

Gender Recognition Using Fusion of Local and Global Facial Features

Anwar M. Mirza¹, Muhammad Hussain², Huda Almuzaini², Ghulam Muhammad¹,
Hatim Aboalsamh², and George Bebis^{2,3}

¹Department of Computer Engineering,

²Department of Computer Science, College of Computer and Information Sciences,
King Saud University, Riyadh 11543, Saudi Arabia

³Department of Computer Science and Engineering, University of Nevada at Reno, USA

Abstract. Human perception of the face involves the observation of both coarse (global) and detailed (local) features of the face to identify and categorize a person. Face categorization involves finding common visual cues, such as gender, race and age, which could be used as a precursor to a face recognition system to improve recognition rates. In this paper, we investigate the fusion of both global and local features for gender classification. Global features are obtained using the principal component analysis (PCA) and discrete cosine transformation (DCT) approaches. A spatial local binary pattern (LBP) approach augmented with a two-dimensional DCT approach has been used to find the local features. The performance of the proposed approach has been investigated through extensive experiments performed on FERET database. The proposed approach gives a recognition accuracy of 98.16% on FERET database. Comparisons with some of the existing techniques have shown a marked reduction in number of features used per image to produce results more efficiently and without loss of accuracy for gender classification.

Keywords: Fusion of features, principal component analysis (PCA), discrete cosine transformation (DCT), local binary pattern (LBP) approach, FERET database.

1 Introduction

Categorization of facial images, before passing it on to a face recognition system, can lead to significant improvements in the performance of the recognition system. This categorization can be achieved using visual cues like gender, race and age[1]. The notion of humans utilizing visual cues for face recognition is supported by a number of studies. Brigham [2], O'Toole et al. [3] and Philips et al. [4] have shown that people are more accurate at recognizing faces from their own race than faces from other races. Cheng et al. [5] and Baudouin et al. [6] have demonstrated that human use gender information along with other face categories for face recognition purposes. Gender recognition is important for its use in numerous applications including identity authentication, demographic data collection, search engine retrieval accuracy, human-computer interaction, access control, and surveillance.

Yamaguchi et al. [7] and Wild et al. [8] created prototypical male and female faces using a morphing technique. Their studies showed global structural differences between prototypical boy and prototypical girl faces. This clearly indicates that the global structural features of the face play an important role in the process of face categorization according to gender.

Determination of the salient features from facial images is one of the most important steps in the process of face and category specific face recognition. If the extracted features do not help in discriminating between different facial images, the use of an excellent classification algorithm can lead to very bad results. The feature extraction can be subdivided into either appearance based or geometric methods. Features including distance between eyes; size and location of the eyes, nose, mouth and ears; overall size of the face etc. are used in the geometric methods. In appearance based methods, features are extracted from the whole facial image.

SEXTNET was the first application based on the artificial neural networks to identify gender from face images [9]. A small database of 90 face images of 45 male and 45 female images was used in this study to train a 2-layer fully connected neural net for classification purposes. Using geometrical features as an input, Brunelli et al. [10] developed a radial basis function (RBF) based competing network for gender classification. They reported an accuracy of 79% in their work. Gutta et al. [11] proposed a hybrid approach based on the use of RBFs and decision trees for gender classification. A very important contribution was made by Moghaddam et al. [12] when they employed support vector machine (SVM) for gender classification and reported a misclassification rate of 3%. Yang et al. [13] improved gender classification using texture normalization. Several weak classifiers were combined using AdaBoost by Baluja and Rowley [14] in their gender classification system. They used small sized normalized images from FERET database. Their system showed an overall 90% recognition rate. Lu and Shi [15] used the fusion of left eye, upper face region and nose in their study. They showed that their fusion approach outperforms the whole face approach. This work was later extended by Alexandre [16] to combine shape and texture features from differenced sized normalized images. They used local binary pattern (LBP) for the texture features.

The rest of the paper is organized as follows. Section 2 presents an overview of the feature extraction techniques. This section describes the mathematical foundations of the global and local feature extraction techniques used in this paper. In Section 3 the proposed methodology adopted in this paper is described. Experimental results and discussion are given in section 4. Concluding remarks about the present work are given in section 5.

