FEATURE IMAGE GENERATION USING LOW, MID AND HIGH FREQUENCY REGIONS FOR FACE RECOGNITION

Vikas Maheshkar[1], Sushila Kamble[2], Suneeta Agarwal[3] and Vinay Kumar Srivastava[4]

[1-3]Department of Computer Science and Engineering, MNNIT, Allahabad
v_mahehskar@yahoo.com , sushila@mnnit.ac.in , suneeta@mnnit.ac.in
[4]Department of Electronics & Communication Engineering, MNNIT, Allahabad
vinay@mnnit.ac.in

ABSTRACT

In this paper, we propose a generation of feature image for face recognition. Feature image is generated using Discrete Cosine Transform (DCT), which tries to best classify different face images by maximizing the individual difference between face images. Low, Mid, and High frequency components of DCT are used to hold different information of face by providing a local description of different facial components such as detailed smooth region description and, effective edge representation. Reduction in different frequency components is achieved using statistical measure such as standard deviation. A comparison of proposed work with other standard methods is done on ORL face datasets. An experimental result shows a significant improvement of the proposed scheme.

KEYWORDS

Discrete Cosine Transform, Low, Mid, and High frequency components, Mahalanobis distance

1. INTRODUCTION

Face recognition (or identification) is a very active research area nowadays due to its importance in both human computer and social interaction which recognizes an individual by matching input face image against face images stored and finds the best match, either to identify a person or verify a person's claimed identity. The general problem of face recognition remains to be unsolved, since most of the systems to date can only successfully recognize faces when images are obtained under prescribed conditions [9]. Their performance will degrade abruptly when face images are captured under varying pose, lighting, with accessories and expression [13]. Another challenging problem is the number of training images used for face recognition.

A face recognition system developed should be strong enough in account of all the above-mentioned problems and cope with them in an effective manner. In order to do so, an efficient and effective representation for face should be generated. Many techniques have been proposed in the literature for representing face images. Principal Components Analysis (PCA) reduce high dimensional data, and is one of the most frequently used technique. D.Swets and J.Weng proposed discrimant eigenfeatures for image retrieva[3]l. Mathematical transforms such as Discrete Fourier Transform (DFT), Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT) has been widely used to generate facial feature image. Weilong Chen, Meng Joo Er, and Shiqian Wu, [11] use DCT to generate feature image by using PCA and Linear Discriminant Analysis. In particular, many data compression techniques employ the DCT, which

DOI : 10.5121/ijma.2012.4107

has been found to be asymptotically equivalent to the optimal Karhunen-Loeve Transform (KLT) for signal decorrelation. In DCT, coefficients with large magnitude are mainly located in the upper-left corner of the DCT matrix. They are the low spatial frequency DCT components in the image. M. J. Er, W. Chen, and S. Wu.[7] states that the DC or the first three low frequency coefficients have been truncated in order to decrease the effects of the illumination variation. C. Sanderson and K. K. Paliwal [1] shows that the polynomial coefficients are derived from the 2D-DCT coefficients obtained from the spatially neighboring blocks. Jiang and Feng [6] showed that removal of DC element enables the reconstructed facial image to be robust to lighting changes and removal of high-frequency DCT coefficients to be robust to scaling variations. X.Y. Jing and D. Zhang [12] proposed an approach to find discriminant bands (a group of coefficients) in the transformed space. Their approach searches the discriminant coefficients in the transformed space group by group. In this case, it is possible to lose a discriminant coefficient placing beside the non- discriminant coefficients in a group as shown by Dabbaghchian, S., Aghagolzadeh, A., and Moin M.S.[2]. DCT obtains the optimal performance of PCA in facial information compression and the performance of DCT is superior to conventional transforms. Jean Choi et al. [5] proposed face recognition using energy probability to generate a frequency mask without attention to the class.

In this paper we propose a new approach to improve the classification for pose and expression invariant face recognition. This approach is based on local fusion of different frequency components using detailed magnitude of Discrete Cosine Transform (DCT). In order to capture better information of different facial regions, the information in different frequency regions is retained separately. Energy distribution is used to set the parameters for the selection of different frequency regions to construct feature image. Reduction in data in different frequency regions is done using statistical measures such as standard deviation.

The organization of the paper is as follows. In Section 2, the different methodologies used in the paper are DCT, energy distribution and Standard Deviation is described. The fusion of different frequency regions and proposed technique is explained in Section 3. Experimental results are presented and discussed in Section 4 followed by conclusion in Section 5.

