Binarized Eigenphases for Limited Memory Face Recognition Applications

Naser Zaeri[1], Farzin Mokhtarian[1]
Abdallah Cherri[2]

[1]Electronic and Electrical Engineering Dept., University of Surrey
GU2 7XH Guildford, United Kingdom
[2] Electrical Engineering Dept., Kuwait University
P.O.Box 5969 Safat, Kuwait
{n.zaeri,f.mokhtarian}@surrey.ac.uk
cherri@eng.kuniv.edu.kw

Abstract. Most of the algorithms proposed for face recognition involve considerable amount of calculations, and hence they can not be used on devices of limited memory constraints. In this paper, we propose a novel solution for efficient face recognition problem for the systems that utilize low memory devices. The new technique applies the principal component analysis to the binarized phase spectrum of the Fourier transform of the covariance matrix constructed from the MPEG-7 Fourier Feature Descriptor vectors of the images. The binarization step that is applied to the phases adds many interesting advantages to the system. It will be shown that the proposed technique maximizes the recognition rate while achieving substantial savings in computational time, when compared to other known systems.

Keywords: Face recognition, limited memory, PCA, and MPEG-7.

1 Introduction

In the last few years, researchers in the area of face recognition have proposed many numerous techniques that achieve high recognition rate. However, there are very little published works in literature that propose solution for devices of limited memory constraints [1-2]. One of the main challenges of face recognition is that there is a considerable amount of calculations and computations involved, and this will slow down the system significantly. For example, in the surveillance systems, where face recognition plays an important role for reliable security issues, we use wireless domain for data transmission and the low memory requirement of the data sent is of very important consideration. The typical face recognition algorithms that found great applications in different fields (banks and airports) may not be executable to applications that are memory-constrained, where the devices that are used have processor of no more than 100-300 MHz and hence they need special treatment and investigation.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

Most of the appearance-based methods for face recognition that deal with the face images as a whole, depend on calculating the eigenvalues and the eigenvectors of a system representing this face space [8-10]. The calculations related to (and consequently the time required for) finding the eigenvectors and eigenvalues of the covariance matrix of the system is relatively huge. In our case, we can not just apply the principal component analysis (PCA) (or even other known face recognition method such as linear component analysis (LDA)) to the face images *directly*.

In this paper, we propose a novel technique for face recognition for the systems that have a limited memory. The novelty of this technique comes from the following ideas:

a) The face space is not represented by the images of the database in use (as it is done usually); rather it is represented by the *Fourier transform* of the covariance matrix constructed from the MPEG-7 Fourier Feature Descriptor (FFD) vectors of these images as a first step. The reason behind using the MPEG-7 face descriptor is that you can end up with a compact vector of small memory size.

b) As Oppenheim *et al.* showed [3] that the phase angle of the Fourier transform retains most of the information about the image, we will represent our system with the *phases* of the Fourier transform of the covariance matrix constructed from the FFD vectors. The phase component acts to position the bright and dark spots in the image in order to form regions that are recognizable by a viewer. Similarly, the phase component retains the most important information about the system of FFD vectors in the covariance matrix.

c) The third idea of our system is the *Binarization* step. In digital image processing, we can find the important features of an image through enhancing the high frequency components of that image. We can look at our system as a matrix that has some elements of high frequency and others of low frequency. The enhancing process of these high frequency components can be done by sharpening the system. This operation can be done by binarization process, as will be shown in Section 3.

d) Now, we seek the transformation that best describe the space in a way that helps us in further reducing the dimensionality of this system. Here comes the final step of the proposed system, where we apply the PCA on this final resultant matrix.

It will be shown that the new technique improves the performance of the recognition rate when compared to the MPEG-7 FFD vector method, the method of direct eigenphase implementation to the face space [4], and the PCA method.

This work is organized as follows. A brief description of MPEG-7 Fourier Feature descriptor is given in Section 2. The formulation of the proposed technique is presented and discussed in Section 3. In Section 4, results of testing and implementing the new technique on two independent databases; the Olivetti Research Ltd. (ORL) database and xm2vts, CVSSP – University of Surrey database, are presented. Concluding remarks are given in Section 5.

2 MPEG-7 Fourier Feature Descriptor

The MPEG-7 , formally named "Multimedia Content Description Interface", objective is to describe the content of multimedia data so that it can be efficiently searched, accessed, transformed or adapted for use by any device and to support different applications [5].