2 Feature Extraction Techniques

Zhao et al. [17], Sinha et al. [18] and Ng et al. [19] have given detailed reviews of some of the most important face recognition techniques in general and gender recognition from facial images in particular. Most of the commonly used techniques either rely on just the global features or local features. The process of gender recognition by

humans involves first scanning of the face and then observing the detailed features of the face. In this paper we employ both global and local features for gender recognition, that are fused together to form a compact feature representation for the facial images.

2.1 Global Feature Extraction

We have employed two global feature extraction techniques, namely, the eigenface approach based on the reconstruction of face images employing principal component analysis (PCA) and the two-dimensional discrete cosine transform (2D-DCT) approach.

In the eigenface approach, a face image $f(x, y)$ of size $R \times C$, is converted into a column vector X of size $N = R \times C$, by concatenating all the columns of $f(x, y)$. The image (now represented by) X is projected onto a low dimensional vector Y of size M , such that ($M < N$) using

$$Y = UX \tag{1}$$

where U is the projection matrix of size $M \times N$. The column vectors of the projection matrix U are the eigenfaces obtained from the PCA approach. The projected vector Y is the global feature vector in this case.

Let X_i be one sample face image in the training dataset of size L images. The mean is given by

$$\mu = \frac{1}{L} \sum_{i=1}^L X_i \tag{2}$$

The scatter matrix S for all sample images is obtained by

$$S = \sum_{i=1}^L (X_i - \mu)(X_i - \mu)^T \tag{3}$$

In the PCA approach, eigenvectors of the scatter matrix S are obtained. These eigenvectors are sorted according to the descending values of the eigenvalues of S . First M eigenvectors of S form the column vectors of the projection matrix U . From equation (1), it could be noticed that each face image X_i now can be represented by a lower dimensional feature vector Y_i .

In the second feature extraction approach we have used 2D-DCT of the sample face image matrix $f_i(x, y)$ having R rows and C columns, using the transformation

$$g_i(u, v) = \frac{2}{\sqrt{RC}} a(u)a(v) \sum_{x=0}^{R-1} \sum_{y=0}^{C-1} f_i(x, y) \cos\left(\frac{(2x + 1)u\pi}{2R}\right) \cos\left(\frac{(2y + 1)v\pi}{2C}\right) \tag{4}$$

where

$$a(u) = \begin{cases} \sqrt{1/R} & \text{for } u = 0 \\ \sqrt{2/R} & \text{for } u = 1, 2, 3, \dots, R - 1 \end{cases} \tag{5}$$

and similarly for $a(v)$. The discrete cosine transformation has a high information packing ability. This is the reason for its use in image compression technologies. The most significant DCT coefficients of Y_i are chosen by determining which of the coefficients have the greatest variance. For natural images, the DCT coefficient's energy drops off very rapidly for higher frequency components. This compaction property is used in the dimension reduction for face images. After obtaining the transformed matrix $g_i(u, v)$, its coefficients are sorted according to the zigzag scanning technique of image compression [Gonzalez ref]. A dimensionally reduced feature vector based on 2D-DCT is obtained by choosing first M coefficients of Y_i after performing zigzag scanning.

2.2 Local Feature Extraction

Local binary pattern (LBP) descriptor computed using the LBP operator was first introduced by Ojala et al. [20]. It was initially used as a texture descriptor giving very promising results in many applications [21], [22], and [23]. Ahonen et al. [24] used it for the first time for face recognition. Sun et al. [25] and Lian et al. [26] extended its use to gender recognition using facial images. The initial LBP operator associates a label with each pixel of an image; the labeling process involves converting each pixel value in the 3×3 neighborhood of a pixel into a binary digit (0 or 1) using the center value as a threshold and concatenating the bits, as shown in Figure 1. Later the operator was extended to general neighborhood sizes, and its rotation invariant and uniform versions were introduced [21].

