# Feature Selection for Efficient Gender Classification

M. Nazir, Muhammad Ishtiaq, Anab Batool, M. Arfan Jaffar, Anwar M. Mirza
National University of Computer and Emerging Science, FAST, Islamabad, Pakistan
{ muhammad.nazir, m.ishtiaq, arfan.jaffar, anwar.m.mirza}@nu.edu.pk, kazmi.shah110@gmail.com,

**Abstract**
The study presents an efficient gender classification technique. The gender of a facial image is the most prominent feature, and improvement in the existing gender classification methods will result in the high performance of the face retrieval and classification methods for large repositories. In this paper a new efficient gender classification method is proposed. First, the face part of the image is segmented using Viola and Jones face detection technique which excludes unwanted area from the image, so reducing image size. Histogram equalization is performed to normalize the illumination effect. Discrete Cosine Transform (DCT) is employed for feature extraction and sorting the features with high variance. K-nearest neighbor classifier (KNN) is used for classification. The face images used in this study were obtained from the Stanford university medical student (SUMS) frontal facial images database. The experimental results on the SUMS face database indicate that the proposed approach achieves as high as 99.3% gender classification accuracy.

**Keywords:** Gender Classification, KNN, DCT, Feature selection.

## 1. Introduction

Gender classification problem is an active research area which has attracted a great deal of attention recently. It is a challenging pattern recognition problem. Generally gender classification involves a process of determining the gender of a subject from face images. The face images analysis plays an important role in computer vision. Face images analysis has been successfully used in many applications ranging from biometric to robotic-human interaction. Current face detection applications operate with high accuracy as compared to gender classifications systems, because gender classification systems do not offer same level of performance and accuracy.

Gender classification is one of the most challenging problems in computer vision. Two key components of gender classification are feature extraction and pattern classification. A number of different techniques based on facial images have been reported in the literature for solving this problem. These techniques include geometrical feature based methods [17], graph matching methods [15], support vector machine [14] and neural network based methods [16]. These approaches are classified into feature based and feature and template based.

Abdul Majid et al. [2] utilized the DCT for facial features extraction and K-means, K-Nearest Neighbors (KNN), LDA, Mahalanobis Distance Based (MDB), and modified KNN classifiers impact on four hundred frontal facial images. They have used complete input image and their method is also prone to varying illumination effects of the images.

In this paper the gender classification problem is solved in an efficient manner. Face portion of the image is extracted from the image using Viola & Jones [1], this helps in reducing the complexity of the proposed method. Then Histogram equalization is performed to reduce the effect of varying illumination. Block based 8x8 DCT zigzag features are then extracted and classification is performed using KNN. Zigzag scan on the DCT coefficients in a way sorts them in the order of interest. The method is tested on Stanford university medical student`s frontal face images dataset. The results obtained are better than other methods in practice.

Rest of this paper is structured as follow. In section 2 a brief description of the image database is provided. In section 3 the proposed method is discussed. Section 4 describes our experimentation and results followed by discussion on the results obtained and section 5 concludes the paper.

## 2. Related Work

Reasonable work has been done on gender classification, but search for improved gender classification is still going on [3]. Shakhnarovich et al. [4] combined the cascaded face detector by Viola and Jones [1] with threshold Adaboost (Freund and Schapire, 1997) trained classifiers for gender and ethnicity classification. Baback et al. (2005) [5] applied the SVMs to gender classification with low-resolution thumbnail faces (21-by-12 pixels) processed from 1,755 images. The performance of SVMs (3.4% error) is shown to be superior to traditional pattern classifiers (Linear, Quadratic, Fisher Linear Discriminant, Nearest-Neighbor). It also outperformed techniques such as Radial Basis Function (RBF) classifiers and large ensemble-RBF networks.

Tolba et al [6] has proposed gender classification using two techniques with different neural network classifiers i.e. Learning Vector Quantization (LVQ) and Radial Basis Functions (RBF). The problem with any neural network method is that they are computationally expensive. Zhiguang YANG et al. (2006) [7] illustrated an experimental study on automatic face gender classification that focused on the different texture normalization methods including Delaunay triangulation warping and two kinds of affine mappings to the performances of gender classification by SVM, Linear Discriminant Analysis (LDA) and Real Adaboost. Results concluded the suitability of Delaunay triangulation normalization for Haar-like feature based Adaboost approach and for global features based method such as Fisher Linear Discriminant (FLD) or SVM.

Ziyi Xu et al. (2008) [8], presented a novel hybrid face coding method for facial feature extraction. They have fused appearance features extracted by AdaBoost algorithm, and geometry features extracted by Active Appearance Model (AAM) from the normalized image to form a feature vector and passed them to SVM classifier for classification.

