A Spectral Domain Local Feature Extraction Algorithm for Face Recognition

Hafiz Imtiaz hafiz.imtiaz@live.com
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh

Shaikh Anowarul Fattah sfattah@princeton.edu
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh

Abstract

In this paper, a spectral domain feature extraction algorithm for face recognition is proposed, which efficiently exploits the local spatial variations in a face image. For the purpose of feature extraction, instead of considering the entire face image, an entropy-based local band selection criterion is developed, which selects high-informative horizontal bands from the face image. In order to capture the local variations within these high-informative horizontal bands precisely, a feature selection algorithm based on two-dimensional discrete Fourier transform (2D-DFT) is proposed. Magnitudes corresponding to the dominant 2D-DFT coefficients are selected as features and shown to provide high within-class compactness and high between-class separability. A principal component analysis is performed to further reduce the dimensionality of the feature space. Extensive experimentations have been carried out upon standard face databases and the recognition performance is compared with some of the existing face recognition schemes. It is found that the proposed method offers not only computational savings but also a very high degree of recognition accuracy.

Keywords: Feature Extraction, Classification, Two Dimensional Discrete Fourier Transform, Dominant Spectral Feature, Face Recognition, Modularization.

1. INTRODUCTION

Automatic face recognition has widespread applications in security, authentication, surveillance, and criminal identification. Conventional ID card and password based identification methods, although very popular, are no more reliable as before because of the use of several advanced techniques of forgery and password-hacking. As an alternative, biometric, which is defined as an intrinsic physical or behavioral trait of human beings, is being used for identity access management [1]. The main advantage of biometric features is that these are not prone to theft and loss, and do not rely on the memory of their users. Among physiological biometrics, face is getting more popularity because of its non-intrusiveness and high degree of security. Moreover, unlike iris or finger-print recognition, face recognition do not require high precision equipment and user agreement, when doing image acquisition, which make face recognition even more popular for video surveillance.

Nevertheless, face recognition is a complicated visual task even for humans. The primary difficulty in face recognition arises from the fact that different images of a particular person may vary largely, while images of different persons may not necessarily vary significantly. Moreover, some aspects of the image, such as variations in illumination, pose, position, scale, environment, accessories, and age differences, make the recognition task more complicated. However, despite many relatively successful attempts to implement face recognition systems, a single approach, which is capable of addressing the hurdles, is yet to be developed.

Face recognition methods are based on extracting unique features from face images. In this regard, face recognition approaches can be classified into two main categories: holistic and texture-based [2]-[4]. Holistic or global approaches to face recognition involve encoding the entire facial image in a high-dimensional space [2]. It is assumed that all faces are constrained to particular positions, orientations, and scales. However, texture-based approaches rely on the detection of individual facial characteristics and their geometric relationships prior to performing face recognition [3], [4]. Apart from these approaches, face recognition can also be performed by using different local regions of face images [5], [6]. It is well-known that, although face images are affected due to variations, such as non-uniform illumination, expressions and partial occlusions, facial variations are confined mostly to local regions. It is expected that capturing these localized variations of images would result in a better recognition accuracy [6]. Hence, it is motivating to utilize local variations for feature extraction and thereby develop a scheme of holistic face recognition incorporating the advantageous properties of texture-based approach.

The objective of this paper is to develop a spectral domain face recognition scheme based on dominant spectral features extracted from local zones instead of using the entire face image as a whole. In order to exploit the high-informative areas of a face image, an entropy based horizontal band selection criterion is presented. Such high-informative bands are further divided into some smaller spatial modules to extract local variations in detail. A spectral domain feature extraction algorithm using 2D-DFT is developed, which operates within those local zones to extract dominant spectral features. It is shown that the discriminating capabilities of the proposed features are enhanced because of modularization of the face images. In view of further reducing the computational complexity, principal component analysis is performed on the proposed feature space. Finally, the face recognition task is carried out using a distance based classifier..

2. BRIEF DESCRIPTION OF THE PROPOSED SCHEME

A typical face recognition system consists of some major steps, namely, input face image collection, pre-processing, feature extraction, classification and template storage or database, as illustrated in Fig. 1. The input image can be collected generally from a video camera or still camera or surveillance camera. In the process of capturing images, distortions including rotation, scaling, shift and translation may be present in the face images, which make it difficult to locate at the correct position. Pre-processing removes any un-wanted objects (such as, background) from the collected image. It may also segment the face image for feature extraction. For the purpose of classification, an image database is needed to be prepared consisting template face poses of different persons. The recognition task is based on comparing a test face image with template data. It is obvious that considering images themselves would require extensive computations for the purpose of comparison. Thus, instead of utilizing the raw face images, some characteristic features are extracted for preparing the template. It is to be noted that the recognition accuracy strongly depends upon the quality of the extracted features. Therefore, the main focus of this research is to develop an efficient feature extraction algorithm.

The proposed feature extraction algorithm is based on extracting spatial variations precisely from high informative local zones of the face image instead of utilizing the entire image. In view of this, an entropy based selection criterion is developed to select high informative facial zones. A modularization technique is employed then to segment the high informative zones into several smaller segments. It should be noted that variation of illumination of different face images of the same person may affect their similarity. Therefore, prior to feature extraction, an illumination adjustment step is included in the proposed algorithm. After feature extraction, a classifier compares features extracted from face images of different persons and a database is used to store registered templates and also for verification purpose.