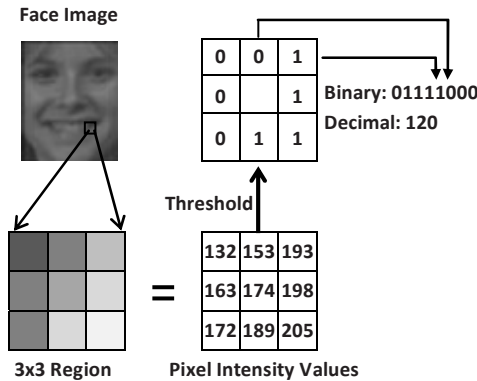


Fig. 1. Example of basic LBP operator

The general LBP operator is denoted by $LBP_{p,R}$ and is defined as:

$$LBP_{p,R} = \sum_{i=1}^{p-1} 2^i S(p_i - p_c) \tag{6}$$

where P is the total number of pixels in the neighborhood and R is its radius, p_c is the center pixel. The thresholding operation is defined as follows:

$$S(p_i - p_c) = \begin{cases} 1 & p_i - p_c \geq 0 \\ 0 & p_i - p_c < 0 \end{cases} \quad (7)$$

After applying the $LBP_{P,R}$ operation on the original image $f(x, y)$, a labeled age $f_l(x, y)$ is obtained. The histogram of the labels is used as a texture descriptor. The histogram $H(i)$ of the labeled image is defined as:

$$H(i) = \sum_{x=1}^R \sum_{y=1}^C I\{f_l(x, y) = i\}, \quad \text{for } i = 0, \dots, n - 1 \quad (8)$$

where n is the number of different labels produced by the LBP operator and

$$I\{x\} = \begin{cases} 1 & x \text{ is true} \\ 0 & x \text{ is false} \end{cases} \quad (9)$$

General LBP operator has three parameters: circular neighborhood (P, R), rotation invariance (ri) and uniformity (u)[21]. For a particular application, it is necessary to explore this parameter space to come up with the best combination of these parameters.

Figure 2 shows the histogram extracted from an image with LBP operator. An LBP histogram in this approach contains information about facial micro-patterns like the distribution of edges, spots and flat areas over the whole image. In case of $(P, R) = (8, 1)$ neighborhood, there are 256 unique labels, and the dimension of LBP histogram descriptor is 256.

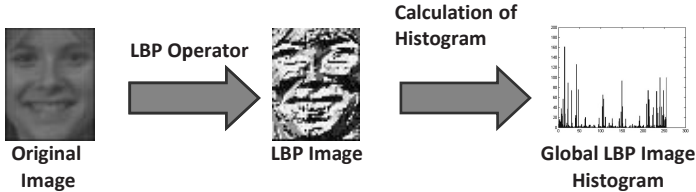


Fig. 2. LBP histogram calculation for the whole image

It could be noticed that, the basic LBP histogram descriptor obtained in this case is for the whole image and can be considered as a global descriptor for the image. The descriptor represents facial patterns but their spatial location information is lost due to the global histogram operation. We, however, have employed block by block version of the LBP labeled image (as explained in the proposed approach of section 4) which gives rise to a local representation of the facial features in the image.

3 Proposed Methodology

We followed the general steps of pre-processing, feature extraction, feature selection and classification of pattern recognition in our work. The main contribution of this

work is in the feature extraction stage of the overall procedure. The pre-processing stage consists of face normalization which crops out the background from each of the face image and centers it according to the location of the left and right eyes of the subject. For feature selection, we use the procedure given in the feature extraction part to form an overall feature vector of the most significant features for discriminating the faces according to gender.

A schematic diagram of our proposed methodology is shown in Figure 3. The input image, after the preprocessing step, passes through the feature extraction stage. We apply 2D-DCT or PCA on the whole image to obtain the global feature vector. Both these techniques are used for dimension reduction. Therefore, only the most significant N_G feature vector values are selected after sorting the original global feature vector.

