

A New Application for Gabor Filters in Face-Based Gender Classification

Ebrahim Al-Wajih and Moataz Ahmed

Information and Computer Science Department, King Fahd University of Petroleum and Minerals, KSA

Abstract: Human face is one of the most important biometrics as it contains information such as gender, race, and age. Identifying the gender based on human face images is a challenging problem that has been extensively studied due to its various relevant applications. Several approaches were used to address this problem by specifying suitable features. In this study, we present an extension of feature extraction technique based on statistical aggregation and Gabor filters. We extract statistical features from the image of a face after applying Gabor filters; subsequently, we use seven classifiers to investigate the performance of the selected features. Experiments show that the accuracy achieved using the proposed features is comparable to accuracies reported in recent studies. We used seven classifiers to investigate the performance of our proposed features. Experiments reveal that *k*-Nearest Neighbors algorithm (*k*-NN), *K*-Star classifier (*K**), and Rotation Forest offer the best accuracies.

Keywords: Gabor filters, gender recognition, statistical features, PCA.

Received September 25, 2017; accepted May 3, 2018

<https://doi.org/10.34028/iajit/17/2/5>

1. Introduction

Gender classification is a research topic with high application potential in areas such as observation, face recognition, witness face reconstruction, image database investigations, and dynamic marketing surveys [35]. Unlike other information inferred from face images, the gender information can be astutely employed to reduce the processing and response times of face-based biometric systems by restricting the face query and matching tasks to a specific gender. Golomb *et al.* [10] did the first study to recognize gender from faces; subsequently, many studies have been proposed to solve this problem. These studies used different approaches as well as different feature extraction techniques. These approaches are classified into three groups: Holistic matching methods, Feature-based matching methods, and hybrid method [17, 40].

These holistic matching methods use the whole face region as the raw input to a recognition system. In this approach, the whole image pixels are used to represent the feature vectors. After that, dimensionality reduction techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) are used to decrease the feature size [18, 24].

In feature-based matching methods, Local features such as nose, eyes, and mouth are extracted and their locations and local statistics are fed into a structural classifier [7, 26]. These methods are based on extracting statistic aggregations from the whole image after filtering the image by Gabor filters. Hybrid methods, just as the human perception system,

use both local features and the whole face region to recognize a face [22, 31].

Gabor filter is a linear filter used for detecting edges. It has been successfully applied to faced-based gender recognition recently [15, 16]. In this paper, we use energies of 24 Gabor filters and extend the set of statistical features reported in previous studies; the extension is mainly meant to overcome the effect of outliers. We applied seven different classifiers to investigate the performance of our proposed features and compared performance.

In section 2 of this paper, we review the prominent studies conducted in gender recognition. Section 3 describes the proposed system including a background on the preprocessing, feature extraction, feature reduction and classifiers used. Section 4 illustrates the setup of our experiments. Results of the experiments are discussed in section 5. Threats to validity and the conclusion along with future work are discussed in sections 6 and 7, respectively.

2. Literature Review

Many studies conducted in the area of gender recognition since the first study was done in 1990 by Golomb *et al.* [10]. In this section, we briefly discuss some prominent studies found in the literature.

Jain *et al.* [18] proposed their technique which extracts ICA features from 500-individuals images to recognize the gender. Support Vector Machine (SVM), LDA, and Cosine Distance Classifier (COS) were used as classifiers. Accuracies reported were 85.33, 93.33, and 95.67% for SVM, COS and LDA, respectively. In the same year, appearance features

and local region-based features were used by Abdelkader and Griffin [2]. These features were extracted using two approaches. The first approach applied PCA on a fixed size canonical region of the face. The second approach applied PCA on N different local regions and then concatenated these region scores to generate a single feature vector. SVM and Fisher Linear+Discriminant (FLD) were used as classifiers. All four possible combinations of classifiers and feature methods were validated and the best accuracy was 94.2% by using SVM classification with the second feature method.

Tivive and Bouzerdoum [37] proposed a technique based on shunting inhibitory convolutional neural networks. Face images of 32x32 pixels from FERET and PHUNG datasets were used to investigate the performance of this technique and 97.1% of accuracy was obtained on average. Toews and Arbel [38] proposed a framework to localize and classify faces from arbitrary viewpoints. Scale-Invariant Feature Transform (SIFT) was used to extract the features and Bayesian method was used as a classifier. In 2010, Discrete Cosine Transform (DCT), and Local Binary Pattern (LBP), and Geometrical Distance Feature (GDF) were used to extract features for gender by Mozaffari *et al.* [29]. These features were validated using Ethnic, AR databases and similarities distance measurement were used as a classifier. A majority rule was used as a last step to classify gender images. This rule allowed the GDF to make decision only and only if the results of DCT and LBP features were not equal.

Recently, Logarithmic Gabor filters combined with LBP features were applied by JafariBarani *et al.* [16] to extract features for gender. Images of 64x64 pixels from AR face database were used and Self-Organizing Map (SOM) was utilized in the classification step. The accuracy of this technique was 90.34%. Moreover, PCA and Discrete Wavelet Transform (DWT) were applied to recognize gender by Biswas and Sil [4]. In their study, two types of features were extracted based on fusion of spatial and temporal features. PCA was used to obtain the spatial features while temporal features were obtained by DWT. The best accuracy was 93% when neural networks were used as classifiers. Another work was proposed by Berbar [3]. Three approaches were proposed to extract features. The first approach used DCT, the second approach was by using the Gray-Level Co-Occurrence Matrix (GLCM), and the third one used 2D-wavelet transform. SVM was used to train models with k-fold cross validation. The best accuracy was given using SVM+DCT.