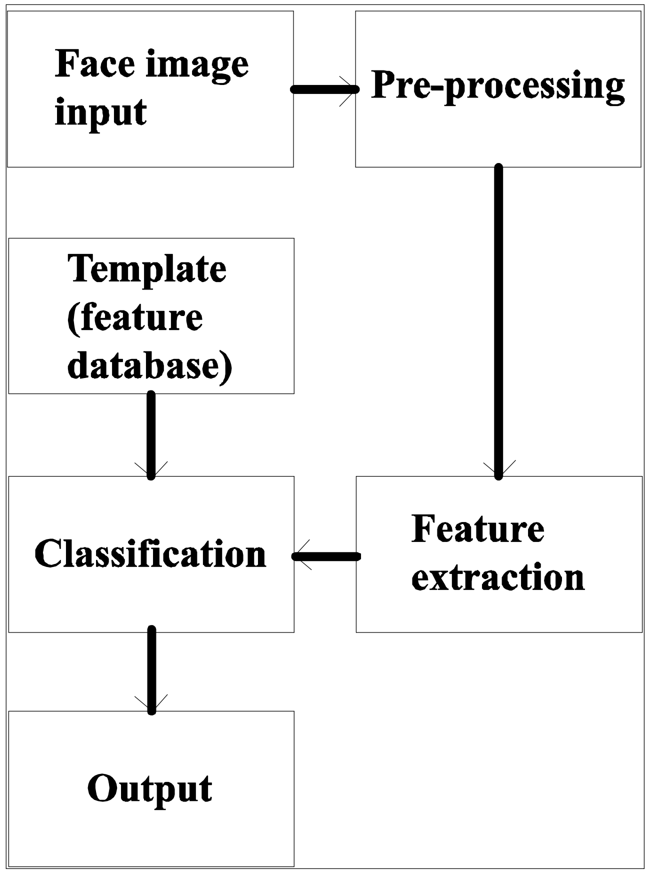

FIGURE 1: Block diagram of the proposed method

3. PROPOSED METHOD

For any type of biometric recognition, the most important task is to extract distinguishing features from the training biometric traits, which directly dictates the recognition accuracy. In comparison to person recognition based on different biometric features, face image based recognition is very challenging even for a human being, as face images of different persons may seem similar whereas face images of a single person may seem different, under different conditions. Thus, obtaining a significant feature space with respect to the spatial variation in a human face image is very crucial. Moreover, a direct subjective correspondence between face image features in the spatial domain and those in the frequency domain is not very apparent. In what follows, we are going to demonstrate the proposed feature extraction algorithm for face recognition, where spatial domain local variation is extracted using frequency domain transform.

3.1 Entropy Based Horizontal Band Selection

The information content of different regions of a human face image varies widely [7]. It can be shown that, if an image of a face were divided into certain segments, not all the segments would contain the same amount of information. It is expected that a close neighborhood of eyes, nose and lips contains more information than that possessed by the other regions of a human face image. It is obvious that a region with high information content would be the region of interest for the purpose of feature extraction. However, identification of these regions is not a trivial task. Estimating the amount of information from a given image can be used to identify those significant zones. In this paper, in order to determine the information content in a given area of a face image, an entropy based measure of intensity variation is defined as [8]

$$H = -\sum_{k=1}^{m} p_k log_2 p_k, \quad (1)$$

where the probabilities $\{p_k\}_1^m$ are obtained based on the intensity distribution of the pixels of a segment of an image. It is to be mentioned that the information in a face image exhibits variations more prominently in the vertical direction than that in the horizontal direction [9]. Thus, the face image is proposed to be divided into several horizontal bands and the entropy of each band is to be computed. It has been observed from our experiments that variation in entropy is closely related to variation in the face geometry. Fig. 2(b) shows the entropy values obtained in different horizontal bands of a person for several sample face poses. One of the poses of the person is shown in Fig. 2(a). As expected, it is observed from the figure that the neighborhood of eyes, nose and lips contains more information than that possessed by the other regions. Moreover, it is found that the locus of entropies obtained from different horizontal bands can trace the spatial structure of a face image. Hence, for feature extraction in the proposed method, spatial horizontal bands of face images are chosen corresponding to their entropy content.

