MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
13-16

A Novel Spatially Confined Non-Negative Matrix Factorization for Face Recognition

HF NEO, BJ TEOH, CL NGO
Faculty of Information Science and Technology
Multimedia University
Jalan Ayer Keroh Lama, Bukit Beruang, 75450 Melaka
MALAYSIA
hfneo@mmu.edu.my, bjteoh@mmu.edu.my, david.ngo@mmu.edu.my

Abstract

In this paper, a novel method for facial representation called Spatially Confined Non-Negative Matrix Factorization (SFNMF) is presented. SFNMF aims to extract more spatially confined, parts-based representation from the NMF based representation by merely removing non-prominent region, and focalize on the salient feature. SFNMF derived a significant set of basis which allows a non-subtractive representation of images and these bases manifest localized features. Experimental results are presented to compare SFNMF with NMF and Local NMF. Advantageous of SFNMF is demonstrated when SFNMF achieves highest verification rate among the other.

1 Introduction

Face recognition is one of the human most remarkable abilities. Human being is able to recognize thousands of faces learned throughout their lifetime. In psychological community, there are huge amount of work done to justify whether face recognition is based on perception of its parts or it is viewed as a holistic process. Farah et. al. [1] found that the features used to recognize faces are “holistic” in nature. On the other hand, Wachsmuth et. al. [2] have drawn phychological and physiological evidence for parts-based objects representations in the brain. Biederman developed with the theory of recognition-by-components (RBC) [3].

A well-known and widely used of holistic paradigm in face recognition is EigenFace [4] which based on Principle Component Analysis (PCA). It operates directly on whole patterns represented as (feature) vectors to extract so-needed global features for subsequent classification by a set of previously found global projectors from a given training pattern set, whose aim is to produce a most expressive subspace for face representation and recognition. Contrast to EigenFace which finds a projection direction that retains maximum variance, LDA or FisherFace [5] – an example of the most discriminating subspace method seeks a projection direction that maximizes the distances between cluster centers.

For the local based approach, Local Feature Analysis (LFA) [6] is devised as a method for extracting, from the holistic PCA basis, local topographic representation in terms of local features. Independent Component Analysis (ICA) [7] is a linear non-orthogonal transform leading to a representation in which unknown linear mixtures of multi-dimensional random variables are made as statistically independent as possible. LFA and ICA’s projection coefficients can be either positive or negative, and such linear combinations generally involve complex cancellations between positive and negative numbers. Therefore, the representations lack the intuitive meaning of adding parts to form a whole.

Recently, a new approach for obtaining a part-based linear representation of facial image has been proposed. This new technique known as Non-Negative Matrix Factorization (NMF) first introduced by Lee et. al. [8]. NMF produces a part-based representation because only additive combination of basis is allowed for entries. In addition, the non negativity constraint is compatible with the intuitive notion of combining parts to form as a whole. One of the most useful properties of NMF is that it usually produces a sparse representation of the data. Such a representation encodes much of the data using few “active” components, which makes the encoding easy to interpret [3]. However, because the sparse property given by NMF is somewhat of a side-effect rather than a goal, one cannot in any way control the degree of sparseness and it is data dependent. Li et. al. [9] found that NMF representation yields low recognition accuracy in ORL Face Database and thus they proposed Local NMF (LNMF) which leads to better classification performance.

However, NMF and LNMF basis do not display perfectly the local characteristics as there are still some non-zero weight values in the feature. These values appear as noise and contribute to the degradation of the recognition performance. In the proposed Spatially Confined NMF (SFNMF) method, basis images contain only spatially confined feature regions. Recognition by component paradigm can be effectively realized for face recognition using SFNMF basis as each SFNMF basis represents only locally salient regions.

The outline of the paper is organized as follow: Section 2 presents the overview of feature extraction techniques. Section 3 is devoted to the experimental results and conclusion is discussed in Section 4.

2 Feature Extraction Schemes

2.1 NMF

NMF finds an approximate factorization, $X \approx WH$ where X is the raw face data into non-negative factors W and H. The non-negativity constraints make the representation purely additive (allowing no subtractions), in contrast to many other linear representations such as PCA. This ensures that the components are combined to form a whole in the non-subtractive way.