2. METHODOLOGY

2.1. Discrete Cosine Transform(DCT)

DCT is a powerful transform in image processing applications include image coding and face recognition. DCT is very close to the KLT and has a strong ability for data decorrelation. Weilong Chen, Meng Joo Er, and Shiqian Wu, [11] prove that DCT is an orthogonal transformation. DCT expresses data as the sum of cosine function for reduced size of data. Under the assumption that gray scale matrix of face image as *f(x, y)* of size NxN, its DCT, *F(u, v)* of size NxN, is obtained by the following equation:

$$F(u,v) = \alpha(u)\alpha(v)\sum_{x=1}^{N}\sum_{y=1}^{N} f(x,y) * \cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

$$where, \alpha(u), \alpha(v) = \begin{cases} \sqrt{\frac{1}{N}}, & u,v = 1 \\ \sqrt{\frac{2}{N}}, & otherwise \end{cases}$$

2.2. Energy Distribution(ED)

Energy is one of the important characteristic of image defined by following equation

$$Energy_C = \sum_{u=1}^{N} \sum_{v=1}^{N} \left| F(u,v) \right|^2$$

Where, *F(u,v)* represents image in transform domain.

Energy Distribution *ED(u, v)* represents the energy contribution of each transformed coefficient and is given by following equation:

$$ED(u,v) = \frac{\left| F(u,v) \right|^2}{Energy_C}$$

The magnitude of *ED(u, v)* gives Energy Distribution matrix.

2.3. Standard Deviation

Standard deviation is given as

$$StdDeviation = \left(\frac{1}{N-1}\sum_{i=1}^{N}(f_k(i) - mean)^2\right)^{1/2}$$

Where mean is given by

$$mean = \frac{1}{N}\sum_{i=1}^{N} f_k(i)$$

Here, f_k is the feature, k is the feature index where k=1,2,3,..,N where N is the available features.

3. PROPOSED TECHNIQUE

In the proposed technique we consider two datasets viz. training datasets containing set of face images for extracting the relevant information and test dataset for the face image to be recognized. The face images in the training dataset are converted to transform domain (DCT, Slant or Walsh). The transformed coefficients after zigzag scan are divided equally into three bands, namely low frequency, middle frequency and high frequency and are represented by *TCvector*. This is shown diagrammatically in Figure 1.

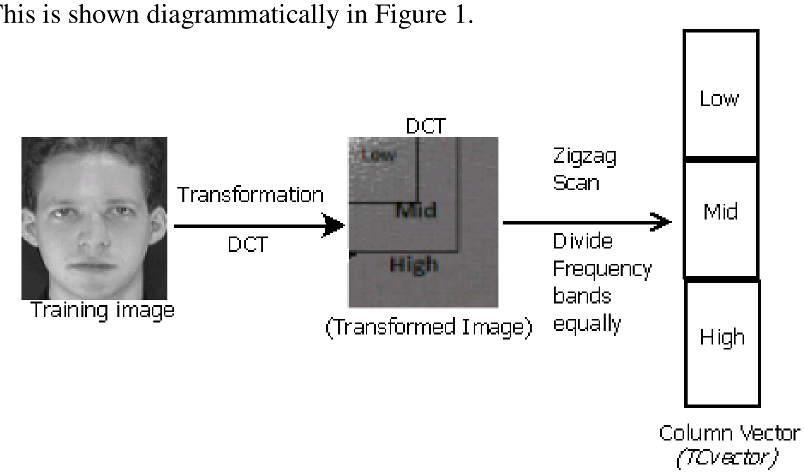

Figure 1. Formation of Column Vector (*TCvector*)

Low frequency coefficients are related to illumination variation and smooth regions (like forehead, cheeks etc) of face. High frequency coefficients represent noise and detailed information of edge. The middle frequency coefficients represents the basic structure of the of face image. It shows that each band contains important information. Hence we cannot ignore the low frequency components to compensate illumination variations, if the image is not so much affected by lighting conditions. Similarly we cannot truncate the high frequency coefficients to remove noise as they are responsible for details and edge information of the image.

Energy is used to describe a measure of "information" in an image. *E*nergy *D*istribution represented by *ED(u, v)* as described in section 2.2 gives the contribution of individual transformed coefficients. Let M be the total number of facial images of training dataset and N as width and height of all images. The *ED*, of size NxN, is converted into the column vector by performing Zigzag scan of length N*N represented by *EDvector*. The *EDvector* is divided into three equal regions similar to the *TCvector*.