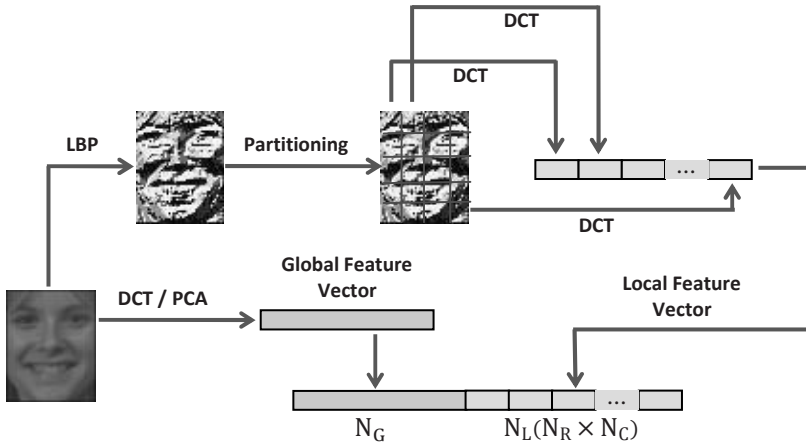


Fig. 3. Schematic diagram of the proposed methodology

For the local feature vector, we first apply LBP operator on the whole image. The labeled LBP obtained is partitioned into $N_R \times N_C$ blocks of pre-defined sizes. We then apply 2D-DCT on each of these LBP blocks to obtain the contribution of local features to the overall local feature vector. Only the most significant N_L 2D-DCT components are selected after zigzag scanning of each of the resulting DCT coefficients. The block-wise local feature are concatenated to obtain the overall local feature vector. Specifically, the following steps are carried out in the feature extraction stage.

- Global Feature Extraction
 - All training and testing images are fed to the global feature extraction algorithm. For each image, do the following steps.
 - Apply either 2D-DCT or PCA to obtain the global features.
 - Select N_G global features to form the global feature vector FG .
- Local Feature Extraction
 - All training and testing images are fed to the local feature extraction algorithm. For each image do the following steps.

- Apply the LBP algorithm to the whole image to obtain the LBP labeled image.
 - Partition the LBP labeled image into $NR \times NC$ blocks.
 - Apply the 2D-DCT on each of the block to get the DCT coefficients.
 - Employ zigzag scanning function to select the low frequency DCT coefficients.
 - Select only NL DCT coefficients from each block.
 - Concatenate the selected DCT coefficients from each block to form the overall local feature vector FL of size $NL \times NR \times NC$ coefficients.
- Fusion of the Features
 - Combine the global and local feature vectors to form the overall feature vector for the whole image $F = FG, FL$.

We employ k-nearest neighbors (kNN) classifier to classify faces according to the overall feature vectors.

4 Experiments and Discussion

We have performed experiments on FERRET database [27], which is considered as a challenging database for face recognition. The database consists of frontal, left or right profile images and could have some variations in pose, expression and lightning. We used 1204 face images consisting of 746 male and 458 female images for training and 1196 images comprising 740 male and 456 female images for testing. Each image is normalized and cropped to the size 60×48 pixels. Some of the images taken from FERET database are shown below.



Different values for the parameters N_G and N_L have been used to check the performance of the algorithm. For the local feature extraction part, blocks of different sizes including 32×24 , 16×16 , 12×12 and 6×6 , have been used. Also LBP variants with uniform mapping, no mapping and $P = 8$, and $R = 1$ have been employed.

Our experiments have shown that the best performance for different block sizes is obtained for different combinations of the global and local features. Figure 4 shows the recognition rates for different block sizes. The lengths of the feature vectors (after performing fusion) were found to be 690 512, 489 and 589 for blocks of sizes 32×24 , 16×16 , 12×12 and 6×6 , respectively. The overall best performance was given by block size 12×12 , yielding a recognition rate of 98.16% with 9 global 2D-DWT features and 24 local LPB+DCT features. A poorer performance was achieved in case of the use of PCA for global features. In all experiments a kNN classifier was used. Comparison of the performance evaluation with different distance measure from including city block, chi-square, cosine and Euclidean showed that the city block distance measure gives the best results in this case. It is noticed that decreasing the block size does not improve the performance of the system. In fact, the

number of features required to produced the best results for small size block also increases, resulting in more computations.

