

# FACE DESCRIPTION BASED ON DECOMPOSITION AND COMBINING OF A FACIAL SPACE WITH LDA

Tae-Kyun Kim<sup>1,2</sup>, Hyunwoo Kim<sup>1</sup>, Wonjun Hwang<sup>1</sup>, Seok-Cheol Kee<sup>1</sup> and Josef Kittler<sup>2</sup>

<sup>1</sup>: HCI Lab, Samsung AIT, Korea, <sup>2</sup>: CVSSP, University of Surrey, U.K.

## ABSTRACT

We propose a method of efficient face description for facial image retrieval from a large data set. The novel descriptor is obtained by decomposing the face image into several components and then combining the component features. The decomposition combined with LDA (Linear Discriminant Analysis) provides discriminative facial features that are less sensitive to light and pose changes. Each component is represented in its Fisher space and another LDA is then applied to compactly combine the features of the components. To enhance retrieval accuracy further, a simple pose classification and transformation technique is performed, followed by recursive matching. The experimental results obtained on the MPEG-7 data set show an impressive accuracy of our algorithm as compared with the conventional PCA/ICA/LDA methods.

## 1. INTRODUCTION

Many algorithms have been developed to deal with face image retrieval from huge databases such as those found in Internet environments[1,2,3]. Such retrieval requires a compact face representation which has robust recognition performance under lighting and pose variations. The partitioning of the face image into components was shown to facilitate efficient and robust recognition in our previous study[7]. It has the advantage that image variation due to pose and/or illumination change within each component patch is more easily compensated than that of the whole image. Furthermore, a facial component can be weighted according to its importance. The component with a large variation is weighted less in the matching stage. Finally, an LDA encoding is more effective at the component level, which has simplified statistics, than for the whole image.

In this paper, we combine the component-based representation with LDA. First, initial pose estimation and compensation are carried out based on the full face image. The image is then partitioned into several facial components to simplify image statistics. The components are encoded by LDA to compensate for the effects of illumination and expression variation. Another LDA is then applied to a collection of the component-based LDA representations yielding a compact description. The decomposition and combining of the facial space with

LDA effectively solves the problems of face retrieval and person identification. Finally a recursive matching is proposed to further improve the retrieval accuracy.

Section 2 explains the approach of component-based description with the cascaded LDA. The pose compensation method and the recursive matching are described in Section 3 and Section 4 respectively. Section 5 presents the experimental results and Section 6 draws the paper to conclusion.

## 2. COMPONENT-BASED LDA FACE DESCRIPTOR

In this paper, LDA applied to a whole face image is called 'the holistic LDA' and LDA applied to image components is called 'the component LDA'. Although the component scheme encodes a face image with the benefit of good linear property and robustness to image variation, it lacks relational information between the components. To overcome this problem, a two stage LDA is considered which combines both sources of information.

### 2.1. Decomposition/Representation of a Facial Space

Given a set of  $N$  training images  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ , a set of LDA transformation matrixes is extracted. First, all the images are partitioned into  $L$  facial components. The image patches of each component are represented in a vector form with the  $k$ -th component being denoted as  $\{\mathbf{c}_1^k, \dots, \mathbf{c}_N^k\}$ . Then, for the  $k$ -th facial component, the corresponding LDA transformation matrix  $\mathbf{W}^k$  is computed.

During testing, the  $L$  vectors  $\{\mathbf{c}^1, \dots, \mathbf{c}^L\}$  corresponding to the facial component patches are extracted from a face image  $\mathbf{x}$  of the test data set. A set of LDA feature vectors  $\mathbf{y} = \{\mathbf{y}^1, \dots, \mathbf{y}^L\}$  is extracted by transforming the component vectors by the corresponding LDA transformation matrixes as

$$\mathbf{y}^k = (\mathbf{W}^k)^T \mathbf{c}^k, \quad k = 1, 2, \dots, L. \quad (1)$$

Thus for the component-based LDA, a face image  $\mathbf{x}$  is represented by a set of LDA feature vectors  $\{\mathbf{y}^1, \dots, \mathbf{y}^L\}$ . This set is augmented by the LDA coefficients  $\mathbf{y}^0$  of the holistic image to yield combined representation  $\{\mathbf{y}^0, \mathbf{y}^1, \dots, \mathbf{y}^L\}$ .