Further, Jia *et al.* [19] Studied the difference between Convolutional Neural Networks (CNNs) and a support vector machine approach. LBP was used as a feature extraction technique on Labelled Faces in the Wild (LFW) dataset. Mansanet *et al.* [27] suggested an approach to recognize gender using the face image called Local Deep Neural Network (Local-DNN).

Local features were extracted by applying Sobel filters and low-pass filter. After that, the yielded image was converted to binary value and then the centers of all local features were extracted.

Our approach uses statistical features extracted from filter images of an image. The dimensionality of these features is very small comparing to [13, 15, 16]. In addition, the performance of these features are compared with [3] which used the same dataset that are used in this study.

3. Face-based Gender Recognition Model

Our face-based gender recognition approach is based on the generic model of the pattern recognition system [36] illustrated in Figure 1. Our work focuses on the feature extraction and model blocks that are specific for face-based gender recognition. In the sequel, we discuss the generic model within the context of our system setup and experimentation. We also introduce the classifiers we used in our recognition system.

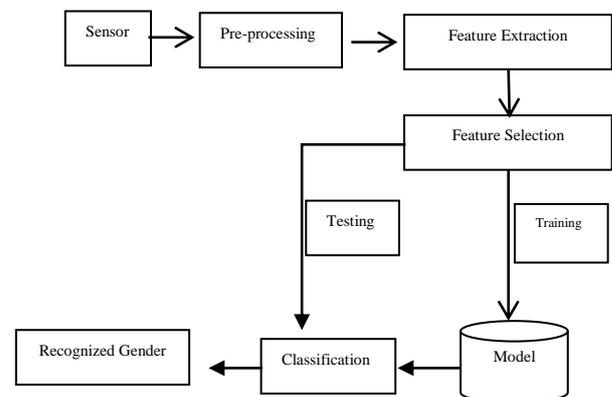


Figure 1. A generic model for gender recognition system.

3.1. Preprocessing

A preprocessing activity is always conducted to prepare the images for further processing. This activity typically includes converting colour images to grayscale images; detecting the face region from the image and remove other objects (e.g., hair) and background from the image; and resizing the image to an appropriate size for feature extraction. In our work, we resize the image to 28x28 pixels.

3.2. Feature Extraction

In real life, each object has its attributes that distinguish it from other objects. These attributes are called features. Several techniques Features are extracted from images using as discussed in section 2.

3.2.1. Gabor Filter

Gabor filter is a linear filter used for detecting the edge. Frequency and orientation representations of Gabor filters are like those of the human visual system, and they have been found to be particularly

appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor kernel is the product of a Gaussian and a cosine or a sine plane wave. The 2D Gabor filter is mathematically formulated as [12]:

$$G_{\gamma,\sigma,\theta,\phi}(x,y) = \frac{f^2}{\pi\gamma\eta} \exp\left(-\frac{(\alpha\bar{x})^2 + (\beta\bar{y})^2}{2\sigma^2}\right) \cdot \exp(j2\pi f\bar{x} + \phi)$$

$$\bar{x} = x \cdot \cos\theta + y \cdot \sin\theta \quad (1)$$

$$\bar{y} = -x \cdot \sin\theta + y \cdot \cos\theta$$

Where x and y represent the values of a filter of size $R \times C$, f is the frequency of the sinusoidal factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ϕ is the phase offset, σ is the standard deviation of the Gaussian envelope and γ is the spatial aspect ratio.

After preprocessing image, Gabor filters are applied to a face image with scales $u=4$, and orientations $v=6$, and the size of filter is 13×13 pixels. The experiments are run by re-implementing Gabor filters several times by optimizing the parameters and the best accuracies are obtained with the following parameters:

$f = 0.25/(\sqrt{2})^{i-1}$, where $i = 1, \dots, u$, $\gamma = \sqrt{2}$, $\theta = m\pi/6$, where $m = 1, \dots, v$, $\phi = 0$, $\eta = \sqrt{2}$, $\sigma = 1/\sqrt{2}$, $\alpha = f/\gamma$, $\beta = f/\eta$. The Twenty-four filtered images are generated from each image. These filtered images are used to extract the features that used in our approach. Figure 2 shows the real parts of Gabor-filtered images.

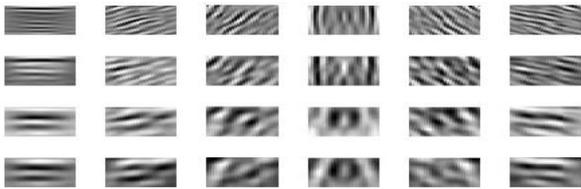


Figure 2. The real parts of gabor-filtered images.