Given an initial database expressed by a n x m matrix X, where each column is an n-dimensional non-negative vector of the original database (m vectors), it is possible to find two new matrices (W and H) in order to approximate the original matrix :

$$X \approx \widetilde{X} \equiv WH \quad \text{where} \quad W \in \Re^{nxr}, H \in \Re^{rxm} \tag{1}$$

We can rewrite the factorization in terms of the columns of X and H as:

$$x_j \approx \widetilde{x}_j = Wh_j \quad \text{where} \quad x_j \in \Re^n, h_j \in \Re^r \quad \text{for} \quad j = 1,\ldots,n \tag{2}$$

The dimensions of the factorized matrices W and H are n x r and r x m, respectively. Assuming consistent precision, a reduction of storage is obtained whenever r, the number of basis vectors, satisfies $(n+m)r < nm$. Each column of matrix W contains basis vectors while each column of H contains the weights needed to approximate the corresponding column in X using the basis from W.

In order to estimate the factorization matrices, an objective function has to be defined. We have used the square of Euclidean distance between each column of X and its approximation of $X=WH$ subject to this objective function:

$$\Theta_{NMF}(W,H) = \sum_{j=1}^{n} \left\| x_j - Wh_j \right\|^2 = \left\| X - WH \right\|^2 \tag{3}$$

This objective function can be related to the likelihood of generating the images in X from the basis W and encoding H. An iterative approach to reach a local minimum of this objective function is given by the following rules [10]:

$$W_{ia} \leftarrow W_{ia} \sum_{\mu} \frac{X_{i\mu}}{(WH)_{i\mu}} H_{a\mu} \tag{4}$$

$$W_{ia} \leftarrow \frac{W_{ia}}{\sum_j W_{ja}} \tag{5}$$

$$H_{a\mu} \leftarrow H_{a\mu} \sum_i W_{ia} \frac{X_{i\mu}}{(WH)_{i\mu}} \tag{6}$$

Initialization is performed using positive random initial conditions for matrices W and H. Convergence of the process is also ensured. Fig. 1 (a) demonstrates the NMF basis figures. These bases provide a sparse and part-based representation of face images.

2.2 LNMF

LNMF [9] aims to improve the locality of the learned features by imposing additional constraints. It incorporates the following three additional constraints into the original NMF formulation.

(1) LNMF attempts to minimize the number of basis components required to represent X. This implies that a basis component should not be further decomposed into more components.

(2) LNMF attempts to maximize the total "activity" on each component. The idea is to retain the basis with the most important information.

(3) LNMF attempts to produce different basis as orthogonal as possible, in order to minimize the redundancy between different basis.

LNMF incorporates the above constraints into the original NMF formulation and defines the following constrained divergence as the objective function:

$$\Theta_{LNMF}(W,H) = \sum_i^m \sum_j^n X_{ij} \log \frac{X_{ij}}{[WH]_{ij}} - X_{ij} + [WH]_{ij} + \alpha C_{ij} - \beta \sum_i D_{ii} \tag{7}$$

where, α , β >0 are constants and $C=W^TW$ and $D=HH^T$. The structure of the LNMF update for W is nearly identical to that in Equation 4, 5; differing only in the coefficient matrix H. The update for H now uses an element by element square root to satisfy the three additional constraints:

$$H_{a\mu} \leftarrow \sqrt{H_{a\mu} \sum_i W_{ia} \frac{X_{i\mu}}{(WH)_{i\mu}}} \tag{8}$$

LNMF basis are illustrated in Fig. 1(b).

2.3 SFNMF

SFNMF method is implemented through a series of simple image processing operations to its corresponding NMF basis image. Firstly, a number of r original NMF basis are selected. Each basis is processed off-line to detect the spatially confined regions. The maximum values of the basis image are identified by thresholding a histogram of pixel values and followed by the morphological dilation operation to find a blob region. As a result, SFNMF basis images where only pixels in the detected regions have grey values copied from the corresponding pixels in the original NMF image are created. The remaining pixels are set to zero.