The high value of *ED(u, v)* means more valid information. To achieves reduction in data while retaining the important features we set different thresholds in three regions of *EDvector* using statistical measure such as Standard Deviation (Standard Deviation1, Standard Deviation2, Standard Deviation3). Energy distributions below the different threshold set by three standard deviations in different regions of *EDvector* are set to zero.

$$Featureimage(u,v) = \begin{cases} 0 & , \quad if \;\; ED(u,v) < STD1, STD2, STD3 \\ Featureimage(u,v) & , \quad otherwise \end{cases}$$

Finally we generate Feature image represented as *Featureimage(u,v)* by retaining corresponding coefficients of *TCvector* that we retained in the *EDvector*. This is shown diagrammatically in Figure 2.

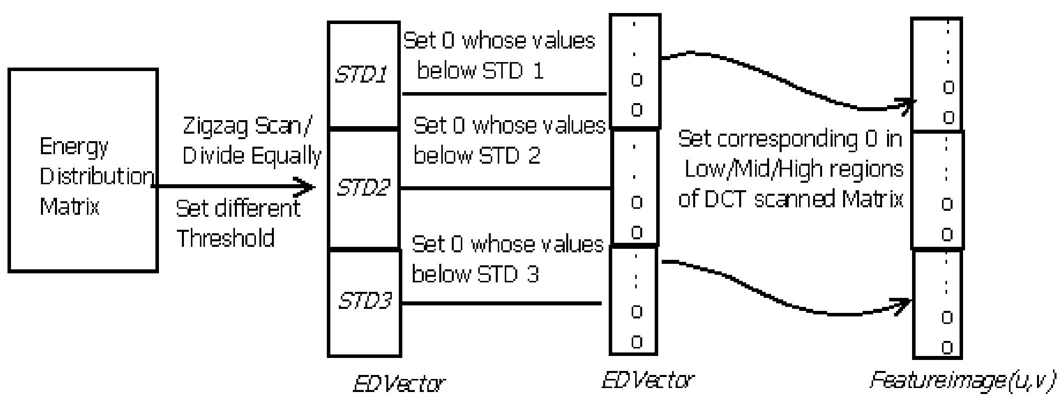

Figure 2. Feature Image Generation

Figure 2 depicts the formation of Feature Image Generation for face recognition. Face images in the training set are converted into Transform domain using DCT. The Energy distribution is defined as criterion of selecting effective facial features. Energy distribution is calculated for the dimension reduction of data and optimization of valid information in three regions is done by calculating three standard deviations. The generated image is called as *Featureimage(u,v)*. Similar procedure is carried out for test image to form *TFeatureimage(u,v)*. Mahalanobis distance between set of *Featureimage(u,v)* and the test image formed as *TFeatureimage(u,v)* is calculated. The *Featureimage(u,v)* which gives minimum distance from the set of training images and test image will give recognized face.

4. EXPERIMENTAL RESULTS

To validate the results we implement the proposed technique in MATLAB 2010. The experimentation is carried out on standard ORL datasets. The images in the ORL training dataset were taken between April 1992 and April 1994 at the AT&T Laboratories Cambridge lab [4]. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses).

4.1. Experimental analysis based on ORL face dataset

The technique is tested on wide set of images taken from ORL datasets. We implemented our proposed technique using datasets of 9 poses per person for training and remaining different one pose for testing. To prove the robustness of the technique, we also implement the technique using only two poses of each individual (out of 10 poses of each individual in the training dataset) i.e., 80 images of 40 individuals are considered. Test dataset contains poses that are different than the training dataset.

The set of features are generated using training dataset. The different test images are given as input to check the recognition rate. Recognition is done on the generated features by Euclidean and Mahalanobis distance. Figure 3 shows test image and its correctly recognized image (2 pose per person). All the input test images correctly recognizes the face from the training dataset even if there is change in emotions like- neutral, smile, laughter, sad/disgust and poses like looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, and looking down.