A. Majid et.al [11] uses SVM classifiers with different kernel functions for training and combination is performed using Genetic Programming (GP). They proved there Optimal Composite Classifier outperform individual SVM classifiers. They claim that combining the different kernel functions can explore feature space efficiently and can select optimal kernel function for SVM through GP. Maximum performance achieved was 98.8% after 60 generations with 1000 features. They have optimized classifiers input parameters using GP. Genetic algorithm is an efficient evolutionary optimization tool to solve complex problems but no matter how good the classifier is its wrong combinations can lead to undesirable results [12].

N.P.Costen, et.al [13] in their paper uses SVM for feature selection and maximization of classification margin. They have used 300 images of different genders of Japanese individuals with age between 15-60 years and with fixed frontal poses. PCA is used for feature selection which is optimized using SVM. They have compared their results with other existing methods and show comparable performance of their method. The whole approach is dependent on a single regularization parameter θ. Success of this method is highly dependent on the appropriate selection of θ because its incorrect value leads to high error rate.

Sun Z. et.al [14] applied PCA to reduce dimension of image as a feature vector. Genetic algorithm (GA) is used to select subset features from reduced dimension which are true representation of gender. They have compared their results with four different classifiers: linear discriminate analysis (LDA), support vector machines (SVM), neural network (NN), and Bayesian decision making and observed that SVM outperform all others. They achieve the accuracy of 91.1%, 82.3%, 85.8, 77.62 using SVM, NN, LDA and Bayes respectively.

F. Scalzo, et.al [10] proposed a new Feature Fusion Hierarchical (FFH) method for gender classification using Genetic algorithm. FFH model is two levels model. In first level, Gabor and Laplace features are extracted and used as input to feature fusion level. In second level, classifiers fusion uses output of future fusion level to produce a result. This paper has made good contribution by increasing the performance of the current problem at hand. But it increases the performance of individual classifiers rather than improve the overall performance of system.

**Figure.1.** A sample of SUMS face database.

## 3. Proposed Method

Our gender classification method consists of three main modules: face detection, feature extraction/selection, and classification. An input facial image is passed to face detector to extract face from the image, Viola and Jones [1] face detection method is used for this purpose. Then histogram equalization is performed to stretch the contrast of the image, this help overcome illumination variation in the images. In [4] they showed that low-resolution images have equal level of classification accuracy, so we can decrease computational cost by reducing the size of the image. After face detection, the image is resized to 32x32. This resized image is divided into 16 8x8 size blocks. Then each 8x8 block is sorted according to zigzag scan order. These sorted coefficients are arranged in a vector and passed to the KNN classifier. Figure 2 shows the general architecture of the proposed gender classification method and figure 7 depicts step by step flow chart of the system.

### *3.1. Face Detection*

Viola and Jones (2001) [1] in their paper presented a new cascade face detection technique. This is a well known and robust frontal face detection method; its calculation is very fast. This detector extract faces from the image by starting from top left corner and ending at bottom right corner of an image. Three main modules of technique are: First images are represented in the form of "Integral Images", which make feature computation very fast. Second module is using adaboost learning algorithm for feature selection. And the third module is using a cascade of AdaBoost classifiers, to quickly eliminate background regions of the image, while spending more computation on promising object-like regions, speed up the process of detection significantly. In figure 3 face extraction results are given of some selected images from the dataset.

### 3.2. *Feature Extraction*

Ahmed, Natarajan and Rao (1974) first introduced the discrete cosine transform (DCT) in the early seventies. DCT is a well known transformation technique used in image compression applications; Majid et. al. use it for face recognition application [2]. DCT can be used for dimension reduction. DCT coefficients are then sorted according to zigzag scan order, this way we sort the coefficients with decreasing importance, i.e. high variance coefficients are picked first. Like other transforms, the Discrete Cosine Transform (DCT) attempts to decorrelate the image data. After decorrelation each transformed coefficient can be encoded independently without losing compression efficiency.

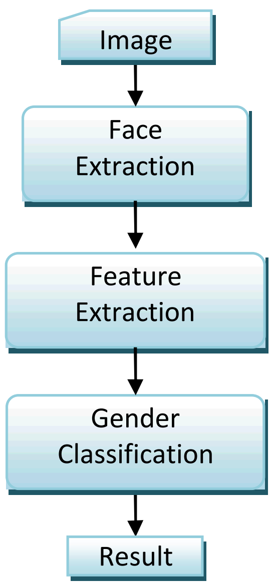


**Figure.2.** Architecture of Gender Classifier

**Figure.3**. A sample of extracted faces from the images.

Given that a gray image is expressed by f(x,y) of size NxN, DCT is defined according to equation 1.