SFNMF basis image only represents spatially confined regions. This is intuitive with the idea of recognition by components where spatially confined regions correspond to the important facial features regions such as eyes, eyebrows, nose and lips. SFNMF basis are shown in Fig. 1(c). Fig. 2 demonstrates the SFNMF process.

2.4 Face Recognition in Subspace

As in most algorithms that employ subspace projection, NMF, LNMF, and SFNMF basis are learned from a set of training images. Let ω denote the projection vector, the columns of W are NMF, LNMF or SFNMF basis images. During recognition, given an input face image, X_{test}, it is projected to $\omega = W^T X_{test}$ and classified by comparison with the vectors $'s$. that were computed from a set of training images by using three different distance metric, namely L1 norm, L2 norm and cosine angle metric

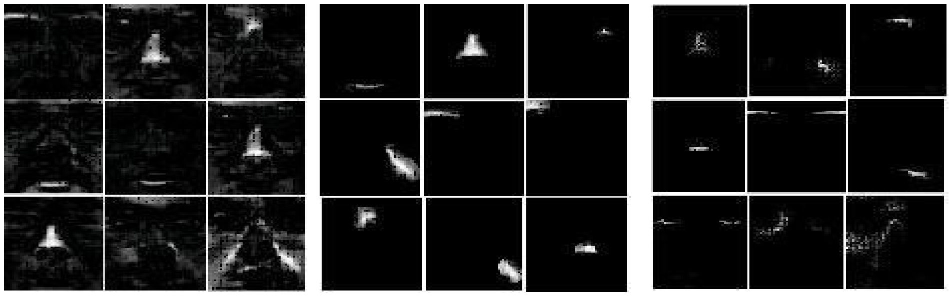
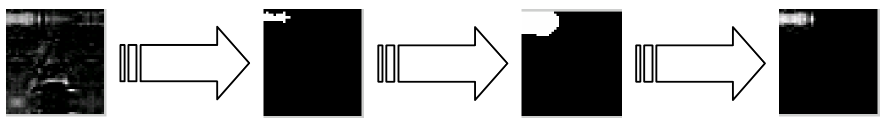


Fig. 1 (a) NMF basis (b) LNMF basis (c) SFNMF basis



NMF basis Maximum intensity detected Morphological Dilation SFNMF basis

Fig. 2 Overview of Spatially Confined NMF

3 Experimental Studies

The experiments are conducted by using *Faces*-94 Essex University Face Database (Essex) [11] and Korean Face Database (KFDB) [12]. There are various aspects in the *Faces*-94 Essex database which made it appropriate to this experiment. Data capture conditions are subject to photograph at fixed distance from camera, and individuals are asked to speak to produce images of the same individuals with different facial expressions. This database consists of 100 subjects with 10 images for each subject. The set of the 3 images for the first 50 subjects are used for training and another 50 subjects with 10 images are used for testing. The image size after cropped is 61 x 73 pixels, 256–level grayscale. The face scale in the images is uniform and there are minor variations in turn, tilt and slant.

On the other hand, KFDB contains grayscale images with eight directions in the illumination conditions,. There are five kinds of expressions under two different illumination colors – neutral, happy, surprise, anger and blink expressions. KFDB consists of 53 subjects with 10 images for each subject. The set of 7 images for 50 subjects are used for training and all subjects with 7 images are used for testing. The image size after cropped is also 61 x 73 pixels.

An experiment is carried out by using a set of r=2,3,4,5,6,7,8,9,10,20,40 to verify the performance rate of NMF, LNMF and SFNMF using three different type of distance metrics. Table 1 summarizes the NMF, LNMF and SFNMF verification rate for Essex and KFDB. LNMF leads to better classification performance in compared to NMF. The proposed method, SFNMF outperformed NMF and LNMF. SFNMF proves to be able to detect more spatially confined features in basis images.