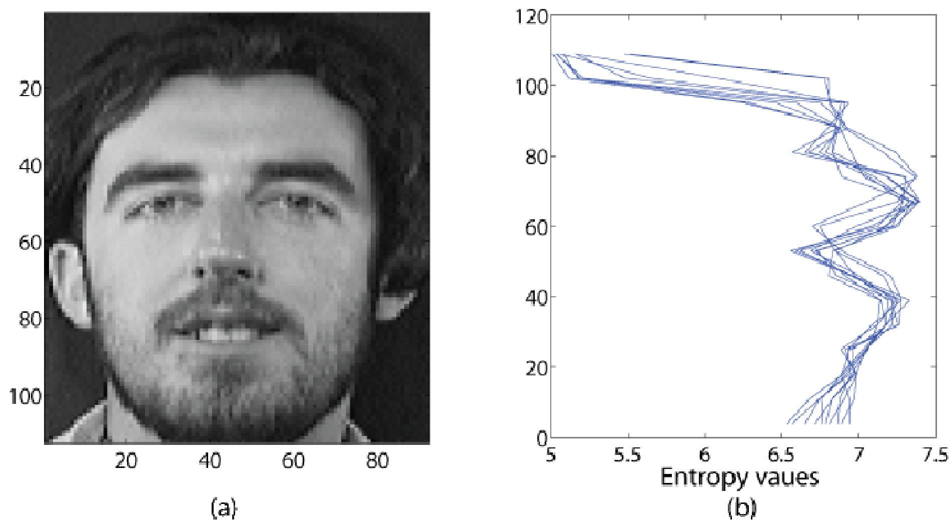

FIGURE 2: (a) Sample face image of a person and (b) entropy values in different horizontal bands of several face poses

3.2 Spectral Feature Extraction and Illumination Adjustment

For biometric recognition, feature extraction can be carried out using mainly two approaches, namely, the spatial domain approach and the spectral domain approach [10]. The spatial domain approach utilizes the spatial data directly from the face image or employs some statistical measure of the spatial data. On the other hand, spectral domain approaches employ some kind of transform over the face images for feature extraction. In case of spectral domain feature extraction, pixel-by-pixel comparison between face images in the spatial domain is not necessary. Phenomena, such as rotation, scale and illumination, are more severe in the spatial domain than in spectral domain. Hence, in what follows, we intend to develop a feature extraction algorithm based on spectral domain transformation. We have employed an efficient feature extraction scheme using Fourier transform, which offers an ease of implementation in practical applications. For a function $f(x,y)$ with two-dimensional variation, the 2D Fourier transform is given by [11]

$$\mathcal{F}(\omega_x,\omega_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(\omega_x x+\omega_y y)}dxdy, \quad (2)$$

where ω_x and ω_y represent frequencies in the two-dimensional space.
It is intuitive that images of a particular person captured under different lighting conditions may vary significantly, which can affect the face recognition accuracy. In order to overcome the effect of lighting variation in the proposed method, illumination adjustment is performed prior to feature extraction. Given two images of a single person having different intensity distributions due to variation in illumination conditions, our objective is to provide with similar feature vectors for these two images irrespective of the different illumination condition. Since in the proposed method, feature extraction is performed in the Fourier domain, it is of our interest to analyze the effect of variation in illumination on the DFT-based feature extraction.

In Fig. 3, two face images of the same person are shown, where the second image (shown in Fig. 3(b) is made brighter than the first one by changing the average illumination level. 2D-DFT is performed upon each image, first without any illumination adjustment and then after performing illumination adjustment. Considering all the 2D-DFT coefficients to form the feature vectors for these two images, a measure of similarity can be obtained by using correlation. In Figs. 4 and 5, the cross-correlation values of the 2D-DFT coefficients obtained by using the two images without and with illumination adjustment are shown, respectively. It is evident from these two figures that the latter case exhibits more similarity between the DFT coefficients indicating that the features belong to the same person. The similarity measure in terms of Euclidean distances between the 2D-DFT coefficients of the two images for the aforementioned two cases are also calculated. It is found that there exists a huge separation in terms of Euclidean distance when no illumination adjustment is performed, whereas the distance completely diminishes when illumination adjustment is performed, as expected, which clearly indicates that a better similarity between extracted feature vectors.

3.3 Proposed Dominant Spectral Feature

It has already been mentioned that, in the proposed algorithm, instead of taking the DFT coefficients of the entire image, the coefficients obtained form each modules of the high-informative horizontal band of a face image are considered to form the feature vector of that image. However, if all of these coefficients were used, it would definitely result in a feature vector with a very large dimension. One advantage of working in the frequency domain is that a few DFT coefficients with higher magnitudes would be sufficient to represent an image or a portion of an image. Hence, in view of reducing the feature dimension, we propose to utilize the magnitudes and 2D-frequencies corresponding to the dominant DFT coefficients as spectral features.

