Local Spectrum Feature Extraction for Face Recognition

Muhammad Imran Ahmad, Ruzelita Ngadiran, Mohd Nazrin Md Isa, Nor Ashidi Mat Isa, Mohd Zaizu Ilyas, Raja Abdullah Raja Ahmad, Said Amirul Anwar Ab Hamid, Muzammil Jusoh

Abstract—This paper presents two techniques, local feature extraction using image spectrum and low frequency spectrum modelling using GMM to capture the underlying statistical information to improve the performance of face recognition system. Local spectrum features are extracted using overlap sub block window that are mapped on the face image. For each of this block, spatial domain is transformed to frequency domain using DFT. A low frequency coefficient is preserved by discarding high frequency coefficients by applying rectangular mask on the spectrum of the facial image. Low frequency information is non-Gaussian in the feature space and by using combination of several Gaussian functions that has different statistical properties, the best feature representation can be modelled using probability density function. The recognition process is performed using maximum likelihood value computed using pre-calculated GMM components. The method is tested using FERET datasets and is able to achieved 92% recognition rates.

Keywords—Local features modelling, face recognition system, Gaussian mixture models.

I. INTRODUCTION

FACE recognition is an active research area and widely used in several applications such as security system, video surveillance, identity authentication and human computer interaction. Currently, various different algorithms are developed that are able to properly recognize human based on their face image. A superior performance is obtained with sufficient amount of training data covering a wide range of variations in pose and illumination. This research is fast developing due to the extensive research focus in the computer vision and pattern recognition. Several new algorithms which are effective and able to achieve a good performance are being studied [1], [2]. Face recognition framework has two main process which is training and recognition stages. Both of these stages involve pre-processing, feature extraction, feature selection, creation of template data and classification. Among these stages, feature extraction will most important and affect the performance of the whole system. A proper feature extraction scheme must be design in order to produce a high discrimination feature space.

M. I. Ahmad, R. Ngadiran, S.A.A. Ab Hamid, M. Jusoh, R. A. Raja Ahmad, M. Z. Ilyas are with the *ENAC* Research Cluster, School of Computer and Communication Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia (e-mail: m.imran@unimap.edu.my).

M. N. Md Isa is with the School of Microelectronic Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia.

N. A. Mat Isa is with the School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia.

Feature extraction in face recognition problem falls in two categories: appearance-based and feature-based, depending on the way the face image is processed. The appearance based method involves encoding the whole facial image and treating the resulting facial new representation as a point in a high dimensional space. Due to a high dimensional feature space, a new vector cannot be compared directly. This approach must use some kind of dimensionality reduction techniques to derive lower-dimensional vectors for subsequent classification. An example of linear projection method is Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). In PCA based technique, the low dimensional face image is called eigenface [3], where the high dimensional vectors are projected onto the subspace defined by the eigenvector that has the highest eigenvalues. LDA technique is a supervised approach also called Fisherface [4], use linear projection based on Fisher's linear discriminant objective to find a new subspace where vectors of the same class are close to each other and at the same time for from to the other classes. A new feature space of PCA and LDA is then used for classification using a simple technique such as Euclidean distance classifier [5]; Mahalanobis distance classifier and support vector machine [6]. The method above is optimum for a linear feature distribution. In the case of non-linear data distribution, a method such as Locality Preserving Projection [7], Kernel PCA [8], [9] and Generalised Discriminant Analysis (GDA) can be used effectively. Among the technique listed above, nonlinear technique based on kernel method is the best for the face recognition task [10]. A subspace based features usually require a large amount of training data to properly learn the variations exist in the face image.

Feature extraction using feature-based technique normally use a set of local annotations obtained from the face image to generate a model for each class, which is used for classification. One of the most well-known systems is Elastic Bunch Graph Matching (EBGM) where a face image is represented by a set of wavelets coefficients arranged in a graph. In the recognition process, the mesh is allowed to be deformable in order to maximize the correlation between the gallery and the probe image. Another feature based approach is by using Local Binary Patterns (LBPs) [11], [12], where the information of each sub block window of face image is represented by a histogram of LBP.