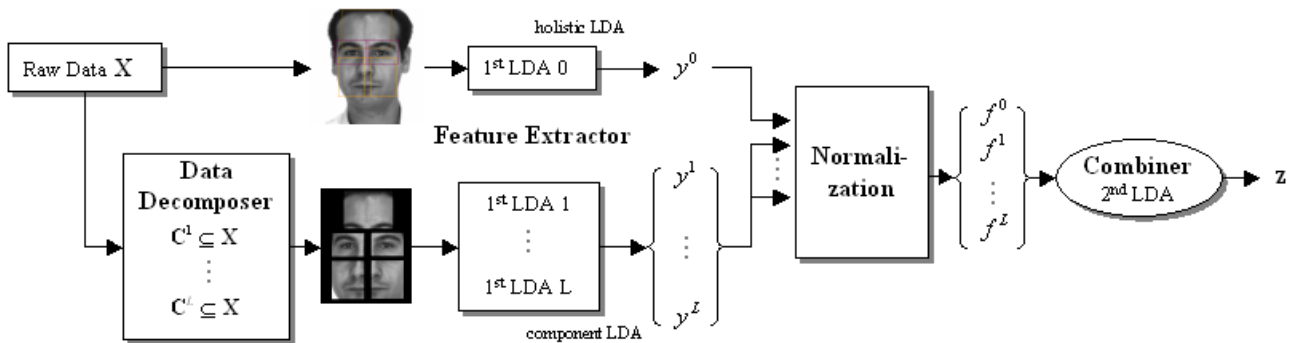


Figure 1. Architecture of the Component-based LDA Face Description

## 2.2. Combining Component Features - Cascaded LDA

The dimensionality of the simple collection of the component features derived in the previous section is quite large. In order to reduce this dimensionality, and to determine the appropriate weighting for these features, the combined representation  $\{y^0, y^1, \dots, y^L\}$  is transformed by another LDA. The resulting two stage process is referred to as “cascaded LDA”. It is illustrated in Figure 1. A face image is represented as a merged vector  $\mathbf{f}$  of the normalized LDA feature vectors  $\mathbf{y}$ . An LDA transformation matrix  $\mathbf{W}_{2nd}$  is computed for the merged vectors of the training face images. This LDA allows us to control the dimension of the final descriptor  $\mathbf{z}$ . Note that large eigenvalues indicate the corresponding transformation vectors to be more discriminative. Moreover, the elements of the LDA transformation matrix define the weights for the component features. The diagonal terms of the matrix reflects the importance of each feature and the off-diagonal terms assign weights for the combinations of the features. The proposed final description  $\mathbf{z}$  is given by

$$\mathbf{z} = (\mathbf{W}_{2nd})^T \mathbf{f}, \quad (2)$$

CTU[6] applied Generalized Discriminant Analysis (GDA), a nonlinear version of LDA, to our component-based LDA descriptor as a combiner. They achieved good retrieval performance by considering the facial space to be nonlinear. However, the complexity of the feature extraction and matching in this method is too high to be applied to large data sets.

## 3. POSE CLASSIFICATION AND COMPENSATION

When the component positions are well aligned, facial pose can be compensated, yielding accuracy improvement [4]. However, facial component detection in natural environments is difficult. Here, a pose classification technique and 2D affine transformation are combined for pose compensation based on a holistic facial image.

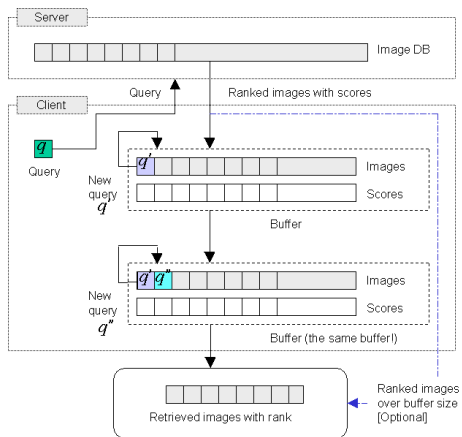
First, for the training of the pose classification stage, the face images are manually clustered according to quantized 5 pose sets (frontal, right, left, up and down). For each pose class, eigenfaces are extracted by PCA. During the pose classification stage a test image is projected into the five different Eigen-subspaces corresponding to the first few eigenfaces of each pose class. The image is classified into the class with the smallest projection error [8].

Second, for pose compensation, the inverse mapping is performed by a pre-computed affine transformation corresponding to the pose class. The transformation from each pose class to the frontal pose class is determined by a matrix that establishes the correspondence of facial feature points between the frontal pose class and each pose class. After the pose transformation, the component-based scheme is utilized for description and similarity computation, resulting in higher face retrieval accuracy for the pose data set.

## 4. RECURSIVE FACE MATCHING

A novel recursive matching shown in Figure 2 has been adopted. When a face image is encoded, the representation contains environmental variation as well as the face characteristic. When a query face is presented, the retrieved faces reflect the encoding error of the query face image injected by the environmental variations. The faces obtained under the similar imaging conditions will be retrieved successfully. Assuming that the most of the top ranked retrieved have been identified by the face characteristic, the rank one face image will encapsulate slightly different environmental variations from that affecting the query face. This means that it can be used as an another query face. Should the first retrieved face be a wrong answer, the effect of the second retrieval by the new query is weakened by weighting the matching scores between the new query and the original query. The process can be conducted recursively.