3.2.2. Statistical Features

Eight different features from each filtered images are extracted. Two of such features the Arithmetic Mean (AM), standard deviation (s)) were used in [15], and [30]. In our research, we added the six features: median, Geometric Mean (GM), Harmonic Mean (HM), Mean Absolute Deviation (MAD), skewness (skew), and Kurtosis (Kur) features.

$$AM = \frac{\sum_{i=1}^n x_i}{n} \quad (2)$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3)$$

Median:

$$\begin{aligned} & \text{-Sort the } m \text{ by } m \text{ pixels values.} \\ & \text{If } n \text{ is odd, Median} = \text{middle value} \\ & \text{Else, Median} = \text{mean of two middle values.} \end{aligned} \quad (4)$$

$$GM = \left[\prod_{i=1}^n x_i \right]^{\frac{1}{n}} \quad (5)$$

$$HM = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (6)$$

$$mad = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (7)$$

$$skew = \frac{1}{ns^3} \sum_{i=1}^n (x_i - \bar{x})^3 \quad (8)$$

$$kur = \frac{1}{ns^4} \sum_{i=1}^n (x_i - \bar{x})^4 - 3 \quad (9)$$

Where m is the width and height of the image, $n = m * m$, and x_i is the intensity value of each pixel in image.

The rationale behind extending the features is that the mean and standard deviation, used in [15, 30], are affected by outliers. In images, the outliers occur due to the real world illumination changes and variations in direction of the light source which causes a misclassification issue. For mean attribute, median is another central tendency measure that is less affected by outliers. For variance attribute, large outliers will create a higher dispersion that can be reduced by using mean absolute deviation. Gabor filters are convoluted with the signal, using Equation (1) that is based on several parameters, resulting in a so-called Gabor spaces. Hence, due to the complicated equation, using harmonic and geometric mean as a feature is reasonable to represent these spaces. Furthermore, skewness, and Kurtosis are used to support the proposed feature with information of the lack of symmetry, and the lack or the abundance of outliers, respectively.

3.3. Feature Selection/Reduction

The objective of this step of feature reduction is to reduce the features size used in training to build a model as well as the testing phase. PCA is one of the most popular statistical procedures used to reduce the number of features [20]. PCA uses an orthogonal transformation to transform correlated variables into uncorrelated variables. Applying PCA on original features of a dataset may allow representing the dataset variables in new component variables [41]. In PCA, the variance of a linear combination of the variables is maximized. The largest variance is included in the first component, the second largest in second components and so on. Generally, in the real problem, the number of components is specified by trying many times and get the best components that give the best result.

The number of features considered in our study is 192. We used PCA to reduce the feature size and the corresponding computational complexity of the machine learning algorithms. The number of Principal

Components (PCs) has been selected based on a cumulative percentage of total variation of variances [21] that is calculated by:

$$C = \frac{\sum_{i=1}^m \lambda^i}{\sum_{i=1}^n \lambda^i} \times 100 \quad (10)$$

Where m is the number of eigenvalues selected, n is the total number of eigenvalues, λ^i is the i^{th} eigenvalue and C is the threshold set which gives cumulative percentage explained by number of selected PCs that is determined by 90%.

3.4. Classifiers Techniques

The extracted features are used to distinguish between the patterns by a classifier. The classifier takes the features as an input and maps them to some discrete label. We have used seven classifiers to investigate the performance of our proposed features. In section, we described these classifiers briefly.

- **Multilayer Perceptron classifier (MLP):** One of the most common classifiers for image classification problem is the MLP built using back-propagation algorithm [34]. The input layer of the network receives the features or the variables extracted from training data. The input of hidden layers and output layer is the weighted sum of the outputs from the previous layer which is calculated using:

$$net_j = \sum_i w_{ji} o_j \quad (11)$$

The “weighted sum:” is then transformed using one of the activation function (such as a sigmoid (12) or hyperbolic tangent (13)) to obtain the output node:

$$o_j = \frac{1}{1 + \exp(-net_j + \theta_j)} \quad (12)$$

$$o_j = m \tanh(k(net_j)) \quad (13)$$

Where θ_j , m , and k are constants.

The weights of each layer are updated during training the model with the generalized delta rule as:

$$\Delta w_{ji}(n+1) = \eta(\delta_j o_j) + \alpha \Delta w_{ji}(n) \quad (14)$$

Where $\Delta w_{ji}(n+1)$ is the change of a weight connecting nodes i and j in two successive layers, at the $(n+1)^{\text{th}}$ iteration, δ_j is the rate of change of error regarding the output from node j , η is the learning rate, and a a momentum term. More details of these networks can be found in [1]. In this work, 30 nodes are used in the hidden layer.

- **Support Vector Machine classifier (SVM):** SVM is a classification approach that is used to classify linear or nonlinear data. The first work using SVM was

proposed by [5]. The idea of using SVM is based on statistical learning theory [39]. In general, the idea of this classifier is by separating data set of two classes with a maximum distance between them. The performance of SVM classifier differs based on a kernel function and the data of the problem itself. The data that are used to build a model can be linearly separable or non-linear separable. The best kernel function is used with SVM based on the nature of data, separable or not separable. There are many kernel function used with SVM classifier such as Linear, Polynomial, Sigmoid and Radial Basis Function kernels [14]. In this study, we used linear kernel function.