We also have compared our approach with the competing techniques including PCA (as a baseline approach), dyadic wavelets (DyWT) approach [28], histogram based LBP (HLBP) [25], and Multi-resolution Decision Fusion method (MDF) [16]. The results shown in Figure 5 indicate that our proposed system gives better recognition rates as compared to PCA, DyWT and HLBP approaches. The MDF approach however is reported to give better accuracies. However, it should be noticed that the proposed approach achieves 98.16% accuracy with fewer number of features as compared to MDF, which works at multiple scales (including 20×20 , 36×36 , and 128×128). It required extraction of shape as well as texture features for each of these scales and independent tuning of the classifier for each scale separately. Therefore, our proposed scheme comparable yields performance at a much lesser computational cost and better efficiency.

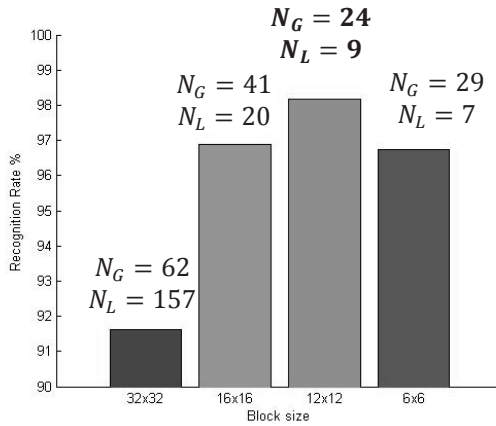


Fig. 4. Effect of the block size size on the recognition rate

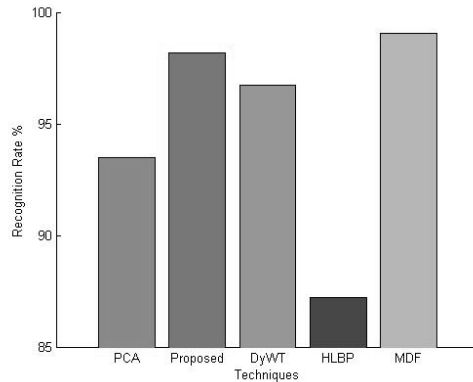


Fig. 5. Comparison of the proposed technique with the competing techniques

5 Conclusion

In this paper, we have presented a fusion strategy for the combination of local and global facial feature to address the gender recognition problem. The new approach make use selected 2D-DCT features for the overall face image. The local features are obtained by applying LBP approach on the facial image, which is then partitioned into blocks, followed by the application of 2D-DCT on each block and selection of the most significant features. The proposed methodology gives a recognition accuracy of 98.16% with only 489 features. We found that block size of 12×12 , with uniform mapping and neighborhood $(P, R) = (8, 1)$ for LBP yield the best accuracy. A comparison of our proposed scheme with the competing techniques indicates a better performance in terms of accuracy as well as the number of feature vector length used in those techniques. In future, we plan to explore the use of this strategy for other face databases and sophisticated classifiers like SVM.

Acknowledgement. The work presented in this paper is supported by the National Plan for Science and Technology, King Saud University, Riyadh, Saudi Arabia under project number 10-INF1044-02.