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

MPEG-7 uses the FFD vector to represent the facial feature of an image. The descriptor represents the projection of a face vector onto a set of basis vectors which span the space of possible face vectors. The Face Recognition feature set is extracted from a normalized face image of size 56×46. The FFD vector that represents the facial feature of an image by a small single vector is derived from two feature vectors; one is a Fourier Spectrum Vector x_1^f , and the other is a Multi-block Fourier Amplitude Vector x_2^f of a normalized face image.

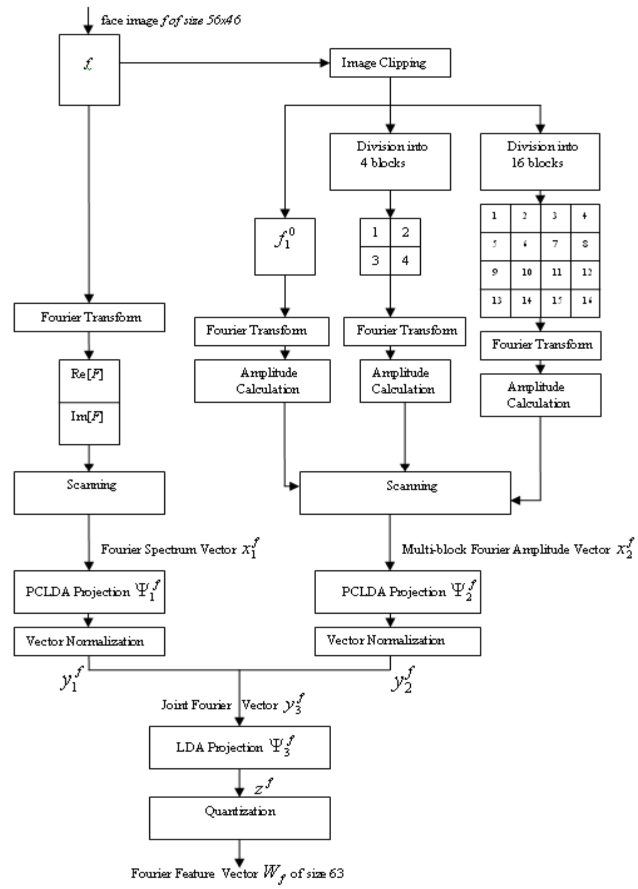


Fig. 1. The process of Fourier Feature extraction for a single image in the MPEG-7 algorithm

Fig. 1 shows the process of extracting the FFD for a single image. The quantized elements representing the FFD vector, *Wf* is of size 63. From Fig. 1 we see that the normalized image of size 56×46 goes through parallel processing or stages: a) the image is taken as a whole and its Fourier spectrum vector x_1^f is found. This vector

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

describes the image globally; b) the image is divided into different blocks and the Multi-block Fourier Amplitude Vector x_2^f is found. This x_2^f vector describes the image locally; c) the x_1^f and x_2^f vectors are projected using *Linear Discriminant Analysis using Principal Component* (PCLDA) to find the normalized vectors y_1^f and y_2^f; d) the vectors y_1^f and y_2^f are joined to form one vector where this vector is projected using *Linear Discriminant Analysis* (LDA); e) finally this resultant vector is quantized, where we end up with a small, efficient, and descriptive vector W_f of size 63. More detailed description of the MPEG-7 FFD can be found in [6-7].

Now, these vectors will span a new space and of course the elements of the vectors are of different weight of importance. It is very beneficial to investigate and find out how are these important elements or features distributed. If we could do so, then we can emphasize the important ones and neglect the relatively less important. Here comes the importance of transforming these vectors into another domain, namely finding the frequency domain of the system constructed from these FFD vectors. Section 3 will explain this part.

3 Binarized Eigenphases of the MPEG-7 Fourier Feature Descriptor Vectors

Most of the appearance-based methods for face recognition that deal with the face images as a whole, depend on calculating the eigenvalues and the eigenvectors of a system representing this face space [8-10]. This calculation consumes the major part of the processing time.