- **K-Star classifier (K*):** K* is one of the lazy type classifier which is an instance-based classifier [35]. The validation phase of using this classifier works based on the training data such as k-nearest neighbors classifier. However, an entropy -based distance function is used as a similarity function. K-star classify an instance by comparing this instance to database of instances that are processed in the training phase.
- **Rotation Forest classifier:** Rotation forest is one of the multiple classifiers that are proposed in machine learning area [33]. Taking decision based on multiple classifiers can produce more accurate performance than one classifier. This classifier works as following,

Let $X = \{(x_1, \dots, x_N)\}$ be the training sample set, $Y = \{(y_1, \dots, y_N)\}$ be the corresponding labels, $\omega = \{(\omega_1, \dots, \omega_c)\}$ be the set of class labels from which Y takes values, and $F = \{(f_1, \dots, f_n)\}$ be the feature set, resulting that X is an $N \times n$ matrix. Assume that the Rotation Forest contains L decision trees denoted by D_1, \dots, D_L . The training set for an individual classifier D_i is processed using the following steps:

- **Step 1:** Divide the feature set F into K disjointed subsets randomly. Assuming that each feature subset contains $M=n/K$ features.
- **Step 2:** Let F_{ij} be the j^{th} subset of features for training classifier D_i , and X_{ij} be the dataset for the features in F_{ij} . For each subset, a nonempty subset of classes is selected from X_{ij} randomly. Then a bootstrap subset of samples is built using 75% of the dataset to create a new training set, which is denoted by X_{ij} . After that, a linear transformation is applied to X_{ij} to produce its coefficient a_{ij} in a matrix c_{ij} .
- **Step 3:** create a sporadic rotation matrix R_i with the obtained coefficients in matrix c_{ij} , as follows:

$$R_i = \begin{bmatrix} a_{i1}^1, \dots, a_{i1}^{M_1} & \{0\} & \dots & \{0\} \\ \{0\} & a_{i2}^1, \dots, a_{i2}^{M_2} & \dots & \{0\} \\ \{0\} & \{0\} & \dots & a_{iK}^1, \dots, a_{iK}^{M_K} \end{bmatrix} \quad (15)$$

The columns of R_i is reordered based on the original feature sequence, and the reordered rotation matrix is denoted as R_i^a . Then the transformed training set for classifier D_i is XR_i^a . In the validation phase, given a test sample, let d_{ij} be the probability obtained using the classifier D_i to the assumption that x belongs to class ω_i . Then the label of x is computed by the average combination method:

$$\mu_i = \frac{1}{L} \sum_{i=1}^L d_{ij}(xR_i^a), j = 1, \dots, c \quad (16)$$

and x is assigned to the class that has largest average.

- **DECORATE Classifier:** Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples (DECORATAE) is another type of the ensemble classifiers [28]. This classifier is developed based on the measure of disagreement called “diversity”. The diversity of i^{th} classifier d_i in classification problem can be calculated as:

$$d_i(x) = \begin{cases} 0, & \text{if } C_i(x) = C^*(x) \\ 1, & \text{otherwise} \end{cases} \quad (17)$$

Where $C_i(x)$ is the prediction of i^{th} classifier, $C^*(x)$ is the prediction of the ensemble. Krogh and Vedelsby [23] generalize error, E , of the ensemble classifier as $E = \bar{E} - \bar{D}$, where \bar{D} is the diversity of the ensemble and \bar{E} is the mean error. The Algorithm of DECORATE classifier was defined in [28].

- **Logistic Model Tree (LMT) Classifier:** LMT is a type of tree classifier that is built based on the logistic regression [25]. The logistic regression model is implemented at the leaf nodes of a tree. As it is known that the logistic regression is constructed using the logistic function and its inverse function formulated as Equations (18) and (19), respectively.

$$f(x) = \frac{1}{1 + \exp(-\beta X)} \quad (18)$$

$$g(X) = \ln \frac{f(x)}{1-f(x)} \quad (19)$$

$$= \beta X \quad (20)$$

$$= \beta_0 + \beta_1 x_1 + \dots + \beta_N x_N \quad (21)$$

Where $X=(1, x_1, \dots, x_N)$ are the dependent variables, $\beta=(\beta_0, \beta_1, \dots, \beta_N)$ are the regression coefficients and N is the number of dependent variables.

- **k-Nearest Neighbour (k-NN):** k-NN is a simple classification approach used commonly in pattern recognition field. k-NN was firstly proposed by Fix and Hodges [9] and modified by Cover and Hart [6].

In this method, the similarity distance between a test sample and each training test samples are computed, and the class label of the sample that produces the smallest distance (the nearest sample) is considered as the class label for the test sample. There are many similarity distance measures, but only one of them has been applied in this study namely Euclidean distance [11] Equation as:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (22)$$

where a_r is the value of the r^{th} feature on sample x .

However, instead of using only the class label of the closest point, k closest points are considered. In this case, k closest points to the test sample are selected and then the most frequent class among them is assigned to test instance. In this paper, k of 3 has been considered after optimizing it by exploring $k=1, 3, 5, 7$.

4. Experiment Design

We conducted some experiments to investigate the performance of our proposed method. In the sequel, we discuss the dataset we used as well as the metrics we employed to evaluate the performance.

4.1. Datasets

We used three different popular databases in our experiments: Faces 94, faces95 and face96, that are available at <http://cswww.essex.ac.uk/mv/allfaces/>. Table 1 describes the characteristics of each dataset. For the sake of validation, we compared our results against Berbar [3] results. Berbar used AT@T, Faces94, UMIST, and colour FERET in his study. Table 1 describes the characteristics of those additional databases we had to use for the sake of comparison.