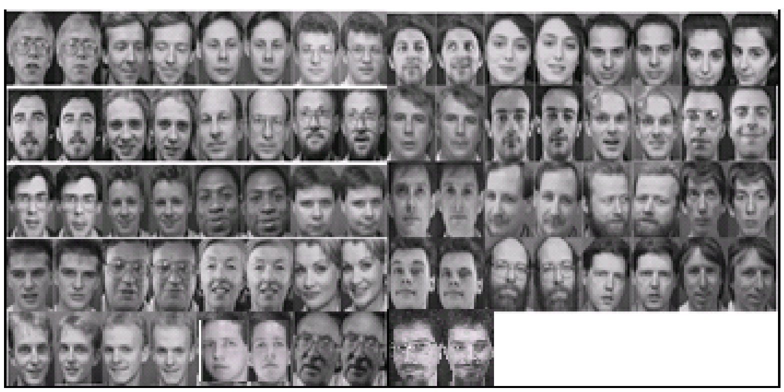

Figure 3. Test image and corresponding recognized faces of ORL dataset (2 pose per person)

Table 1 summarizes the face recognition rate of proposed technique with the correctly recognized image using standard techniques. The experimentation is carried out on dataset considering 9 Poses per Person i.e., 360 images and 2 Poses per Person i.e., 80 images in the training dataset and different remaining pose in the test dataset. The proposed technique is also compared with standard PCA as introduced by M. Turk and A. Pentland [8] and DCT normalization technique as used by Štruc, V. and Pavešić, N [10]**.** Table.1 shows that the proposed technique has high recognition rate when compared to other techniques. The recognition rate is encouraging as 37 correct for 2 pose per person and 100% recognition for 9 pose per person i.e., all 40 test faces were correct.

Table 1: Face recognition rate using ORL dataset

Technique	Training Images	Test Images	Correctly Recognized Faces
Eigenface technique(PCA), DCT Normalization	80	40	**33**
Proposed technique (Euclidean distance)	80	40	**35**
Proposed technique (Mahalanobis distance)	80	40	**36**
Proposed technique (Mahalanobis distance)	360	40	**40**

Figure 4 shows the comparison plot of different face recognition technique in terms of number of correctly recognized faces (considering 80 training images).

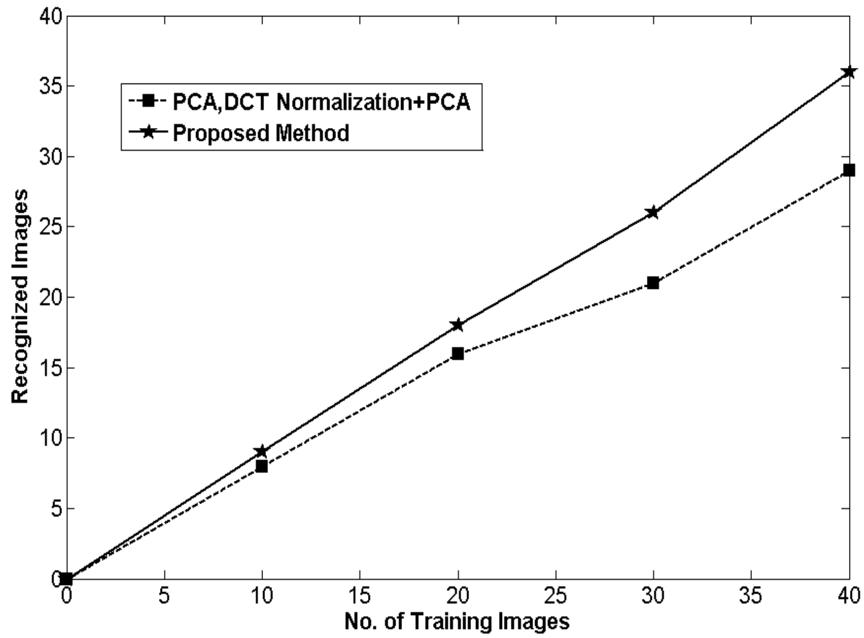

Figure 4. Plot of No. of training images tested Vs. recognition rate (ORL Datasets)

5. CONCLUSION

The generated feature image is robust to orientation, expression, illumination, pose and angle. Compared to the traditional methods, the proposed method not only performs better, but also inherits low computational complexity, as recognition is done directly i.e., without applying Principal Component Analysis approach. Recognition is done directly on the generated feature image using distance measure such as Mahalanobis. The results of face recognition using this approach are very good and encouraging. Furthermore, only two poses per person is used for training which makes it useful for practical face recognition applications.

REFERENCES

[1] C. Sanderson and K. K. Paliwal (2003), "Features for Robust Face-Based Identity Verification", Journal *of Signal Processing*, Vol. 83, pp. 931-940.