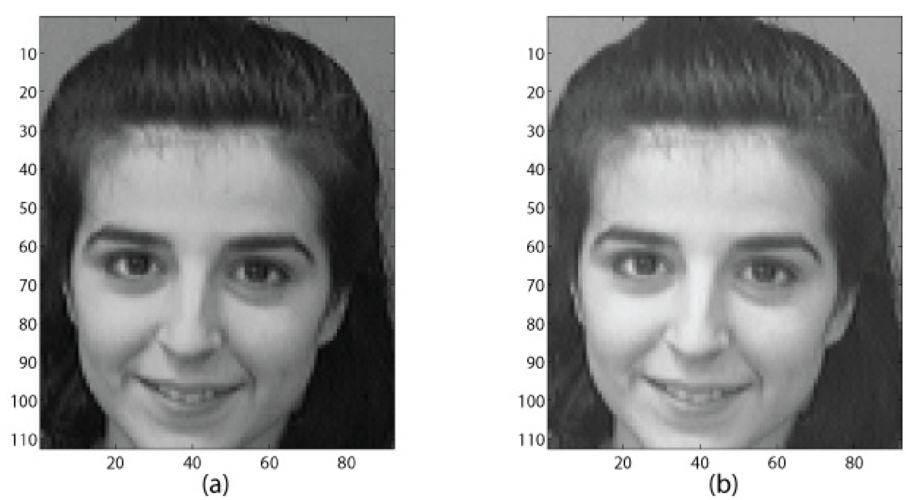

FIGURE 3: Two face images of the same person under different illumination

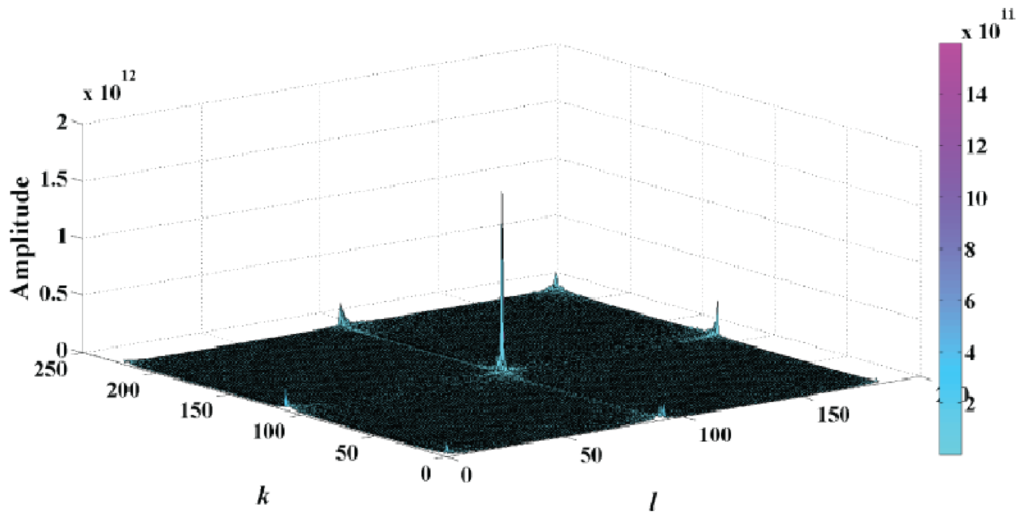

FIGURE 4: Correlation of the 2D-DFT coefficients of the sample images: no illumination adjustment

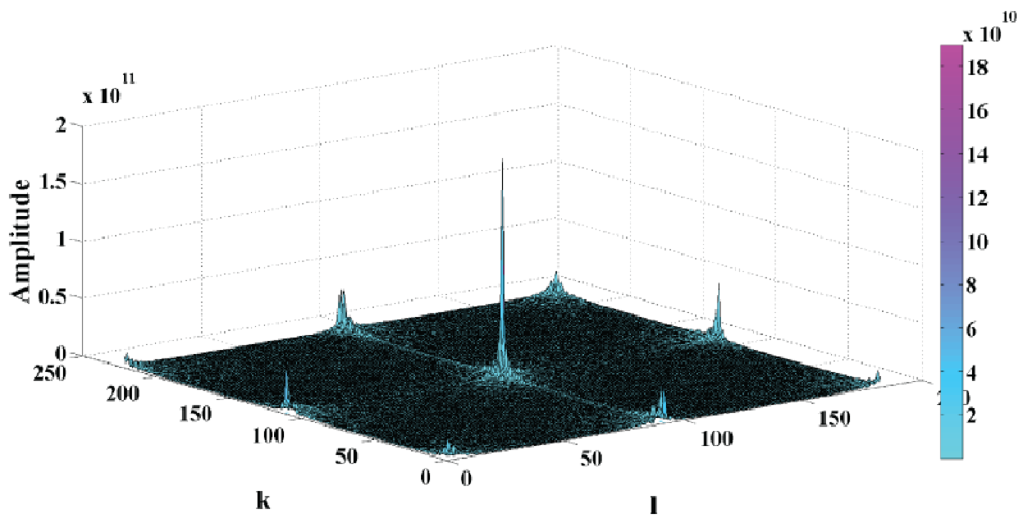

FIGURE 5: Correlation of the 2D-DFT coefficients of the sample: illumination adjusted

The 2D-DFT coefficient corresponding to the maximum magnitude is treated as the dominant coefficient ($D1$). Considering the magnitudes of the 2D-DFT coefficients in descending order, magnitude values other than the dominant one may also be treated as possible candidates for desired features. In accordance with their magnitude values, these dominant magnitudes are termed as second-dominant $(D2)$, third-dominant $(D3)$, and so on. If the magnitude variations along all the segments for the case of different dominant magnitudes remain similar, it would be very difficult to select one of those dominant magnitudes as a desired feature.

In order to demonstrate the characteristics of the dominant magnitudes in different modules, sample face images of two different persons are shown in Fig. 6. In Fig. 7, four dominant magnitudes $(D1, D2, D3, and D4)$ obtained from all the modules of the image of Person 1, appeared in Fig. 6(a), are shown. The face image is divided into 16 modules. It is found that different dominant magnitudes obtained from the spatial modules exhibit completely different characteristics. However, the magnitude value for the first dominant $(D1)$ is found reasonably higher than other dominant magnitudes. An analogous behavior, as shown in Fig. 7, is obtained for Person 2 of Fig. 6(b). It is evident from Fig. 7 that $D1$ is the most significant among all the dominant magnitudes and thus, it is sufficient to consider only $D1$ as a desired feature, which also offers an advantage of reduced feature dimension. Computing $D1$ in each segment of the high-informative horizontal bands of a face image, the proposed feature vector is obtained.