The first three datasets, Faces94, Faces95, and Faces96 have been used in our experiments in different ways. First, each dataset has been used individually. Then two and three combinations of these datasets have been merged to be one dataset. There are 3 single and 4 combination datasets. The first combination contains (Faces94+Faces95) datasets, the second contains (Faces94+Faces96) datasets, the third contains (Faces95+Faces96) datasets, and the fourth contains (Faces94+Faces95+Faces96) datasets. We denote each combination of (Faces94+Faces95), (Faces94+Faces96), (Faces95+Faces96), and (Faces94+Faces95+Faces96) by “Comb1”, “Comb2”, “Comb3”, and “Comb4” respectively.

Table 1. Databases.

Characteristic Database	# individuals	# Female	# Male	Image resolution:	# images per individual
Faces94	133	20	113	180x200	20
Face95	72	14	58	180x200	20
Face96	151	23	128	196x196	20
AT@T	40	5	35	92x112	10
UMIST	20	4	16	220 x 220	> 24
FERET	644	104	540		

4.2. Performance Metrics

To evaluate the performance of our feature set, we used several measures from Fawcett [8]. These measures are classification accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and confusion matrix. A confusion matrix includes information about actual and predicted classifications done by a classification system is also used. Table 2 shows the confusion matrix of a binary classifier.

Table 2. The confusion matrix of a binary classifier.

Actual	Predicted	
	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

By using the elements of the confusion matrix, another measurement can be calculated.

$$\text{True positive rate (TPR)} = \frac{TP}{TP + FN} \quad (23)$$

$$\text{False positive rate (FPR)} = \frac{FP}{FP + TN} \quad (24)$$

The accuracy (AC) is the proportion of the total number of predictions that are correct.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (25)$$

We also use the Receiver Operating Characteristic (ROC), or ROC curve, graphs. ROC is two-dimensional graphs in which *TPR* is plotted on the Y axis and *FPR* is plotted on the X axis. A ROC graph shows the relative trade-offs between benefits (true positives) and costs (false positives).

5. Experiments and Results

In this study, seven classifiers are applied on different databases to show the performance of the proposed features. Table 3 shows the accuracies of the first experiment that used 3 datasets and their 4 combinations. These accuracies are calculated by applying 10-fold cross validation. The value in each cell is the mean accuracy of the 10-fold \pm the confidence interval with respect to 0.95 confidence level. In this work, all 7 classifiers are applied using Weka 3.6 framework. The default values of each

classifier's parameters are used except MLP and k-NN classifiers. In MLP, the number of hidden layers is 30 and the k is equal 3 in k-NN (3-NN). As we mentioned before, all images in all datasets are resized to the same size (28x28 pixels) and the number of features is 192 features. The results of each type of combination are discussed separately.

5.1. Results of Single Datasets

Three single datasets (Faces94, Faces95, and Faces96) are applied in this study. These datasets contain 2631, 946 and 1803 images, respectively. The number of male images of each dataset is 2231, 701, and 1573 respectively and the number of female images is 400, 245, and 230 respectively. The highest accuracies of using Faces94, Faces95, and Faces96 datasets are 99.96 ± 0.09 , 98.73 ± 1.17 and 99.83 ± 0.19 respectively. These accuracies are obtained using 3-NN and K* classifier on Faces94 dataset, Rotation Forest on Faces95 dataset, and 3-NN with Faces96 dataset. However, the difference between the accuracies of the classifiers with the same dataset is too small. So, it can be concluded that the good results are due to the significant of the proposed features to determining gender.

5.2. Results of Comb1, Comb2 and Comb3 Datasets

In this section, three combination datasets are used. The two-combination datasets contain 3577, 4434, and 2549 for Comb1, Comb2, and Comb3 respectively. The number of male images is 2932, 3804, and 2274, and the number of female is 645, 630, and 475 respectively. The highest obtained accuracy using these combinations is 99.64 ± 0.28 , 99.89 ± 0.11 , and 99.42 ± 0.33 for Comb1, Comb2, and Comb3 respectively. These results are obtained using 3-NN classifier on Comb1 and Comb3 while K* is used on Comb2. However, the difference between the accuracies of the classifiers with Comb1 and Comb2 is smaller than Comb3. This is because the variation of illumination of Faces94 and Faces96 datasets is the smallest, and the variation of the illumination of Faces95 and Faces96 is the most. The difference may be reduced if an illumination normalization technique is applied to enhance the image.

5.3. Results of Comb4 Dataset

The last combination Comb4 contains 5380 images. The number of male images is 4505, and the number of female images is 875. The highest accuracy obtained using this combination is 99.61 ± 0.23 . This accuracy is obtained using K* classifier. In this combination, the difference between the accuracies of the classifiers with Comb4 is the highest comparing to all other combinations. This informs that the variation

of the illumination of Faces94, Faces95 and Faces96 datasets is the highest.

In addition, we use the Area Under the Curve (AUC) performance metric here to show that the performance of our model is not by chance. The area under the curve of each classifier with each dataset are illustrated in Figures 3, 4, and 5, for 10-fold, 5-fold, and 2-fold cross

validation, respectively. From these figures, it is noticed that all models built using the 7 classifiers in 10-fold, 5-fold, and 2-fold cross validation produce AUC between 90% and 100%. This implies that if the number of images in each class, male or female, is not balanced, the performance may not be effected.

Table 3. The accuracy of 10-fold cross validation.