References

1. Yang, S., Bebis, G., Hussain, M., Muhammad, G., Mirza, A.M.: Unsuper-vised discovery of visual face categories. *International Journal on Artificial Intelligence Tools* (May 2012) (accepted) doi: 10.1142/S0218213012500297
2. Brigham, J.: The influence of race on face recognition. In: Ellis, H., Jeeves, M., Newcombe, F. (eds.) *Aspects of Face Processing*, pp. 170–177 (1986)
3. O'Toole, A., Peterson, J., Deffenbacher, K.: An other-race effect for classifying faces by sex. *Perception* 25, 669–676 (1996)
4. Phillips, P.J., Jiang, F., Narvekar, A., Ayyad, J., O'Toole, A.: An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception* 8(2) (2011)
5. Cheng, Y.D., O'Toole, A., Abdi, H.: Classifying adults' and children's faces by sex: Computational investigations of subcategorical feature encoding. *Cognitive Science* 25(5), 819–838 (2001)
6. Baudouin, J.Y., Tiberghien, G.: Gender is a dimension of face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(2), 362–365 (2002)
7. Yamaguchi, M.K., Hirukawa, T., Kanazawa, S.: Judgement of gender through facial parts. *Perception* 24, 563–575 (1995)
8. Wild, H., Barrett, S., Spence, M., O'Toole, A., Chenh, Y., Brooke, J.: Recognition and sex classification of adults' and children's faces: examining performance in the absence of sex-stereotypes cues. *Journal of Experimental Child Psychology* 77, 269–291 (2000)
9. Golom, A., Lawrence, D.T., Sejnowski, T.J.: SEXNET: A neural network identifies gender from human faces. In: *Advances in Neural Information Processing Systems*, vol. 3, pp. 572–577 (1991)
10. Brunelli, R., Poggio, T.: HyperBF network for gender classification. In: *DARPA Image Understanding Workshop*, pp. 311–314 (1992)

11. Gutta, S., Wechsler, H., Phillips, P.: Gender and ethnic classification of face images. In: 3rd IEEE International Conference on AUtomatic Face and Desture Recognition, FG 1998, pp. 194–199 (1998)
12. Moghaddam, B., Yang, M.-H.: Gender classification with support vector machines. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 306–311 (2000)
13. Yang, Z., Li, M., Ai, H.: An experimental study on automatic face gender classification. In: Proc. IEEE Int. Conf. on Pattern Recognition, pp. 1099–1102 (2006)
14. Baluja, S., Rowley, H.: Boosting sex identification performance. *International Journal of Computer Vision* 71(1), 111–119 (2007)
15. Lu, L., Shi, P.: Fusion of multiple facial regions for expression-invariant gender classification. *IEICE Electronics Express* 6(10), 587–593 (2009)
16. Alexandre, L.A.: Gender recognition: A multiscale decision fusion approach. *Pattern Recognition Letters* 31(11), 1422–1427 (2010)
17. Zhao, W., Cellappa, R., Rosenfeld, A., Phillips, P.J.: *Face Recognition: A Literature Survey*. ACM Computing Surveys, 399–458 (2003)
18. Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face Recognition by Humans: 199 Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE* 94(11), 1948–1962 (2006)
19. Ng, C.B., Tay, Y.H., Goi, B.-M.: Recognizing Human Gender in Computer Vision: A Survey. In: Anthony, P., Ishizuka, M., Lukose, D. (eds.) PRICAI 2012. LNCS, vol. 7458, pp. 335–346. Springer, Heidelberg (2012)
20. Ojala, T., Pietkainen, M., Harwood, D.: A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognition* 29(1), 51–59 (1996)
21. Ojala, T., Pietkainen, M., Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
22. Zhang, G., Huang, X., Li, S.Z., Wang, Y., Wu, X.: Boosting local binary pattern (LBP)-based face recognition. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) SINOBIOMETRICS 2004. LNCS, vol. 3338, pp. 179–186. Springer, Heidelberg (2004)
23. Liu, H., Sun, J., Liu, L., Zhang, H.: Feature selection with dynamic mutual information. *Journal of Pattern Recognition* 42(7) (July 2009)
24. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004, Part I. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
25. Sun, N., Zheng, W., Sun, C., Zou, C., Zhao, L.: Gender Classification Based on Boosting Local Binary Pattern. In: Wang, J., Yi, Z., Žurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3972, pp. 194–201. Springer, Heidelberg (2006)
26. Lian, H., Lu, B.: Multi-view gender classification using multi-resolution local binary patterns and support vector machines. *International Journal of Neural Systems* 17(6), 479–487 (2007)
27. Phillips, P.J., Hyeonjoon, M., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(10), 1090–1104 (2000)
28. Abdulkirim, T., Hussain, M., Nijjima, K., Takano, S.: The Dyadic Lifting Schemes and the Denoising of Digital Images. *International Journal of Wavelets, Multiresolution and Information Processing* 6(3), 331–351 (2008)