Equal Error Rate (EER = (FAR+FRR)/2) is a measure to determine the accuracy of biometric system. The lower the EER, the better performance of the system. Fig.3 shows the EER comparison among NMF, LNMF and SFNMF for L1 norm and L2 norm. Cosine metric is excluded as the results obtained were incomparable to L1 norm and L2 norm.

The optimum verification rates for each method (NMF, LNMF and SFNMF) with each distance metric are compared using the Receiver Operating Characteristic (ROC) Curve. This is depicted in Fig. 4 and 5. A ROC curve plots the FRR against FAR at various thresholds. The closer the plot lies to the axis, the better the performance.

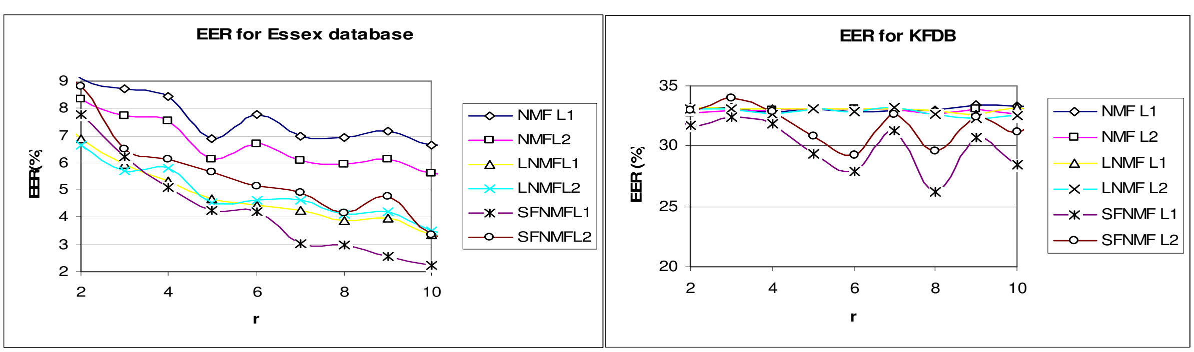


Fig. 3 EER diagrams for Essex and KFDB

Table 1 NMF, LNMF and SFNMF verification rate

Database	Metric	NMF				LNMF				SFNMF			
		r	FAR(%)	FRR(%)	TSR(%)	r	FAR(%)	FRR(%)	TSR(%)	r	FAR(%)	FRR(%)	TSR(%)
Essex	L1	20	6.60	6.66	93.38	40	3.14	3.11	96.86	40	2.45	2.44	97.54
	L2	40	5.56	5.55	94.43	40	3.11	3.11	96.88	40	3.29	3.28	96.70
	Cosine	40	5.13	6.71	94.62	40	5.07	4.84	94.95	40	5.95	6.00	94.03
KFDB	L1	4	32.42	33.96	67.41	20	32.36	33.96	67.48	9	26.10	26.24	73.88
	L2	9	32.62	32.70	67.36	10	32.23	32.35	67.76	7	29.26	29.29	70.74
	Cosine	3	38.36	38.27	61.64	2	34.72	37.65	64.98	9	30.91	31.27	69.05

