposed optimized-based system's accuracy. First, the optimized features are passed to single linear classifiers and the accuracy is evaluated. Next, we ensemble the classifiers using the weight majority technique and the received results are comparatively better (max. 92%) than the single classifier results (max. 84%). We again compute the accuracy after the ensemble classifiers' weights are optimized by the GA and the received results are even better (max. 94).

FS1 = 40 features, FS2 = 60 features, FS3 = 75, and FS4 = 90 features.

3.2. Experiments using local features

We optimize the LBP extracted features (geometric-based) using the BA and PSO. First, we pass the facial features to the BA.

3.2.1. Experiments on the BA

First, 50 LBP geometric-based extracted features are selected randomly. Next, the fitness (accuracy) for each feature set is found. Table 4 represents these features.

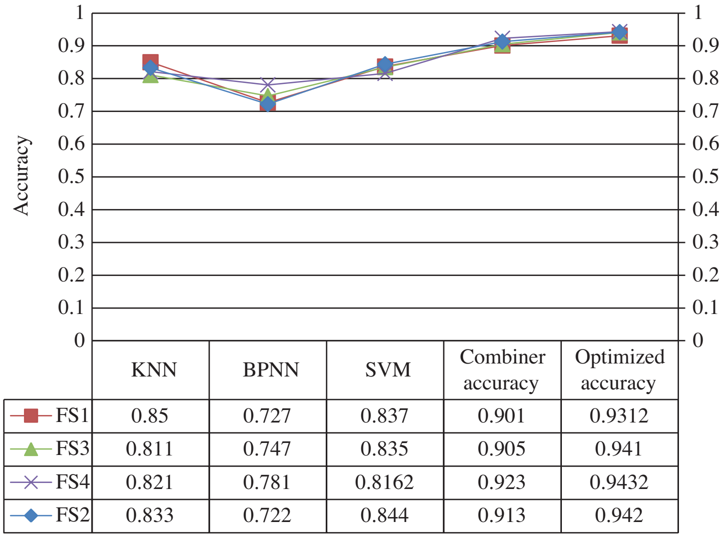


Figure 11. PSO-based optimized DCT features' accuracy comparison using the single and ensemble classifiers.

Table 4. Fitness of the LBP geometric features.

1	2	3	4	5	......	......	......	49	50
5 %	6.6%	10%	9%	70%				86%	84.3%

Second, 20 features with high accuracy are selected and sorted in descending order, as shown in Table 5.

Table 5. Twenty high-accuracy features sorted in descending order.

1	2	3	4	5	......	......	......	19	20
86%	84.6%	84.3%	82.9%	82%				81.5%	80%

The best 4 features having higher accuracy are then selected from these 20 features. Table 6 describes these features.

Table 6. The 4 best selected features.

1	2	3	4
86%	84.6%	84.3%	82.9%

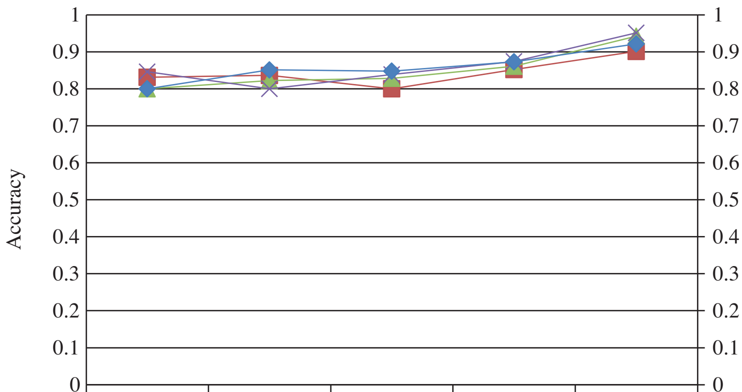
The 4 selected feature sets are passed to the ensemble of classifiers. Their weights are then optimized using the GA. The optimized features with a high accuracy are shown in Figure 12.

First, the 4 feature sets are separately used to train and test a single classifier. The received results are not encouraging (max. 85%). We have examined the fact that the accuracy increases (max. 87%) after we ensemble the classifiers, but when the weights of the ensemble of classifiers are optimized using the GA, we then receive very encouraging results (max. 95%) in comparison to the single and ensemble classifiers' accuracy rates.