It should be noted that, for a high-informative horizontal band of dimension $N \times N$ with M number of modules (of dimension $n \times n$), considering only $D1$ will reduce the length of feature vector from $M \times n \times n$ to M, an order of n^2 reduction.

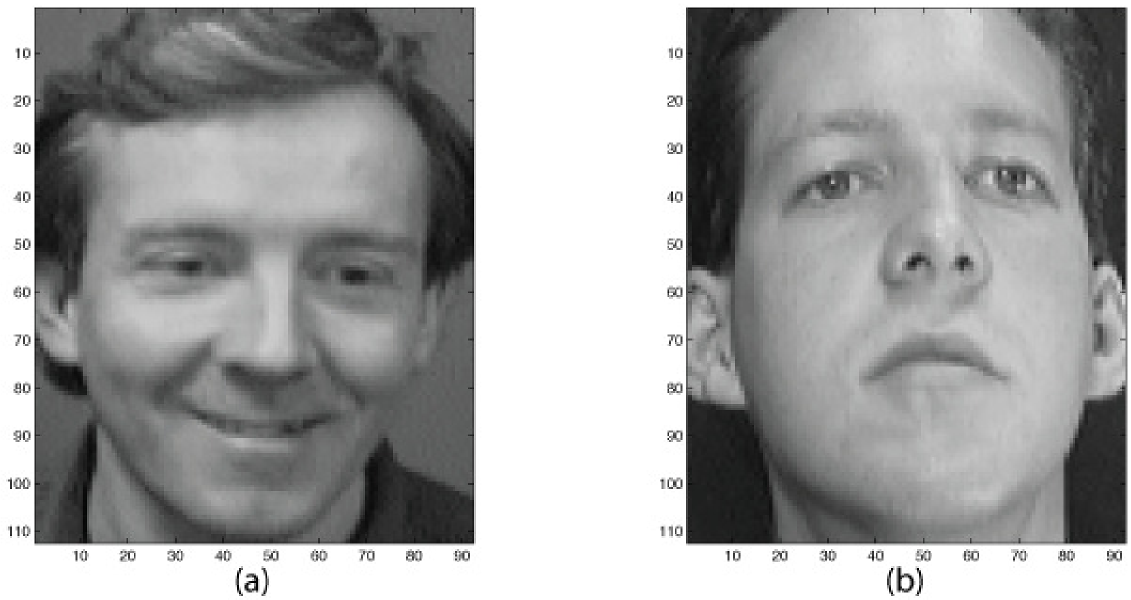

FIGURE 6: Sample face images of two persons

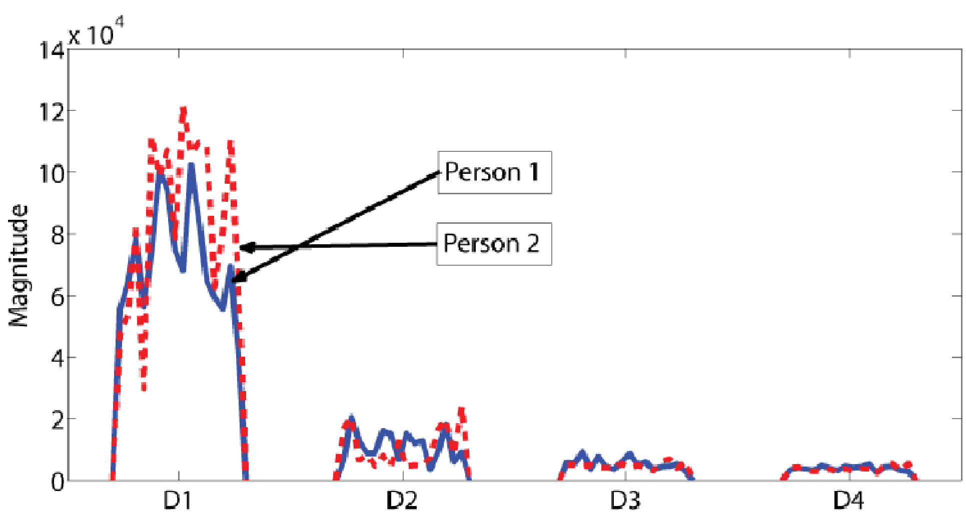

FIGURE 7: Proposed dominant magnitude-features

In order to investigate the effect of division of horizontal bands into smaller segments on the characteristics of dominant magnitude features, Fig. 8 is presented, where the dominant magnitudes obtained from a single high-informative horizontal band of the different poses of the same two persons shown in Fig. 6 are shown. It is observed from the figure that the dominant feature magnitudes for different poses of a particular person are significantly scattered resulting in poor within-class compactness. Moreover, the dominant feature magnitudes of two different persons are substantially overlapped resulting in poor between-class separation.