[2] Dabbaghchian, S., Aghagolzadeh, A., Moin, M.S.,(2008), "Reducing the effects of small sample size in DCT domain for face recognition", Telecommunications, 2008. IST 2008. International Symposium on , vol., no., pp.634-638.

[3] D. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions. PAMI,* 18(8):831-836.1996.

[4] Ferdinando Samaria, Andy Harter(1994), "Parameterisation of a Stochastic Model for Human Face Identification", Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL.

[5] Jean Choi, Yun-Su Chung, Ki-Hyun Kim, Jang-Hee Yoo, (2006), "Face Recognition using Energy Probability in DCT Domain", *Multimedia and Expo, 2006 IEEE International Conference on*, pp.1549-1552.

[6] Jiang, J., Feng, G. , (2007), "Robustness analysis on facial image description in DCT domain", *Electronics Letters* , vol.43, no.24, pp.1354-1356.

[7] M. J. Er, W. Chen, and S. Wu.,(2005), "High Speed Face Recognition Based on Discrete Cosine Transform and RBF Neural Networks", *IEEE Transactions on Neural Networks*, Vol. 16, No. 3, pp. 679-691.

[8] M. Turk, A. Pentland (1991), "Eigenfaces for Recognition", Journal of Cognitive Neurosicence, Vol. 3, No. 1, pp. 71-86.

[9] Ramanathan N, Chellappa R, Roy Chowdhury, A.K, (2004), "Facial similarity across age, disguise, illumination and pose", *In Proceedings of the International Conference on Image Processing ICIP'04*, 1999 - 2002.

[10] Štruc, V. and Pavešić, N(2009), "Illumination Invariant Face Recognition by Non-Local Smoothing", Proceedings of the international Cost 2101& 2102 conference BioID_MultiComm, pp. 1-8.

[11] Weilong Chen, Meng Joo Er, and Shiqian Wu, (2005), "PCA and LDA in DCT domain", *Pattern Recognition Letters*, vol. 26, pp. 2474–2482.

[12] X.Y. Jing, D. Zhang, (2004), "A face and palmprint recognition approach based on discriminant DCT feature extraction", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 34 (6), pp.2405–2415.

[13] Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face Recognition: A Literature Survey. Technical Report CAR-TR-948. Univ. of Maryland, CfA.(2000)

Authors

Vikas Maheshkar received B.E. in Computer Technology from MIET Gondia, Maharashtra, India in 2002, M. Tech in Computer Science & Engineering from SATI Vidisha, MP, India in 2007 and presently pursuing Ph.D from Motilal Nehru National Institute of Technology, Allahabad, UP, India. He is having 7 years of teaching experience. His current research interest includes Pattern Recognition, Computer Vision, Algorithms, Compression, Biometrics, Face recognition, Digital watermarking.

Sushila Kamble received B.E. in Computer Technology from MIET Gondia, Maharashtra, India in 2003, M. Tech in Computer Science & Engineering from SATI Vidisha, MP, India in 2007 and presently pursuing Ph.D from Motilal Nehru National Institute of Technology, Allahabad, UP, India. She is having 4 years of teaching experience. Her current research interest includes Digital watermarking, Pattern Recognition, Computer Vision, Algorithms, Compression, Biometrics, Face recognition.

Suneeta Agarwal received B. Sc degree in 1973 from university of Allahabad, M.Sc degree in 1975 from university of Allahabad, Ph. D in 1980 from IIT Kanpur and M. Tech degree in 2007 from AAIDU. She is having 31 years of Teaching Experience and currently professor in the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology, Allahabad. Her current research interest includes Pattern Recognition, Computer Vision, Theory of Computation Science, Algorithms, Automata Theory, Compression, Patten matching, Finger print recognition.

Vinay Kumar Srivastava received the BE in ETC from GEC Rewa, MP, India in 1989, the M Tech in Communication from IT-BHU, Varanasi, India in 1991 and PhD in Electrical Engineering from IIT Kanpur, India in 2001. Presently he is a Professor in the department of ECE, MNNIT, Allahabad, India. He has about 18 years of teaching and research experience in the area of Signal and image processing. He has chaired many sessions in conferences. He has authored or coauthered around 20 publications. His current research interest includes image compression, digital watermarking, stability 2D PSV system, DSP methods for identification of protein-coding regions, design and analysis of IDMA systems.