3.2.2. Experiments on PSO

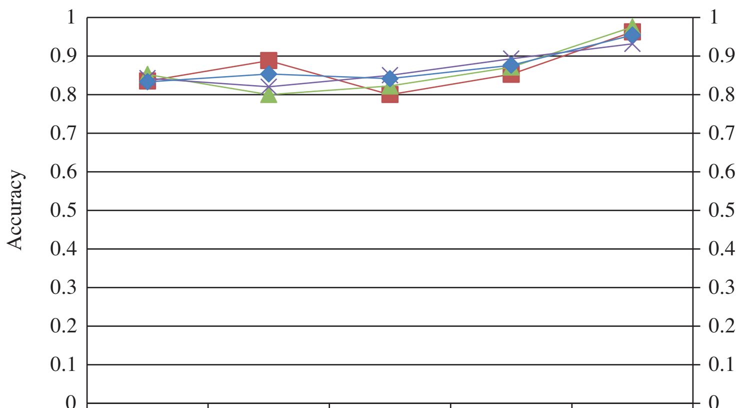
LBP geometric-based features that are extracted from the 5 facial components (eyes, nose, forehead, mouth, and eyebrows) are optimized through the BPSO. Four feature sets (20, 30, 40, and 50) are passed to the BPSO, which results in the optimized features (7, 10, 15, and 20). Our ensemble classifiers are trained and tested by these features. The weights of the ensemble classifiers are optimized through the GA. Figure 13 provides the comparison of the single classifier accuracy rates with our ensemble optimized classifiers. The single classifier high accuracy rate that we have achieved using 10 features is 88.8%, but the ensemble classifiers do not provide noticeable results when compared to the single classifier using the LBP features. We optimize the weights of the single classifier and achieve a 97.5% accuracy, which is probably the highest accuracy rate, utilizing a minimum number of features (i.e. 15) compared to the other techniques' accuracy rates.

FS1 = 7, FS2 = 10, FS3 = 15, and FS4 = 20.



	KNN	BPNN	SVM	Combiner accuracy	Optimized accuracy
FS2	0.8311	0.836	0.8	0.852	0.901
FS3	0.8	0.822	0.8284	0.861	0.9421
FS4	0.846	0.8	0.8385	0.874	0.9512
FS1	0.8	0.851	0.8478	0.873	0.9213

Figure 12. BA-based optimized LBP feature accuracy comparison using the single and ensemble classifiers.



	KNN	BPNN	SVM	Combiner accuracy	Optimized accuracy
FS2	0.835	0.888	0.8	0.8523	0.9623
FS3	0.852	0.8	0.822	0.871	0.975
FS4	0.8423	0.82	0.85	0.893	0.9312
FS1	0.833	0.853	0.841	0.876	0.9541

Figure 13. PSO-based optimized LBP feature accuracy comparison using the single and ensemble classifiers.

Finally, we compare our method with some state-of-the-art gender classification techniques. The results are shown in Table 7. Li et al. [3] used LBP to extract the facial features; however, their data dimensions were very high compared to our proposed technique. Due to the low dimensions, our proposed technique has a low time complexity. Nazir et al. [5] used DCT to extract the important facial features, and after passing these features to the KNN classifier, they achieved a 99.30% classification accuracy, which is relatively higher than that of our proposed technique, but the problem with their technique is that it is not robust to variations in the pose (e.g., mask and glasses). In comparison with all of these state-of-the-art gender classification techniques, we overcome problems like high dimensions (i.e. by utilizing the minimum number of features) and variation in the pose or occlusions (i.e. by extracting the geometric facial features).

Table 7. Proposed technique comparison with other techniques.

Methods		Database	Data dimensions	Pose variation support	Recognition rate
Proposed		FERET, SUMS	15	Yes	97.5%
Li et al. [3]	ASM, LBP	FERET	2891	Yes	95%
Nazir et al. [5]	DCT, K-NN	SUMS	256	No	99.30%
Mousavi et al. [6]	Neuro-fuzzy	IMM	240	Yes	87.5%
Rai et al. [7]	Radon + DCT, K-NN	SUMS	35	No	90%
Shobeirinejad et al. [9]	IDP	FRGC	128	No	91.2%
Sajid et al. [25]	IDP + PCA	SUMS	20	No	97%
Shan [26]	LBP, SVM	LFW	2891	Yes	95%

4. Conclusion and future work

Gender classification is one of the active areas of research in pattern recognition and image processing. Currently, most of the acclaimed work in this domain revolves around frontal facial image-based classification. In this paper, we have focused on reducing the data dimensions and have tried to produce a more optimal feature set that more accurately represents a gender's face. If inadequate features are used, then even the best classifier will fail to achieve a higher accuracy. Therefore, in this paper, we have tried to optimize the features using the BA and PSO. After the optimization phase, a large number of redundant and irrelevant features are eliminated, resulting in a reduction of the data dimensions. We extracted both the geometric- and appearance-based features for our experiments. We achieved a 97.5% classification accuracy rate using only 10 geometric features. We extracted the geometric-based features (i.e. locating facial points such as the nose, chin, and eyes), which makes our system more stable and able to support variations of facial expressions, illumination, and poses. We compared our results with some state-of-the-art gender classification techniques and found that our proposed technique provides a high classification accuracy by utilizing the minimum number of features, which also reduces the time complexity. Thus, we can conclude that the accuracy rate can be remarkably increased if the ensemble classifiers' weights are optimized using the GA. We intend to explore more swarm-based optimization algorithms, like ant colony optimization, and make our system more accurate and stable.