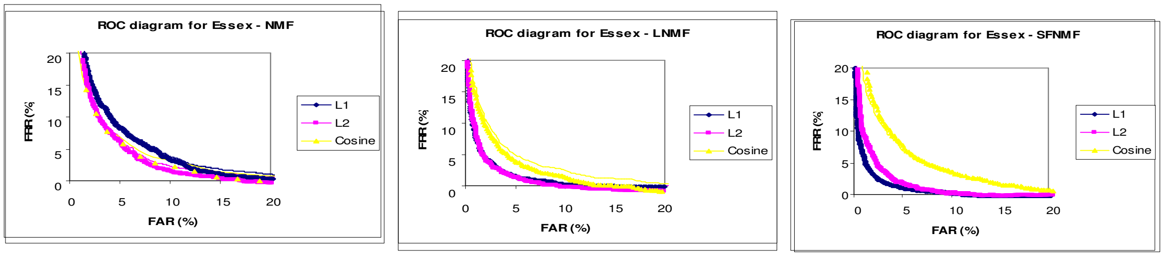


Fig. 4 ROC diagrams compare NMF, LNMF and SFNMF for Essex database

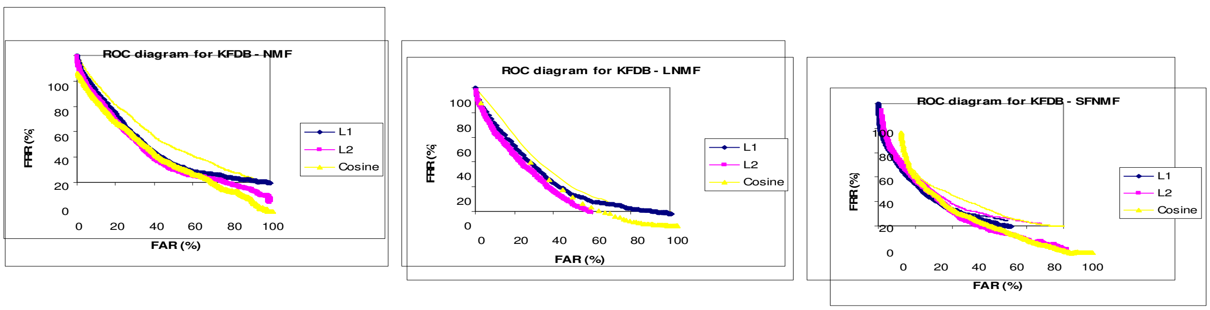


Fig. 5 ROC diagram compares NMF, LNMF and SFNMF for KFDB

4 Conclusion

In this paper, SFNMF is proposed to achieve better face verification rate. This method is intuitive with the notion of "recognition by components" which is beneficial to the problem of face recognition under partial occlusion and local distortion. The aim is to learn more spatially confined features in NMF basis components suitable for the task of face recognition. The basis derived by SFNMF is more significant compared to the original NMF and LNMF. Additionally, SFNMF achieves higher verification rate.

References

[1] M.J.Farah, K.D.Wilson, M.Drain, and J.N.Tanaka. "What is special about facial perception?", Psychological Review, 105(3):482-498, 1998.

[2] E.Wachsmuth, M.W.Oram, and D.I.Pettett. "Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque." Cereb. Cortex, 22(4):509-522, December 1994.

[3] I.Biederman. "Recognition-by-Components: A Theory of Human Image Understand", Psychological Review, 94(2):115-147, 1987.

[4] M.Turk, A.Pentland, Engenfaces for recognition, J. Cognitive Neurosci. 3(1) 71-86.

ıan, Eigenfaces vx. fic Linear Projection, IEEE Trans. On PAMI, Vol. 19, No. 7, pp. 711-720, July 1997.

[6] P.Penev and J.Atick, "Local feature analysis: A general statistical theory for object representation", Neural Systems, vol.7, no.3, 477-500, 1996.

[7] P.Comon, "Independent component analysis - a new concept?", Signal Processing, vol.36, pp.287-314, 1994.

[8] D.D.Lee, H.S.Seung, Learning the Parts of Obejcts by Non-Negative Matrix Factorization, Nature, vol. 401, pp. 788-791, 1999.

[9] S.Z.Li, X.W.Hou, H.J.Zhang, Q.Cheng, Learning Spatially Localized, Parts-Based Representation, IEEE CVPR, 2001.

[10] D.D.Lee, H.S.Seung, Algorithms for Non-Negative Matrix Factorization, Proceedings of Neural Information Processing Systems, vol. 13, pp. 556-562, 2001.

[11] Vision Group of Essex University Face Database http://cswww.essex.ac.uk/mv/allfaces/index.html

[12] B.W.Huang, J.S.Park, Y.K.Ham, S.Y.Bang, J.H.Lee, M.C.Roh, J.S.Moon and S.W.Lee, "Design and Construction of a Korean Face Database for Research and Development," Proc. Of the Korea Information Science Society Workshop on Computer Vision and Pattern Recognition, Seoul, Korea, pp.161-163, 2000. (in Korea)