Recognition is performed based on the similarity measure between the histograms. Feature based features are widely used in the statistical modelling method such as Gaussian

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:9, No:1, 2015

Mixture Model (GMMs) [13] and Hidden Markov Model (HMM) [14]-[16]. This approach decomposes the face image into several sub block and model the distribution of each block using the above model. Feature based system has several advantages over holistic approach: they are more robust to variations and pose, illumination and expression.
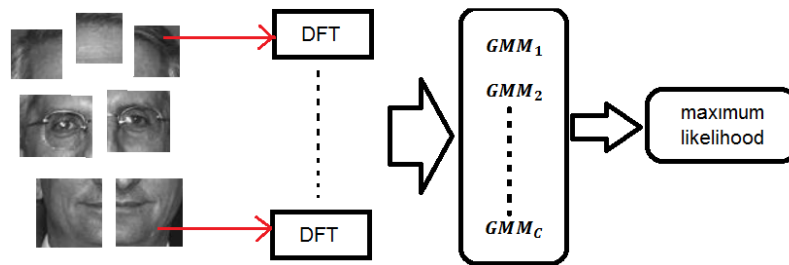

Fig. 1 Block diagram of the propose local spectrum modelling and classification

In this paper, we propose to extract local information by using sub block window mapping on the face image. Each image in spatial domain is transformed to a frequency domain where some image information can be removed to reduce redundant features and producing low dimensional feature vector for each blocks. Only low frequency information is preserve to model a feature distribution of each class. In this paper, GMM is used to capture non Gaussian feature distribution. The proposed method is shown in Fig. 1.

II. LOCAL FEATURE EXTRACTION TECHNIQUE

Feature extraction is a process to reduce number of coefficients required to describe the face image accurately. The face image has low frequency information which is very important for recognition and high frequency information which do not contribute effectively to the recognition process. Transformation of face images to frequency domain and then choose low frequency components can be used to reduce image information redundancy because only several number of coefficients are necessary to preserve and important for recognition. In this paper we proposed to use DFT to extract low frequency information exist in each local region. Face image are divided by several number of local region as shown in Fig. 1. DFT transformation is computed in each of these regions. DFT is a powerful tool used in frequency analysis. Its inherent properties like translation invariance, and symmetry, play a major role in the proposed extraction method. It is observed that the low frequency components of an image contain most of the information that is required to represent that particular image and has high discrimination power. High frequency components which contribute towards finer details are less significant for recognition. When DFT is computed on the block of local region, low frequency information is concentrated at the top corners and then shifted to the center of the spectrum. The first quadrant is swaps with third and the second quadrant with the forth as shown in Fig. 2. In order to get a clear visualization, low frequency coefficients are shifted to the center, then a DFT coefficient can be preserved using a masking method in order to obtain the best feature set. Since all the low frequency components are exists at the center of spectrum, the center region is extracted using rectangular mask. Different size of mask will give different amount of information.
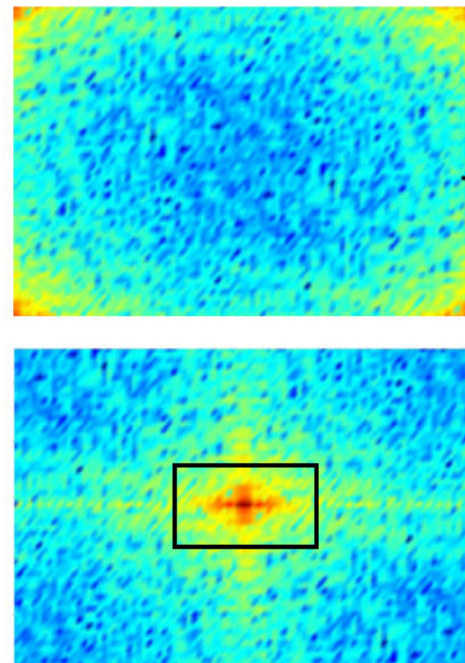