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

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

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

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

The DCT coefficients with high variance are mainly located in the upper-left corner of the DCT matrix. Accordingly, we scan the DCT coefficient matrix in a zigzag manner starting from the upper-left corner and subsequently convert it to a one-dimensional (1-D) vector. This is similar to sorting according to importance. High importance coefficients are located in the top-left corner of the block. When a total of 16 coefficients are selected from an image, only 1st coefficient of each of 16 DCT blocks is selected. As the no. of selected coefficients increases so does the size of the feature vector. For 32 size feature vector first 2 coefficients from each DCT block are selected, and in the same manner 48, 64, 128 and 256 size feature vectors were created.

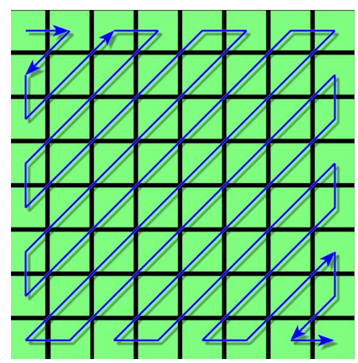

**Figure 4:** Zigzag scan of DCT coefficients

### 3.3. *Classifier*

KNN is a supervised learning classifier. For 1-NN we assign test sample to the class of its closest neighbor, and for KNN we assign the majority class of its K closest neighbors where K parameter is number of neighbors. It is usual to use the Euclidean distance to find closest neighbors, though other distance measures such as the Manhattan distance could in principle be used instead.

## 4. Experimental Results & Discussion

To get the face portion the image is passed to the face detector. Histogram equalization was applied to the extracted face images to normalize for different lighting conditions. Then we resize the image to the size of 32x32 and pass it to feature extractor. Feature extractor first divides the image to the blocks of size 8x8, then picks each block of face image and applies DCT on it. Then coefficients with high variance are selected in a zigzag manner. We continue this process until DCT is applied to all the blocks. In our case we experimented on [256, 128, 64, 48, 32, 16] DCT coefficients. For 256 coefficients we take 16 zigzag coefficients from every block. These selected coefficients are then passed to classifier for classification.

Three types of experiments were performed to see the effect of (i) training set vs test set size, (ii) DCT coefficients normalization effect of classification and (iii) no. of features (DCT coefficients) to be selected for classification. The results for each of these experiments are given in fig. 5 and fig. 6. With our experimentation we concluded that with KNN classifier 50 to 50 training to testing ratio of data set is suitable, it leads to 99.3% of classification accuracy. In table I the proposed methods is compared with [14]. The results obtained better then [14] for all classifiers. In table II the methods is compared with different combination of classifiers. LDA didn't perform well with DCT, But while using Viola and Jones technique with DCT and LDA, drastic change in results were noticed and we can see that as the training ratio increases results also get improved. The proposed method is not affected by the training and testing set size. Also the classification accuracy is not affected even if we select very few of the total DCT features.

**Table I.** Comparison of techniques

| Method | | Accuracy |
|---|---|---|
| Proposed | | **99.30%** |
| Sun et. al [14] | SVM | 91.10% |
| | NN | 82.30% |
| | LDA | 85.80% |
| | Bayes | 77.62% |

**Table II.** Comparison of techniques

| Techniques | Training to Test Ratio | | | |
|---|---|---|---|---|
| | **90 to 10** | **75 to 25** | **25 to 75** | **10 to 90** |
| Proposed | **97.99** | **98.96** | **98.85** | **98.9** |
| DCT , Mahalanobis | 97.47 | 94.76 | 95.8 | 95.87 |
| Viola & Jones, DCT,LDA | 97.5 | 98 | 91.33 | 88.12 |
| DCT , LDA | 72.97 | 69.62 | 67.63 | 68.75 |
| DCT , Kmeans | 54.67 | 61.32 | 61.98 | 58.81 |

## 5. Conclusion

In this paper, problem of gender classification has been investigated. It has been observed that the use of selected face portion in combination with DCT features and KNN The proposed method is robust to varying illumination effects and un-even size images. The method is very efficient because high classification accuracies can be obtained with very few DCT features.

The experimentation was carried out on the Stanford university medical student (SUMS) frontal facial images database. The performance of the proposed method is tested against different selected training to test set sizes and the results obtained are better than other methods.