It is observed that a significant variation may occur in the poses of a single person taken under different conditions. In view of demonstrating the effect of such variations on the proposed dominant features, we consider ten sample poses for each of the two persons as appeared in Fig. 6. In Fig. 9, the proposed dominant features obtained from different modules of the high-informative horizontal bands of all the face poses of two different persons are shown. For each person, the centroid of the proposed feature vectors is also shown in the figure (in thick continuous lines). It is to be noted that the feature centroids of the two different persons are well-separated. It is also observed that a low degree of scattering exists among the features around their corresponding centroids. Hence, the dominant features extracted locally within a high-informative horizontal band of a face image offer not only a high degree of between-class separability but also satisfactory within-class compactness.

3.4 Reduction of Feature Dimension

Principal component analysis (PCA) is an efficient orthogonal linear transform to reduce the feature dimension [12]. In the proposed method, considering the magnitudes of the dominant DCT coefficients as features results in a feature space with large dimension. Thus, implementation of PCA on the derived feature space could efficiently reduce the feature dimension without losing much information. Hence, PCA is employed to reduce the dimension of the proposed feature space.

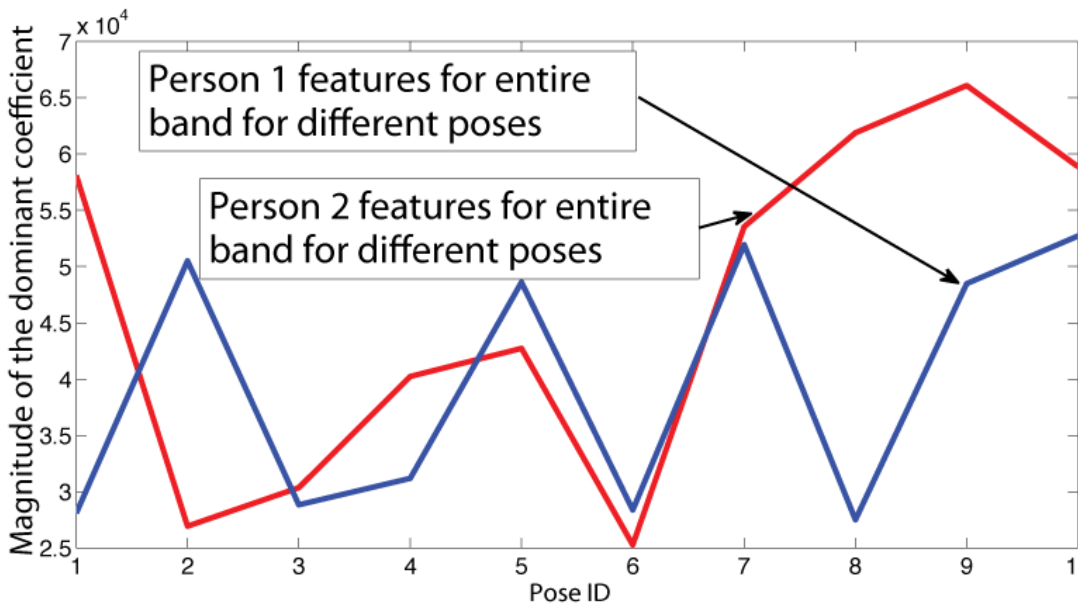

FIGURE 8: Variation of dominant magnitude features with poses using entire horizontal band

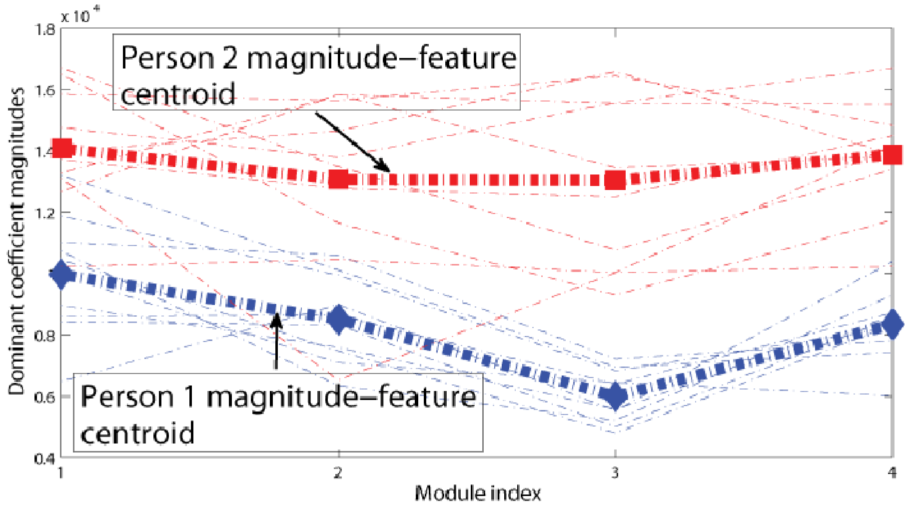

FIGURE 9: Variation of magnitude features with vertical segments of horizontal bands for several poses