Classifier	Faces94	Faces95	Faces96	Comb1	Comb2	Comb3	Comb4
MLP	99.92±0.12	97.89± 1	99.67±0.33	98.88±0.52	99.66±0.26	99.05±0.52	99.39±0.43
SVM	99.89±0.18	98.41±1.02	99.11±0.28	97.68±0.73	98.74±0.28	96.36±0.96	96.08±0.56
K-NN(k=3)	99.96±0.09	98.63±1.12	99.83±0.19	99.64±0.28	99.77 ± 0.19	99.42±0.33	99.57±0.24
K*	99.96±0.09	98.42± 1.02	99.72±0.28	99.55±0.19	99.89±0.11	99.09± 0.45	99.61±0.23
Decorate	99.7±0.28	97.89 ± 1.18	98.84±0.80	99.05±0.35	99.32±0.36	98.22±0.58	98.75±0.25
Rotation Forest	99.58±0.24	98.73± 1.17	99.45±0.26	99.16±0.23	99.17±0.20	98.58±0.67	98.92±0.39
LMT	99.39±0.23	97.89± 1.12	99.28±0.35	98.66±0.32	98.78±0.19	98.14±0.50	98.27±0.47
Mean	99.77	98.27	99.41	98.95	99.33	98.41	98.66
Variance	0.05	0.14	0.13	0.43	0.22	1.04	1.52

5.4. Comparison with Previous Studies

We conducted experiments using the datasets AT@T It is the another name of AT@T dataset, (ORL) and UMIST in addition to Faces94 to compare with Berbar work [3]. Table 4 displays the number of images that were used in [3] compared to the number of images used in this study. We did not use all images because the face detection algorithms could not detect the face region from the original image. Moreover, we could not apply the proposed approach on FERET dataset because we do not have access to this dataset.

Table 5 displays the comparison of the average of the accuracies of 2-, 5-, and 10-fold cross validation between our approach and Berbar approach. SVM classifier was used in Berbar study while 3-NN classifiers is used in this work. Table 5 shows the comparison between our work and Berbar work before and after reducing the number of features. From Table 5, it can be concluded the following; before reducing the feature size, our approach produces better performance in all experiments except when using ORL dataset in 2-fold, and Faces94 dataset in 5-fold. When the feature size is reduced, our approach produces better performance when we used Faces94 dataset in 2-fold and 10-fold. However, Berbar approach produced better than ours when ORL and Faces94 datasets are used in 2-fold and 5-fold, respectively. Moreover, the two approaches produced the same performance when UMIST dataset is used. Table 5 shows that our proposed features have produced a better accuracy with respect to the feature size.

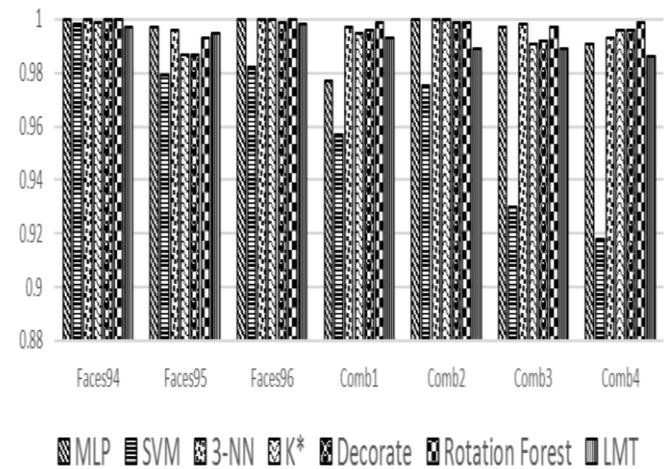


Figure 3. The area under the curve using 10-fold cross validation.

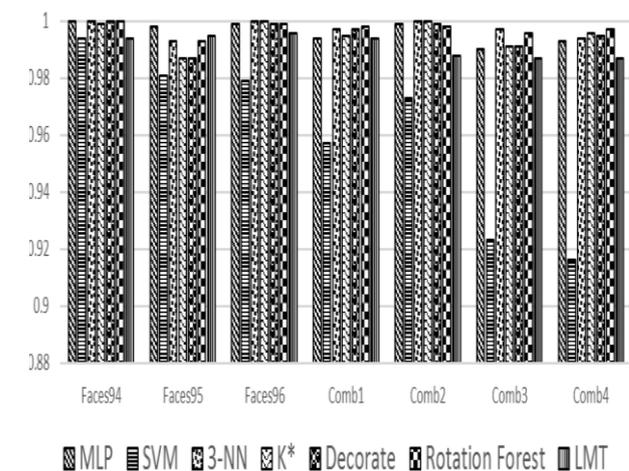


Figure 4. The area under the curve using 5-fold cross validation.

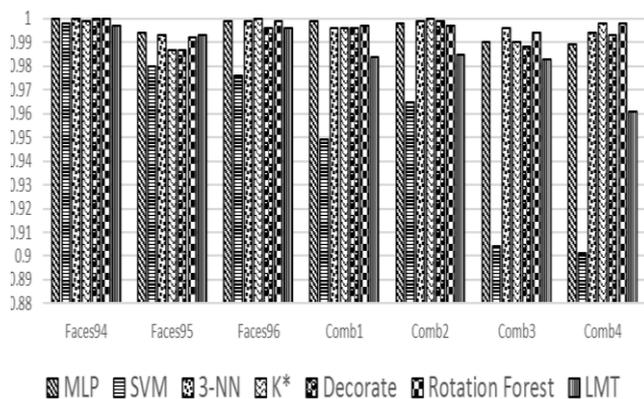


Figure 5. The area under the curve using 2-fold cross validation.