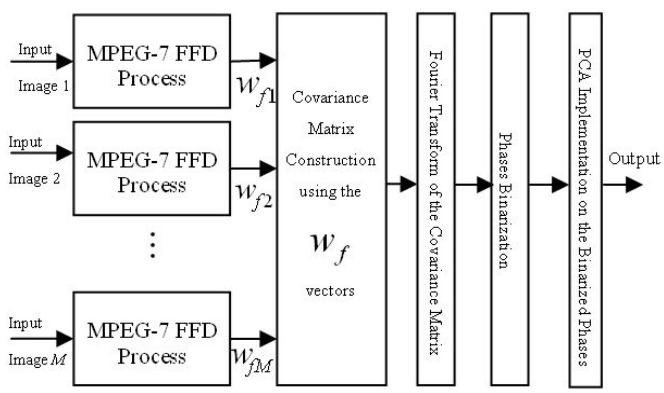


Fig. 2: Block diagram for the new system.

To overcome this problem, we propose our new system as shown in Fig. 2. First, MPEG-7 FFD vector is found for every image in data space. We will end up with *M*

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

vectors, where M is the number of images in the database. Then, for a set of FFD vectors $W_{f1}, W_{f2}, \ldots, W_{fM}$, the average FFD vector of the set is defined by

$$m = \frac{1}{M}\sum_{i=1}^{M} W_{fi} \tag{1}$$

Each W_{fi} vector differs from the average by the vector $\Phi_i = W_{fi} - m$ and the covariance matrix is then given by

$$C_s = AA^T \tag{2}$$

where $A = [\Phi_1\ \Phi_2 \ldots\ldots\Phi_M]$. Eq. (2) can be rewritten as

$$C_s = \sum_{i=1}^{M} (W_{fi} - m)(W_{fi} - m)^T \tag{3}$$

If we denote C_f as the covariance matrix in the frequency domain, then C_f is given by

$$C_f = F_{DFT} C_s F_{DFT}^{-1} \tag{4}$$

where F_{DFT} is the Fourier transform matrix containing the Fourier basis vectors. C_f is a complex matrix whose elements are represented by magnitude and phase. Finding the frequency domain of a system should provide us with the frequency contents of the system under study, where some of these components are more important than others.

As the angles of these elements retain most of the information of the images, we will concentrate on the phase information. In other words, the phases of C_f are extracted and then the following binarization procedure is applied: each one of these phases is replaced by a binary value (either 1 or 0), that is:

$$\Phi(u,v) = \begin{cases} 1 & \textit{for } \Phi(u,v) \geq T_H \\ 0 & \textit{otherwise} \end{cases} \tag{5}$$

where T_H is the threshold used for binarization. This threshold value has been set to the median of the phases of the 3×3 neighborhood elements surrounding the targeted phase. More specifically, we will find the following value:

$$\begin{aligned} &\textit{Median}\,\{\Phi(u,v),\ \Phi(u+1,v),\ \Phi(u-1,v),\ \Phi(u,v+1),\ \Phi(u,v-1),\\ &\qquad \Phi(u+1,v+1),\ \Phi(u-1,v-1),\ \Phi(u+1,v-1),\ \Phi(u-1,v+1)\} \end{aligned} \tag{6}$$

then we compare this value with $\Phi(u,v)$ to see if it is larger or smaller, and consequently we assign a value of 1 or 0 to the corresponding phase component. The

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8

reason behind choosing 3×3 size is that it is the smallest filter that we can apply since we are looking for the minimum number of computations and operations. This binarization step will help in locating the features of interest (the phases that contribute most of the discriminative information about the system). It is similar to providing the silhouette of the object when finding the binary of an image. By applying this step, we gain three important advantages:

- Low storage: no more than 1 bit/pixel, instead of representing the pixel with 8 bits
- Simple processing: the algorithms are in most cases much simpler than those applied to grey-level images
- Enhancing the performance of the system, as will be seen in Section 4.

Since we are concerned with processing capacity, we would like to further reduce the dimensionality of the system in order to cope with the system constraints and requirement. Each feature adds to a computational burden in terms of processing and storage, and this is the reason behind applying the PCA at final step to further reduce the dimensionality of our system, while we have preserved its features and characteristics as much as possible through the implementation of the previous stages.

A value of M' eigenvectors with the largest associated eigenvalues is sufficient for recognition process. In our experiments, we have found that the eigenvectors corresponding to the largest 10 eigenvalues are enough to represent the system.