3.5 Distance Based Face Recognition

In the proposed method, for the purpose of recognition using the extracted dominant features, a distance-based similarity measure is utilized. The recognition task is carried out based on the distances of the feature vectors of the training face images from the feature vector of the test image. Given the m-dimensional feature vector for the k-th pose of the j-th person be $\{\gamma_{jk}(1), \gamma_{jk}(2), \ldots, \gamma_{jk}(m)\}$ and a test face image f with a feature vector $\{v_f(1), v_f(2), \ldots, v_f(m)\}$, a similarity measure between the test image f of the unknown person and the sample images of the j-th person, namely *average sum-squares distance,* Δ, is defined as

$$\Delta_j^f = \frac{1}{q}\sum_{k=1}^{q}\sum_{i=1}^{m} |\gamma_{jk}(i) - v_f(i)|^2, \tag{3}$$

where a particular class represents a person with q number of poses. Therefore, according to (3), given the test face image f, the unknown person is classified as the person j among the p number of classes when

$$\Delta_j^f \leq \Delta_g^f, \forall j \neq g \, and \, \forall g \varepsilon \{1,2,\ldots,p\} \tag{4}$$

4. EXPERIMENTAL RESULTS

Extensive simulations are carried out in order to demonstrate the performance of the proposed feature extraction algorithm for face recognition. In this regard, different well-known face databases have been considered, which consist a range of different face images varying in facial expressions, lighting effects and presence/absence of accessories. The performance of the proposed method in terms of recognition accuracy is obtained and compared with that of some recent methods [13, 14].

4.1 Face Databases

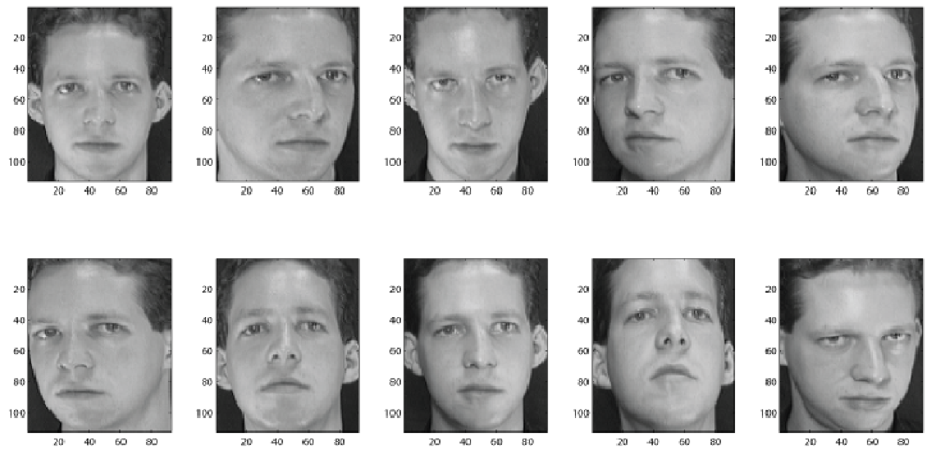

FIGURE 10: Sample poses of a person from the ORL database

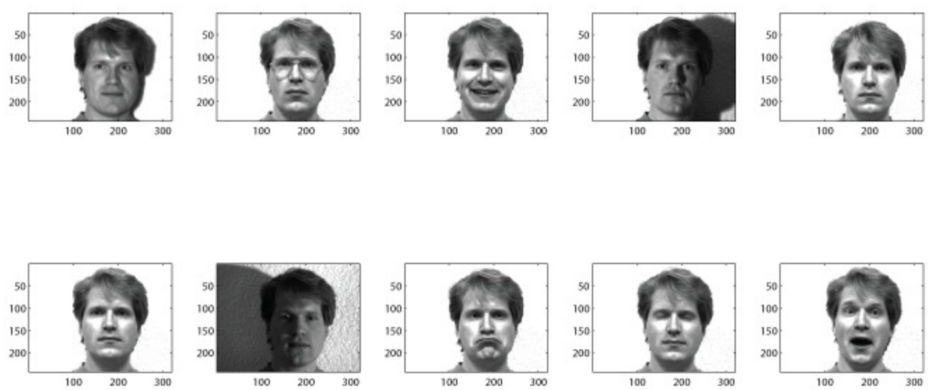

FIGURE 11: Sample poses of a person from the Yale database

In this section, the performance of the proposed face recognition scheme has been presented for two standard face databases, namely, the ORL database (available at http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/) and the Yale database (available at http://cvc.yale.edu/projects/yalefaces/yalefaces.html). In Figs. 10 and 11, sample face images of different poses of two different persons taken from the ORL and the Yale databases, respectively, are shown. The ORL database contains a total of 400 images of 40 persons, each person having 10 different poses. Little variation of illumination, slightly different facial expressions and details are present in the face images. The Yale database, on the other hand, consists a total of 165 images of 15 persons, each person having 11 different poses. The poses exhibit large variations in illumination (such as central lighting, left lighting and right lighting, dark condition), facial expressions (such as wink, sad, happy, surprised, sleepy and normal) and other accessories (such as with glasses and without glass).

4.2 Performance Comparison

In the proposed method, dominant spectral features (magnitudes and frequencies) obtained from all the modules of high-informative horizontal bands of a face image are used to form the feature vector of that image and feature dimension reduction is performed using PCA. The recognition task is carried out using a simple Euclidean distance based classifier as described in Section 3.5. The experiments were performed following the leave-one-out cross validation rule.