Fig. 2 DFT image before and after centralize

A. Local Feature Learning

In general any supervised classifier method can be used to recognise or classify the proposed compact feature representation into a specific class by using a given test feature vector. The use of parametric statistical methods is preferable due to its ability to provide the likelihood score value of the observed features which could be important for the end processing task such as likelihood score normalization and maximum likelihood classification process. GMM can be used to estimate the true pdf when modelling multimodal fused feature vector distribution, whose values are generated by one of several randomly occurring independent sources. GMM probability density function can be defined as weighted summation of Gaussian function given as:

$$p(x; \theta) = \sum_{c=1}^{C} \alpha_c N(x; \mu_c, \Sigma_c) \qquad (1)$$

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:9, No:1, 2015

where α_c is the weight of the *c-th* component. The weight can be interpreted as a priori probability that a value of the random variable is generated by *c-th* source and the value is subject to $0 \le \alpha_c \le 1$ and $\sum_{c=1}^{C} \alpha_c = 1$. $N(x; \mu_c, \Sigma_c)$ is the normal distribution given by

$$N(x; \mu_c, \Sigma_c) = \frac{1}{(2\pi)|\Sigma_i|} e^{-\frac{1}{2}(x-\mu)^T \Sigma_i^{-1}(x-\mu)} \qquad (2)$$

where $\sum$ and μ denote the covariance matrix and mean of feature vector x. A complete Gaussian mixture density model is parameterized by mean vector, covariance matrices and mixture weights from all components densities. These parameters can be completely defined by a parameter list as follows:

$$\theta = \{\alpha_1, \mu_1, \Sigma_1, \ldots, \alpha_c, \mu_c, \Sigma_c\}$$

The aim of the training process is to estimate the parameter of GMM, which best match the distribution of the training feature vectors. One popular method used to estimate model parameters that maximize the likelihood of the GMM based on a given training data is maximum likelihood (ML). The GMM likelihood function for a sequence of N independent training features $X = \{x_1, x_2, \ldots, x_N\}$ is given by

$$L(X; \theta) = \prod_{n=1}^{N} p(x_n; \theta) \qquad (3)$$

The aim of ML is to find $\hat{\theta}$ that maximize the likelihoods as follows:

$$\hat{\theta} = \arg\max_{\theta} L(X; \theta) \qquad (4)$$

Direct maximization of (4) is not possible due to the nonlinearity of function of parameters θ. In practice, an iterative method such as the expectation maximization (EM) algorithm is used for calculating maximum likelihood distribution parameter estimates from incomplete data. The algorithm can also be used to handle cases when an analytical approach for ML estimation is not feasible, such as Gaussian mixtures with unknown covariance matrices and means. The main idea of EM algorithm is starting off with an initial model parameter θ^{init}, and a new model parameter θ^{new} is estimated, in such that the iteration will stop after achieving $p(X; \theta^{new}) \ge p(X; \theta^{init})$. The new model will become the initial model for the next iteration and the process is repeated until some convergence threshold is reached. Re-estimation formulas for weight, means and covariance in EM iterations are given as:

$$\alpha_i = \frac{1}{T} \sum_{n=1}^{N} p(i|x_n, \theta) \qquad (5)$$

$$\mu_i = \frac{\sum_{n=1}^{N} p(i|x_n, \theta) x_n}{\sum_{n=1}^{N} p(i|x_n, \theta)} \qquad (6)$$

$$\sigma_i^2 = \frac{\sum_{n=1}^{N} p(i|x_n, \theta) x_n^2}{\sum_{n=1}^{N} p(i|x_n, \theta)} - \mu_i^2 \qquad (7)$$

where $p(i|x_n, \theta)$ is a posteriori probability for class i and is given as:

$$p(i|x_n, \theta) = \frac{\alpha_i N(x; \mu_c, \Sigma_c)}{\sum_{m=1}^{M} \alpha_m N(x; \mu_c, \Sigma_c)} \qquad (8)$$

During the classification process, the aim is to assign a class of test feature vectors to the model parameter that has the highest value of log-likelihood. Classification process in GMM can be seen as the maximum likelihood rule (ML), where the given test feature vector will produce the highest values of likelihood. Let a group of S feature vectors belong to S class given by $X = \{X_1, X_2, \ldots X_s\}$ and each of the X_i has a non-stationary fused feature vector given by $x_t = \{x_1, x_2, \ldots, x_T\}$. The S class is represented by a Gaussian mixture model using S model parameters given by $\theta_1, \theta_2, \ldots, \theta_s$ where each model parameter $\theta_i = \{w_k, \mu_k, \Sigma_k\}$, and $k = 1, 2, \ldots M_c$ where M_c is the number of GMM components used to model the distribution. The objective of classification is to find the model θ_i which has the maximum a posteriori probability of a given observation feature vector X_i. This criterion can be written as:

$$\hat{S} = \arg\max_{1 \le k \le s} p(\theta_k | X) \qquad (9)$$

By using Bayes's rule, $p(\theta_k | X)$ can be written as:

$$\hat{S} = \arg\max_{1 \le k \le s} p(X | \theta_k) \qquad (10)$$

III. EXPERIMENTAL RESULT

The proposed method is tested using FERET face dataset. In this experiment 100 subjects in FERET dataset were randomly selected, each with six frontal images used for training and testing. The facial images in this datasets were captured under various levels of illumination, different facial expressions and poses ranging from angles of $\pm 15^o$, to $\pm 60^o$. Each image has 384x256 pixel size with 256 grayscales. Most of the images have different size of frontal image including the background and body chest region. The six images per subject used in this experiment are selected among the "fa", "fb", "rb" and "rc" which belong to different pose, aging, expression and pose angle. The image is randomly chosen for testing and training with different ratio of these two groups.

A. Analysis of Different Size of Local Region

In this analysis, we examine the performance of the recognition rate with different block size of local region. The size of local region is varied from 8x8 to 32x32 and each local region overlapped with different size. Three images are randomly chosen for training and testing. The recognition performance is shown in Fig. 3. The best recognition rate is 92% when the local region is extracted using 20x20 blocks with 50% pixel overlap. DFT transform is computed in this block and the coefficients are separated by low and high frequency information. Using smaller block will provide less information and using bigger block will not effective for GMM feature modelling. Bigger blog will produce high dimensional feature vector requiring many data points to properly estimate the probability density function. In this analysis, we found 50% overlap will give the best

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:9, No:1, 2015

performance. Overlapping window will integrate the information between different local regions and this approach is able to increase the discrimination power in the feature space.

B. Analysis Using Different GMM Components and Feature Size

In this experiment, the number of GMM components is examined in order to achieve the best result. The number of GMM components will affect the performance because a perfect model can only be achieved when enough Gaussian function is used to model data distribution. However, overfitting problem might occur if the number of GMM components is too much. In this analysis, we vary the GMM components between 3 and 20. Different size of rectangular window is also analysed. When rectangular window is large, higher number of GMM components is required to achieve the best result. The result is shown in Fig. 4.
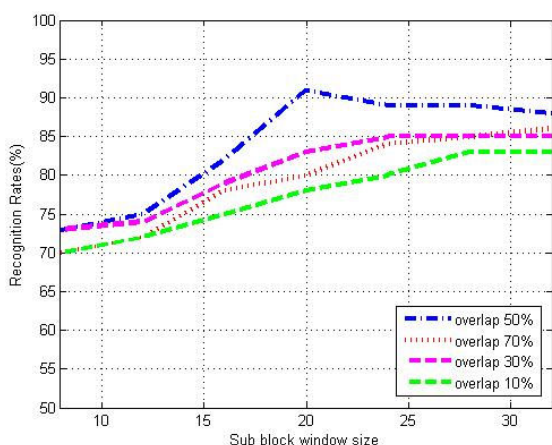

Fig. 3 Recognitions rates with different size of sub block window and different window overlap

TABLE I
ANALYSIS WITH DIFFERENT NUMBER OF TRAINING AND TESTING IMAGE

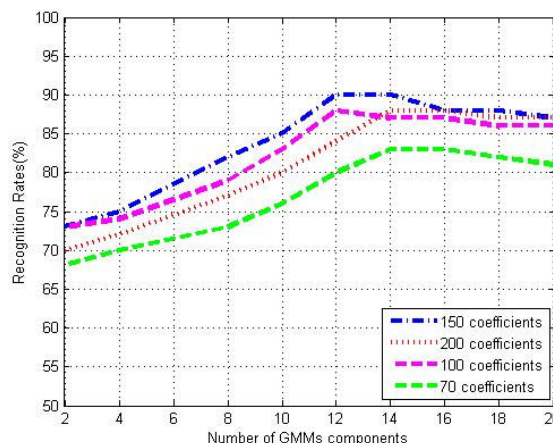| Ratio (train:test)	Holistic[17]	DCT local[18]	Propose DFT local
5:1	87%	90%	92%
4:2	84%	88%	91%
3:3	80%	85%	90%
2:4	75%	83%	88%
1:5	73%	80%	85%