Now, the FFD vector (W_f) is transformed into its eigenvector components by the following weight equation,

$$\omega_k = u_k^T (W_f - m) \tag{7}$$

where u_i is the i^{th} eigenvector and $k = 1, \ldots, M'$. The weights form a vector $\Omega^T = [\omega_1, \omega_2, \ldots\ldots, \omega_{M'}]$ that describes the contribution of each eigenvector in representing the input *binarized phase* FFD vector, and consequently the Ω^T represents the corresponding (original) face image.

We have used the Euclidian distance to determine which face class provides the best description of an input face image by finding the face class *k* that minimizes the Euclidian distance

$$\varepsilon_k = \|(\Omega - \Omega_k)\|^2 \tag{8}$$

where Ω_k is a vector describing the k^{th} *binarized phase* FFD of a face class and it is calculated by averaging the results of the eigenvector representation over a small number of *binarized phase* FFD vectors of each individual. A face is classified as belonging to class *k* if the corresponding ε_k is the minimum among all other ε_k 's.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

4 Experiments

Many different experiments were performed to evaluate our system. Note that as a part of pre-processing step, eye extraction was applied on the images and the images were normalized to 56×46 size. We carried out experiments on two independent and different databases, one is the ORL and the other is the xm2vts. Both sets include a number of images for each person, with variations in pose, expression and lighting. The ORL set includes 400 images of 40 different individuals where each individual is represented by 10. The system was trained using 5 images for each person from this set. For the xm2vts set, we have used 1180 images for 295 different individuals with each individual represented by 4 different images. In this case, 2 images of each individual were used for training. The xm2vts database images are taken at different sessions (different days). The experiments on this database test the robustness of the proposed system under the variation in time conditions of the images. Different timing means different hair style, different clothes and different "moods". Fig. 3 shows examples of the xm2vts database used for this experiment. The recognition rate was 93.5% for this experiment, while under the same conditions the MPEG-7 face recognition method achieved 89.5%.

Fig. 3. Examples from the xm2vts database

The other experiment was to test the proposed technique under other different circumstances. The ORL face database is used in this experiment. This database include images with different poses, different illuminations, different expressions (open or closed eyes, smiling or non-smiling), different facial details (glasses or no glasses), and some of them were taken at different times. Examples of the ORL database used are shown in Fig. 4. The proposed technique achieved 95.5% correct classification, while under the same conditions the MPEG-7 face recognition method achieved 91.5% correct classification.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

Fig. 4. Examples from the ORL database

A summary of the results of the recognition test on both databases is given in Table 1. Note that in this table, the results of applying the PCA method and the eigenphase method [4] directly to the face images are also shown.

Table 1. Summary of the results**.**

	PCA	Eigenphase	MPEG-7	New method
ORL	78%	81%	91.5%	95.5%
xm2vts	75%	79%	89.5%	93.5%

To prove the efficiency of our system, we have applied an evaluation methodology proposed by the developers of FERET where the performance statistics are reported as cumulative match scores. In this case, identification is regarded as correct if the true object is in the top Rank *n* matches. The results for the ORL and xm2vts databases are shown in Fig. 5.

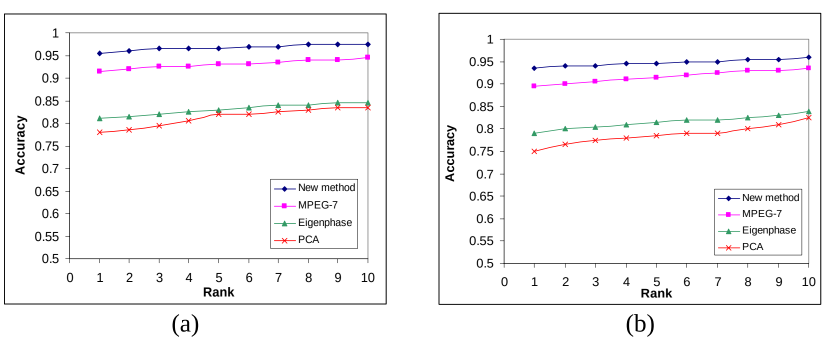
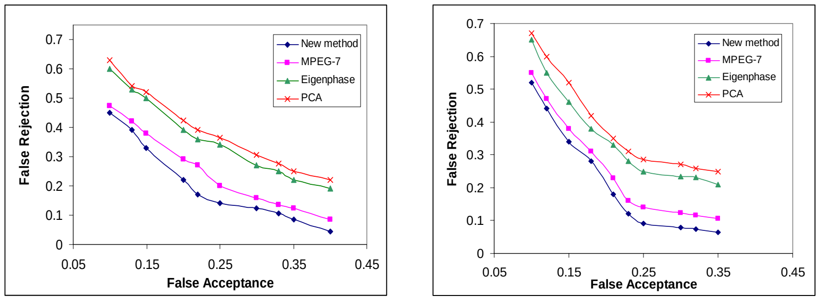