References

[1] A. Burton, V. Bruce, N. Dentch, "What's the difference between men and women? Evidence from facial measurements", Perception, Vol. 22, pp. 153–176, 1993.

[2] J. Fellous, "Gender discrimination and predication on the basis of facial metric information", Vision Research, Vol. 37, pp. 1961–1973, 1997.

[3] B. Li, X.C. Lian, B.L. Lu, "Gender classification by combining clothing, hair and facial component classifiers", Neuro Computing, Vol. 76, pp. 18–27, 2011.

[4] B. Golomb, D. Lawrence, T. Sejnowski, "SexNet: a neural network identifies sex from human faces", Advances in Neural Information Processing Systems, Vol. 3, pp. 572–577, 1991.

[5] M. Nazir, M. Ishtaiq, A. Batool, A. Jaffar, M. Mirza, "Feature selection for efficient gender classification", Proceedings of the 11th WSEAS International Conference, pp. 70–75, 2010.

[6] B. Somayeh Mousavi, A. Hirad, "Automatic gender classification using neuro fuzzy system", Indian Journal of Science and Technology, Vol. 4, pp. 1198–1201, 2011.

[7] P. Rai, P. Khanna, "Gender classification using Radon and wavelet transform", Proceedings of the IEEE International Conference on Industrial and Information System, pp. 448–451, 2010.

[8] S. Ravi, S. Wilson, "Face detection with facial features and gender classification based on support vector machine", Proceedings of IEEE International Conference on Computational Intelligence and Computing Research, pp. 125–130, 2010.

[9] A. Shobeirinejad, Y. Gao, "Gender classification using interlaced derivative pattern", Proceedings of IEEE international Conference on Pattern Recognition, pp. 1509–1512, 2010.

[10] Z. Xu, L. Lu, P. Shi, "A hybrid approach to gender classification from face images", Proceedings of IEEE International Conference on Pattern Recognition, pp. 1–4, 2008.

[11] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511–518, 2001.

[12] Y. Su, S. Shan, X. Chen, W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition", Proceedings of International Conference on Computer Vision, pp. 1–8, 2007.

[13] E. Brown, D. Perrett, "What gives a face its gender?", Perception, Vol. 22, pp. 829–840, 1993.

[14] T. Cootes, C. Taylor, D. Cooper, J. Graham, "Active shape models-their training and application", Computer Vision and Image Understanding, Vol. 61, pp. 38–59, 1995.

[15] X. Lian, B. Lu, "Gender classification by combining facial and hair information", Proceedings of Advances in Neuro-Information Processing, pp. 647–654, 2009.

[16] J. Kennedy, R. Eberhart, "Particle swarm optimization", Proceedings of IEEE International conference on neural networks, pp. 1942–1948, 1995.

[17] J. Kennedy, R.C. Eberhart, "A discrete binary version of the particle swarm algorithm", Proceedings of the IEEE International Conference on System, Man and Cybernetics, pp. 4104–4108, 1997.

[18] D.T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, M. Zaidi, "The bees algorithm – a novel tool for complex optimization problems", Proceedings of Virtual International Conference on Intelligent Production Machines and Systems, pp. 254–461, 2006.

[19] T.Z. Zhao, Y.H. Li, J.J. Lu, Y.F. Zhang, "Classifier ensemble based on bi-coded chromosome genetic algorithm for automatic image annotation", Proceedings of the 7th Conference on Machine Learning and Cybernetics, pp. 12–15, 2008.

[20] B. Gabrysa, D. Ruta, "Genetic algorithms in classifier fusion", Applied Soft Computing, Application of Soft Computing in Information & Communication Technology, Vol. 6, pp. 337–347, 2006.

[21] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, New York, Wiley, 2004.

[22] Z. Sun, G. Bebis, X. Yuan, S.J. Louis, "Genetic feature subset selection for gender classification: a comparison study", Proceedings of the IEEE Workshop on Applications of Computer Vision, pp. 165–170, 2002.

[23] B. Gabrysa, D. Ruta, "Genetic algorithms in classifier fusion", Applied Soft Computing, Application of Soft Computing in Information & Communication Technology, Vol. 6, pp. 337–347, 2006.

[24] P. Phillips, H. Wechsler, J. Huang, P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms", Image and Vision Computing, Vol. 16, pp. 295–306, 1998.

[25] S. Ali Khan, M. Nazir, U. Asghar, "Efficient gender classification using interlaced derivative pattern (IDP) and principal component analysis", Proceedings of the 9th International Conference on Frontiers of Information Technology, pp. 270–274, 2011.

[26] C. Shan, "Learning local binary pattern for gender classification on real-world face images", Pattern Recognition Letter, Vol. 33, pp. 421–437, 2011.