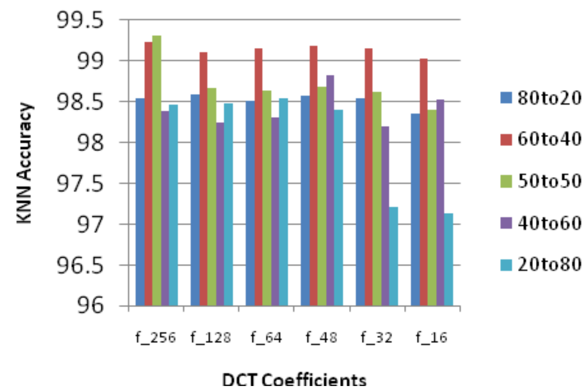


**Figure.5**. DCT Coefficients with KNN classifier

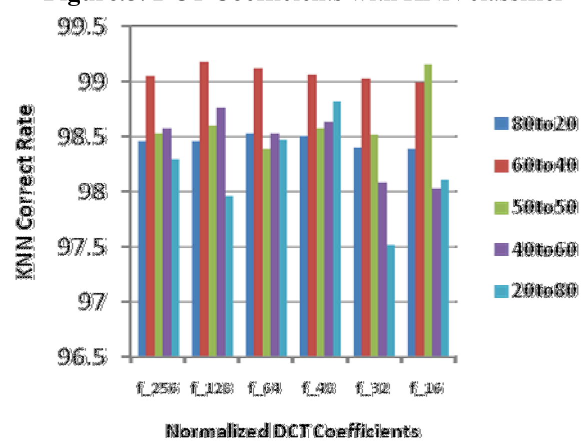


**Figure.6**. Normalized DCT Coefficients with KNN classifier

## References

[1] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01), p. 511, Volume 1, 2001.

[2] Majid, A.; Khan, A.; Mirza, A.M., "Gender classification using discrete cosine transformation: a comparison of different classifiers," Multi Topic Conference, 2003. INMIC 2003. 7th International, vol., no., pp. 59-64, 8-9 Dec. 2003

[3] Asifullah Khan , Abdul Majid , Anwar M. Mirza, Combination and optimization of classifiers in gender classification using genetic programming, International Journal of Knowledge-based and

Intelligent Engineering Systems, v.9 n.1, p.1-11, January 2005

[4] Shakhnarovich, G., Viola, P.A., Moghaddam, B., 2002. A unified learning framework for real time face detection and classification. In: Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR'02). IEEE, pp. 14–21.

[5] Baback M. Ming-Hsuan Y., "Learning Gender with Support Faces", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.24, 2005.

[6] Tolba S., "Invariant gender identification", Digital Signal Processing, vol 11, no. 3, pp.441-470, 2001.

[7] Zhiguang YANG, Ming LI, and Haizhou AI, "An Experimental Study on Automatic Face Gender Classification," in proceedings of 18th International Conference on Pattern Recognition (ICPR), Hong Kong, volume 3, pp 1099-1102, 24 August 2006.

[8] Z.Y. Xu, L. Lu, and P.F. Shi, "A hybrid approach to gender classification from face images," ICPR, 2008, pp.1-4.

[9] Z. Pan, R. Adams and H. Bolouri, "Dimensionality Reduction for face Images Using Discrete Cosine Transformation for Recognition" Technical Report, Science and Technology Research Centre (STRC)

[10] Scalzo, F. Bebis, G. Nicolescu, M. Loss, L. Tavakkoli, A. ," Feature Fusion Hierarchies for gender classification ", 19th International Conference on Pattern Recognition, ICPR 2008, Pages 1-4,2008.

[11] Abdul Majid, Asifullah Khan and Anwar M. Mirza, "Intelligent combination of Kernels information for improved classification", International conference on machine learning and its applications, Los Angeles, California, USA, Dec. 2005.

[12] Bogdan Gabrysa, and Dymitr Ruta, "Genetic algorithms in classifier fusion", Applied Soft Computing, Pages 337-347, Application of Soft Computing in Information & Communication Technology (ICT), Volume 6, Issue 4, August 2006.

[13] Costen, N.P. Brown, M. Akamatsu, S. , "Sparse models for gender classification", Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FG'04), Pages 201- 206, 2004.

[14] Sun Z., Bebis G., Yuan X. and Louis S.J., "Genetic feature subset selection for gender classification: a comparison study", in IEEE Proceedings on Applications of Computer Vision, pag. 165-170, 2002.