Fig. 5. Comparison of cumulative match scores between the four methods for the ORL database and the xm2vts database is shown in Fig. (a) and (b) respectively

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de



(a) (b)

Fig. 6. Comparison of ROC between the four methods for the ORL database and the xm2vts database is shown in Fig. (a) and (b) respectively

We have also assessed the performance of our new technique by comparing between the methods using the receiver operating characteristic (ROC) curve. The ROC curves for the ORL and xm2vts databases are shown in Fig. 6. It is clear from these figures that our proposed method outperforms the other methods.

Table 2. Summary of calculations and consumed time for a certain example**.**

	PCA	**Eigenphase method**	**MPEG-7**	**New method**
τ* = time	37 sec.	42 sec.	.009 sec.	0.012 sec.

***Using Pentium 4 – 2.8 GHz – 512M RAM.**

Table 2 demonstrates the amount of time consumed by each method to perform the recognition process. In table 2, τ is the time taken by the machine to perform the recognition process for a specific example (image size $N \times N$ =32×32 and number of images M used for this demonstration is 100 images). Table 2 illustrates clearly the dramatic reduction in time for our technique (an approximate ratio of more than 1:3000 is achieved when compared to the PCA method). So, if it takes a Pentium 4 machine with 2.8 GHz processor that much time, we can imagine how costly it will be for small memory devices that have a speed of only 100-300 MHz. Note that the MPEG-7 method also provides good time saving, but the recognition performance is very bad when compared to our method, as we have shown in our experiments.

5 Concluding Remarks

In this paper, a novel method is proposed for efficient face recognition that can be implemented in systems that have limited memory capabilities and have low speed

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

processors. This new technique exploits the characteristics and advantages of the MPEG-7 FFD vectors, frequency domain, and the PCA. As the covariance matrix of the face space in our system is represented in an efficient way, the calculations related to the eigenvectors of the covariance matrix are dramatically reduced. Further, a binarization step was added to the system in order to enhance the performance of the system. Moreover, the new system provides much better recognition rate compared to other techniques when applied to two independent databases.

References

[1] K. Lee, H. Byun, "A New Face Authentication System for Memory-Constrained Devices", *IEEE Transactions on Consumer Electronics*, Vol. 49, No. 4, November 2003.

[2] C. Ng, M. Savvides, P. K. Khosla, "Real-Time Face Verification System on a Cell-Phone using Advanced Correlation Filters", *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, Page(s):57 – 62, Oct. 2005, New York, USA.

[3] A. V. Oppenheim and J. S. Lim.,"The importance of phase in signals", *in Proceedings of the IEEE*, volume 69, pages 529–541, May 1981.

[4] M. Savvides, B.V.K. Vijaya Kumar and P.K. Khosla. "Eigenphases vs. Eigenfaces*", IEEE-17th International Conference on Pattern Recognition,* August 2004.

[5] Farzin Mokhtarian and Miroslaw Bober, *Curvature Scale Space: Theory, Applications, and MPEG-7 Standardization*, Kluwer Academic Publishers, Netherlands, 2003.

[6] MPEG-7 Face recognition technique, International Organization for Standarization, ISO/IEC JTC1/SC29/WG11, Coding of Moving Pictures and Audio, NPEG03/N6188, December 2003.

[7] T. Kamei, A. Yamada, Report of Core Experiment on Fourier Spectral PCA Based Face Description, ISO/IEC JTC1/SC21/WG11 M8277, Fairfax, VA, May 2002.

[8] Turk and Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, March 1991.

[9] H.C. Kim, D. Kim, S.Y. Bang, "Face Retrieval Using 1st- and 2nd- order PCA Mixture Model", *International Conference on Image Processing*, Rochester, NY, 2002.

[10] H. Kong, X. Li, J.-G. Wang, E.K. Teoh, C. Kambhamettu, "Discriminant Low-dimensional Subspace Analysis for Face Recognition with Small Number of Training Samples", *British Machine Vision Conference (BMVC)*, Oxford, UK, Sept. 5-9, 2005.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8