For simulation purposes, N number of horizontal bands are selected based on the entropy measure described in Section 3.1 and divided further into small modules. Module height is the same as that of the horizontal band and module width is chosen based on the face image width. In our simulations, $N = 2$ for the ORL database and $N = 3$ for the Yale database are chosen and the module sizes are chosen as 28×23 pixels and 27×20 pixels, respectively.

In order to show the effectiveness of the proposed local dominant feature extraction scheme, where each modules within the high-informative horizontal bands are considered separately, the recognition task is also carried out by considering the entire horizontal bands as a whole using the same feature extraction algorithm. We refer to the later scheme as *Proposed Scheme Without Modularization* (PSWOM) method. For the purpose of comparison, recognition accuracies obtained using the proposed method (*Proposed Scheme With Modularization* or PSWM) along with those obtained by the PSWOM method, and methods reported in [13] and [14] are listed in Table 1. Here, in case of the ORL database, the recognition accuracy for the method in [14] is denoted as not available (N/A). It is evident from the table that the recognition accuracy of the proposed method is comparatively higher than those obtained by the other methods for both the databases. It indicates the robustness of the proposed method against partial occlusions, expressions and nonlinear lighting variations. It is to be noted that the recognition accuracy is drastically reduced for the PSWOM method, where unlike the proposed method, feature extraction is carried out without dividing the horizontal bands into modules.

Method	Yale Database	ORL Database
PSWM	99.20%	99.65%
PSWOM	87.27%	65.75%
Method [13]	98.18%	99.00%
Method [14]	97.70%	N/A

TABLE 1: Comparison of recognition accuracies.

5. CONCLUSIONS

A spectral feature extraction algorithm based on 2D-DFT is proposed for face recognition. Instead of using the whole face image for feature extraction, first, certain high-informative horizontal bands within the image are selected using the proposed entropy based measure. In order to capture spatial information of face images locally, modularization of the horizontal bands is performed. The dominant spectral features are then extracted from the smaller modules within those horizontal bands using 2D-DFT. It has been found that the proposed feature extraction scheme offers an advantage of precise capturing of local variations in the face images, which plays an important role in discriminating different faces. Moreover, it utilizes a very low dimensional feature space, which ensures lower computational burden. For the task of classification, an Euclidean distance based classifier has been employed and it is found that, because of the quality of the extracted features, such a simple classifier can provide a very satisfactory recognition performance and there is no need to employ any complicated classifier. From our extensive simulations on different standard face databases, it has been found that the proposed method provides high recognition accuracy even for images affected due to partial occlusions, expressions and nonlinear lighting variations.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude towards the authorities of the Department of Electrical and Electronic Engineering and Bangladesh University of Engineering and Technology (BUET) for providing constant support throughout this research work.

6. REFERENCES

[1] A. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” IEEE Trans. Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4 – 20, 2004.

[2] C.Villegas-Quezada and J. Climent, “Holistic face recognition using multivariate approximation, genetic algorithms and adaboost classifier: Preliminary results,” World Academy of Science, Engineering and Technology, vol. 44, pp. 802–806, 2008.

[3] L. L. Shen and L. Bai, “Gabor feature based face recognition using kernal methods,” in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, vol. 6, 2004, pp. 386–389.

[4] M. Zhou and H. Wei, “Face verification using gabor wavelets and adaboost,”in Proc. Int. Conf. Pattern Recognition, vol. 1, 2006, pp. 404–407.

[5] C. BenAbdelkader and P. Griffin, “A local region-based approach to gender classification from face images,” in Proc. IEEE Comp. Society Conf. Computer Vision and Pattern Recognition, vol. 3, 2005, pp. 52–57.

[6] R. Gottumukkal and V. K. Asari, “An improved face recognition technique based on modular PCA approach,” Pattern Recognition Lett., vol. 25, pp. 429–436, 2004.

[7] S. Alirezaee, H. Aghaeinia, K. Faez, and F. Askari, “An efficient algorithm for face localization,” Int. Journal of Information Technology, vol. 12, pp. 30–36, 2006.

[8] E. Loutas, I. Pitas, and C. Nikou, “Probabilistic multiple face detection and tracking using entropy measures,” IEEE Trans. Circuits and Systems for Video Technology, vol. 14, pp. 128–135, 2004.

[9] S. C. Dakin and R. J. Watt, “Biological ‘bar codes’ in human faces,” World Academy of Science, Engineering and Technology, vol. 9, pp. 1–10, 2009.

[10] X. Zhang and Y. Gao, “Face recognition across pose: A review,” Pattern Recogn., vol. 42, pp. 2876–2896, 2009.

[11] R. C. Gonzalez and R. E. Woods, Digital Image Processing. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1992.

[12] I. Jolloffe, “Principal component analysis,” Springer-Verlag, Berlin, 1986.

[13] F. M. deS. Matos, L.V. Batista, and J.v.d. Poel, “Face recognition using DCT coefficients selection,” in Proc. ACM symp. Applied computing, 2008, pp. 1753–1757.

[14] X.Y. Jing and D. Zhang, “A face and palm print recognition approach based on discriminant DCT feature extraction,” IEEE Trans. Systems, Man, and Cybernetics, vol. 34, pp. 2405–2415, 2004.