[15] Mäkinen, E. and Raisamo, R. An experimental comparison of gender classification methods. Pattern Recogn. Lett. 29, 10 (Jul. 2008), pp. 1544-1556

[16] Zehang Sun Xiaojing Yuan Bebis, G. Louis, S.J. Neural-network-based gender classification using genetic search foreign-feature selection, International Joint Conference on Neural Networks, IJCNN '02. Volume: 3, On page(s): 2433-2438.

[17] Xia Han, Hassan Ugail, Ian Palmer, "Gender Classification Based on 3D Face Geometry Features Using SVM," Cyberworlds, International Conference on, pp. 114-118, 2009 International Conference on CyberWorlds, 2009.

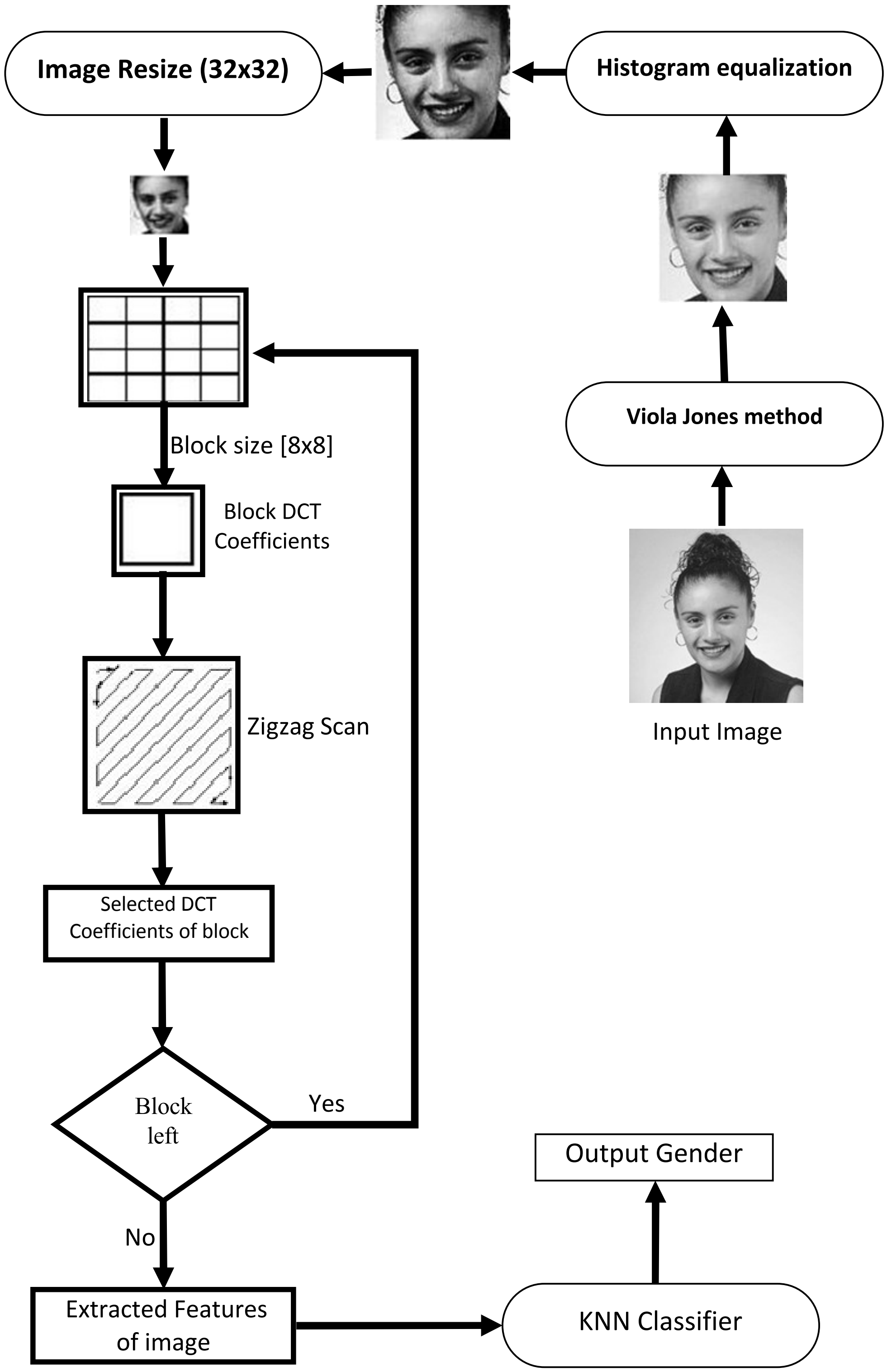


Figure 7. Flowchart of the proposed system.