Table 4. The number of images used in Berbar study and our study.

Studies	Gender	ORL	UMIST	Faces94
Berbar	Male	350	840	2660
	Female	50	172	399
Our work	Male	350	840	2631
	Female	50	172	400

6. Threat to Validity

- *External validity:* In our experiment, we only used 5 small datasets which represent only few humans. Accordingly, we cannot generalize our conclusions from the experiment because the characteristics of human faces differ based on races, regions, etc. In addition, the tool we used to detect and extract the face from original images produced few mistakes with some images. It extracted few incorrect faces from original images.
- *Internal validity:* The face databases that we used have similar characteristics regarding: head scale, image lighting variation, expression variation, turn, tilt and slant, translation and glasses/no glasses. However, other datasets may have different noise such as wearing scarf or some incomplete images. So, if the input features are changed, the results may be changed.

Table 5. Comparing our work with Berbar work.

Study	Classifiers	# features	Database	2-fold	5-fold	10-fold
Berbar	SVM	64	ORL	98.6	98.93	98.5
			UMIST	99.90	99.90	99.90
			Faces94	99.93	100	-
Our proposed	3-NN	192	ORL	96.75	99.25	99.25
			UMIST	100	100	100
			Faces94	99.96	99.93	99.96
		30	ORL	96.5	99.5	99.5
		23	UMIST	99.90	99.90	99.90
		23	Faces94	99.96	99.92	99.96

7. Conclusions Future Work

In this paper, a modified feature extraction technique is proposed based on the statistical aggregation. These features are extracted from energies generated using Gabor filters. The performance of these features is investigated using 5 face datasets and 7 classifiers. These features provide significant performance to recognize gender from facial images with accuracy between 96% and 100%. The best performance is obtained by using 3-NN, K* and Rotation Forest classifiers. In addition, when the datasets are combined with more than one dataset, the performance is almost not affected.

Our future work will consider using illumination normalization technique before applying our proposed work. We will also consider using another larger dataset for training and testing such as GBU [32] dataset.

Acknowledgment

The authors wish to acknowledge King Fahd University of Petroleum and Minerals (KFUPM) for utilizing the various facilities in carrying out this research.

References

- [1] Atkinson P. and Tatnall A., "Introduction Neural Networks in Remote Sensing," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 699-709, 1997.
- [2] Abdelkader C. and Griffin P., "A Local Region-Based Approach to Gender Classification from Face Images," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, pp. 52-52, 2005.
- [3] Berbar M., "Three Robust Features Extraction Approaches for Facial Gender Classification," *The Visual Computer: International Journal of Computer Graphics*, vol. 30, no. 1, pp. 19-31, 2014.
- [4] Biswas S. and Sil J., "Advanced Computing, Networking and Informatics-Volume 1," Springer International Publishing, 2014.
- [5] Boser B., Guyon I., and Vapnik V., "A Training Algorithm for Optimal Margin Classifiers," in *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, Pittsburgh, pp. 144-152, 1992.
- [6] Cover T. and Hart P., "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.
- [7] Cox I., Ghosn J., and Yianilos P., "Feature-Based Face Recognition Using Mixture-Distance," in *Proceedings CVPR IEEE*

- Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, pp. 209-216, 1996.
- [8] Fawcett T., "An Introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006.
- [9] Fix E. and Hodges J., "Discriminatory Analysis-Nonparametric Discrimination: Consistency Properties," *International Statistical Review/Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238-247, 1951.
- [10] Golomb B., Lawrence D., and Sejnowski T., "SEXNET: A Neural Network Identifies Sex from Human Faces," in *Proceedings of the 3rd International Conference on Neural Information Processing Systems*, Denver, pp. 572-577, 1990.
- [11] Gower J., "Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis," *Biometrika*, vol. 53, no. 3-4, pp. 325-338, 1966.
- [12] Haghghat M., Zonouz S., Abdel-Mottaleb M., "Identification Using Encrypted Biometrics," in *Proceedings of International Conference on Computer Analysis of Images and Patterns*, Berlin, pp. 440-448, 2013.
- [13] Haider K., Nawaz T., Habib H., Maqsood M., and Amin T., "Gender Classification of Consumer Face Images Using Gabor Filters," *International Journal of Computer Science and Network Security*, vol. 16, no. 2, pp. 46-53, 2016.
- [14] Hsu C., Chang C., and Lin C., "A Practical Guide to Support Vector Classification," Technical Report, National Taiwan University, 2003.
- [15] Ignat A. and Coman M., "Gender Recognition with Gabor Filters," in *Proceedings of E-Health and Bioengineering Conference*, Iasi, pp. 1-4, 2015.
- [16] JafariBarani M., Faez K., and Jalili F., "Implementation of Gabor Filters Combined with Binary Features for Gender Recognition," *International Journal of Electrical and Computer Engineering*, vol. 4, no. 1, pp. 108-115, 2014.
- [17] Jafri R. and Arabnia H., "A Survey of Face Recognition Techniques," *Journal of Information Processing Systems*, vol. 5, no. 2, pp. 41-68, 2009.
- [18] Jain A., Huang J., and Fang S., "Gender Identification Using Frontal Facial Images," in *Proceedings of IEEE International Conference on Multimedia and Expo*, Amsterdam, pp. 1-4, 2005.
- [19] Jia S., Lansdall-Welfare T., and Cristianini N., "Gender Classification by Deep Learning on Millions of Weakly Labelled Images," in *Proceedings of IEEE 16th International Conference on Data Mining Workshops*, Barcelona, pp. 462-467, 2016.
- [20] Jolliffe I., *Principal Component Analysis*, Wiley Online Library, 2002.
- [21] Kaymak Ç., Sarıcı R., and Uçar A., "Illumination Invariant Face Recognition Using Principal Component Analysis-An Overview," *Machine Vision and Mechatronics in Practice*, Berlin, pp. 269-285, 2015.
- [22] Khan S., Nazir M., Riaz N., and Khan M., "Optimized Features Selection Using Hybrid PSO-GA for Multi-View Gender Classification," *The International Arab Journal of Information Technology*, vol. 12, no. 2, pp. 183-189, 2015.
- [23] Krogh A. and Vedelsby J., "Neural Network Ensembles, Cross Validation, and Active Learning," *Advances in Neural Information Processing Systems*, vol. 7, pp. 231-238, 1995.
- [24] Kumari S. and Majhi B., "Classifying Gender from Faces Using Independent Components," in *Proceedings of International Conference on Computational Science, Engineering and Information Technology*, Berlin, pp. 589-598, 2011.
- [25] Landwehr N., Hall M., and Frank E., "Logistic Model Trees," *Machine Learning*, vol. 59, no. 1-2, pp. 161-205, 2005.
- [26] Manjunath B., Chellappa R., and Malsburg C., "A Feature Based Approach to Face Recognition," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, pp. 373-378, 1992.
- [27] Mansanet J., Albiol A., and Paredes R., "Local Deep Neural Networks for Gender Recognition," *Pattern Recognition Letters*, vol. 70, pp. 80-86, 2016.
- [28] Melville P. and Mooney R., "Constructing Diverse Classifier Ensembles Using Artificial Training Examples," in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, Acapulco, pp. 505-510, 2003.
- [29] Mozaffari S., Behravan H., and Akbari R., "Gender Classification Using Single Frontal Image Per Person: Combination of Appearance and Geometric Based Features," in *Proceedings of 20th International Conference on Pattern Recognition*, Istanbul, pp. 1192-1195, 2010.
- [30] Mu M. and Ruan Q., "Mean and Standard Deviation as Features for Palmprint Recognition Based on Gabor Filters," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 4, pp. 491-512, 2011.
- [31] Penev P. and Atick J., "Local Feature Analysis: A General Statistical Theory for Object Representation," *Journal Network: Computation in Neural Systems*, vol. 7, no. 3, pp. 477-500, 1996.
- [32] Phillips P., Beveridge J., Draper B., Givens G., O'Toole A., Bolme D., Dunlop J., Lui Y., Sahibzada H., and Weimer S., "The Good, the

- Bad, and the Ugly Face Challenge Problem,” *Image Vision Computing*, vol. 30, no. 3, pp. 177-185, 2012.
- [33] Rodriguez J., Kuncheva L., and Alonso C., “Rotation Forest: A New Classifier Ensemble Method,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1619-1630, 2006.
- [34] Rumelhart D., Hinton G., and Williams R., “Learning Internal Representations by Error Propagation,” Technical Report, California University, 1985.
- [35] Shah D., “The Exploration of Face Recognition Techniques,” *International Journal of Application or Innovation in Engineering and Management*, vol. 3, no. 2, pp. 238-346, 2014.
- [36] Theodoridis S. and Koutroumbas K., *Pattern Recognition*, Academic Press, 2009.
- [37] Tivive F. and Bouzerdoum A., “A Shunting Inhibitory Convolutional Neural Network for Gender Classification,” in *Proceedings of 18th International Conference on Pattern Recognition*, Hong Kong, pp. 421-424, 2006.
- [38] Toews M. and Arbel T., “Detection, Localization, and Sex Classification of Faces from Arbitrary Viewpoints and under Occlusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1567-1581, 2009.
- [39] Vapnik V. and Chervonenkis A., “On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities,” *Theory of Probability and its Applications*, vol. 16, no. 2, pp. 264-280, 1971.
- [40] Zhao W., Chellappa R., Phillips P., and Rosenfeld A., “Face Recognition: A Literature Survey,” *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.
- [41] Zhao W., Krishnaswamy A., Chellappa R., Swets D., and Weng J., “Discriminant Analysis of Principal Components for Face Recognition,” in *Proceedings 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, pp. 73-85, 1998.



Ebrahim Al-Wajih, received his BSc in computer science from Hodeidah University, 2007. He received his MSc in computer science from Hodeidah University and King Fahd University of Petroleum and Minerals, 2016, Saudi Arabia. He is currently a Ph.D. student. His areas of interests are machine learning, image processing, and deep learning.



Moataz Ahmed received his PhD in computer science from George Mason University in 1997. Dr. Ahmed is currently a faculty member with the Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Saudi Arabia. He also serves as Adjunct/Guest Professor in a number of universities in the US and Italy. During his career, he worked as a software architect in several software houses. His research interest includes artificial intelligence and machine learning; and automated software engineering, especially, artificial intelligence based software testing, software reuse, and cost estimation. He has supervised a number of theses and published a number of scientific papers in refereed journals and conferences in these areas.