Fig. 4 Recognition rates with different number of GMMs components

C. Analysis with Different Number of Training Image and Comparison with the Existing Method

In this analysis, different number of training image is used to examine the performance of the proposed method. The training image is varied from 1 to 5 where the images are randomly selected for training process. The proposed method is compared with the existing holistic and local feature face recognition. The result is as given in Table I. In this analysis, local features perform better than the holistic approach when low training image is used in the recognition process. The proposed method is compared with holistic approach that used Gabor and PCA for feature representation. The existing local features used in this table applied DCT in each local region of face image and GMM to model the feature distribution.

IV. CONCLUSION

This paper presents local feature extraction for face recognition using DFT. The spectrum of local features is extracted using several number of sub blocks windows with 50% pixel overlap. Independent feature from each sub block window has high discrimination power when they are overlapped with their neighbour. The low frequency information is preserved and modelled using GMM. GMM is able to capture the underlying statistical information that exists in the low frequency information. Using suitable number of GMM components able to construct the best model of non-Gaussian feature distribution in feature space. The best recognition rate is 92% when tested using FERET dataset.

REFERENCES

[1] G. Shakhnarovich, V. Moghaddam, Face recognition in subspaces, Chapter 2, in: Handbook of Face Recognition, Springer-Verlag London Limited, 2011, pp. 19–49.

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:9, No:1, 2015

[2] A. M. Patil, S. R. Kolhe, P. M. Patil, 2D face recognition techniques: a survey, International Journal of Machine Intelligence 2 (2010) 74–83.

[3] M. Turk, A. Pentland, Face recognition using Eigenfaces, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 1991, pp. 586–591.

[4] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

[5] M. I. Ahmad, W. L. Woo, S. S. Dlay, "Multimodal biometric fusion at feature level: Face and palmprint," Proc. of International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP), pp. 801-804, 2009.

[6] Bhat, V. S., Pujari, J. D.,"Face Recognition Using Holistic Based Approach". International Journal of Emerging Technology and Advanced Engineering, 4(7), pp. 134-141, 2014.

[7] X. He, S. Yan, Y. Hu, N. Partha, H.-J.Zhang, Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.

[8] M.-H. Yang, N. Ahuja, D. Kriegman. Face recognition using kernel eigenfaces, in: IEEE International Conference on Image Processing (ICIP), vol. 1, 2000, pp. 37–40.

[9] K.-I. Kim, K. Jung, H.-J.Kim, Face recognition using kernel principal component analysis, IEEE Signal Processing Letters 9 (2) (2002) 40–42.

[10] M.-H. Yang, Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods, in: IEEE International Conference on Automatic Face and Gesture Recognition (AFGR), 2002, pp. 205–211.

[11] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 2037–2041.

[12] Y. Rodriguez, S. Marcel, Face authentication using adapted local binary pattern histograms, in: European Conference on Computer Vision (ECCV), 2006, pp. 321–332.

[13] F. Cardinaux, C. Sanderson, S. Marcel, Comparison of MLP and GMM classifiers for face verification on XM2VTS, in: International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), Springer, Berlin, 2003

[14] F. Samaria, S. Young, HMM-based architecture for face identification, Image and Vision Computing 12 (8) (1994) 537–543.

[15] A. Nefian, M. Hayes, Hidden Markov models for face recognition, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, 1998, pp. 2721–2724.

[16] A. Martinez, Face image retrieval using HMMs, in: IEEE Workshop on Content- Based Access of Image and Video Libraries, 1999, pp. 35–39.

[17] Hongmei Li, Dongming Zhou, RencanNie, Analysis of Face Recognition Methods in Linear Subspace, Lecture Notes in Electrical Engineering Vol.269, 2013, pp 3045-3051

[18] S. Lucey and T. Chen, "A GMM parts based face representation for improved verification through relevance adaptation," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2, 2004, pp. II855-II861.