To reflect the matching by the new query  $q'$ , the score



**Figure 2.** Block Diagram of Recursive Matching

array  $\{s(i) | i = 1, \dots, M\}$  with a buffer of size  $M$ , is updated by

$$s'(i) = s(i) + w_1 \cdot D(q', q'_i), \quad (3)$$

where  $w_1$  denotes a weighing constant.  $D(z_1, z_2)$  denotes Euclidean distance of feature vectors  $z_1$  and  $z_2$  in (2). As a result, we obtain a re-sorted face image set  $\{q'_i | i = 1, \dots, M\}$  and the corresponding score array  $\{s'(i) | i = 1, \dots, M\}$ . This procedure is performed recursively. Note that the additional computational complexity of the recursive matching can be ignored in comparison with that of the simple matching which only utilizes an original query because the new matching and sorting procedure is repeated only within the buffer, not the whole database.

## 5. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental face database obtained by MPEG7 standardisation effort is used in this study. It consists of five databases: the extended version 1 MPEG-7 face database (E1), Altkom database (A2), MPEG-7 testset in XM2VTS database (M3), FERET database (F4), and MPEG-7 testset in the Banca database (B5). The details are described in Table 1. The total number of images is 11,845. All the images in the database are manually normalized to 46x56 pixels, fixed eye positions. The experiments were strictly separated for the training and test part as shown in Table 2. All the parameters such as the basis vector are extracted from the training set and all test images are utilized as a query image. As a measure of retrieval performance, we use ANMRR (Average Normalized Modified Retrieval Rate) specified in [9]. ANMRR is 0 when all ground truth images are ranked on top, and it is 1 when all ground images are ranked out of the first  $m$  images. The retrieval performance of each component is shown in Table 3 in terms of ANMRR.

Three kinds of sub-space methods for the whole face image, PCA, PCA-ICA, LDA were compared. The Bartlett's PCA-ICA technique [5] was adopted w/o the first 8 eigenfaces to remove illumination effects. From the Table 4, we see that the PCA-ICA largely outperforms PCA and the proposed component scheme also significantly enhances the performance of the PCA-ICA. The results of Ex. 1-1 in Table 5 shows that the supervised learning LDA outperforms both PCA and PCA-ICA. This is because the class specific learning is much more profitable to eliminate various changes while keeping identity information. It is noted that the proposed component/or combined scheme significantly enhances the performance for both unsupervised and supervised feature extraction methods.

Table 5 summaries the generalization performance of the holistic LDA, the component LDA and the combined LDA. Note that the component LDA highly outperforms the holistic LDA in the case of small training data. Two have similar performance in Ex. 1-1, which uses a half of data set for training and the other half for the test. While the holistic LDA can be over-trained giving a poor generalization, the component LDA learns evenly from the whole region of a face image by separating the components. The proposed combined LDA scheme further improves the retrieving accuracy.

The pose estimation/affine transformation provided the additional enhancement in facial image retrieval as shown in Table 6. The overall performance was enhanced with a bigger improvement for the data sets which have large pose variation like Altkom(A2) and XM2VTS(M3) data set. However, we see that the linear affine transformation could not solve the pose problem basically due to the non-linearity of the face pose set. The results for A2 and M3 are much worse than that of E1. Any benefit of non-linear feature extraction or transformation should be further investigated for the pose problem taking the complexity into account.

The weighted sum of cross-correlations is defined by

$$d(\mathbf{y}_1, \mathbf{y}_2) = \frac{1}{L} \sum_{k=1}^L w_k \frac{\mathbf{y}_1^k \cdot \mathbf{y}_2^k}{\|\mathbf{y}_1^k\| \|\mathbf{y}_2^k\|} \quad (4)$$

where  $\mathbf{y}_1, \mathbf{y}_2$  are the component LDA feature vector sets of two face images. The weights  $w_k$  were heuristically chosen reflecting the performance of each component shown in Table 3. The proposed LDA which is applied to the merged vector yields an efficient representation in terms of both accuracy and descriptor size. In Table 7, compared with the weighted sum method (4), the cascaded LDA has much smaller descriptor size with a similar retrieval performance. The performance of the cascaded LDA was comparable to that of GDA [6] with approximately one tenth of GDA complexity of feature extraction and matching. Finally, the proposed recursive

Table 1. Face Dataset		Table 2. Experimental Protocol								
Ref		Ex.No.	Train Image			Test Image				
			DB	Person	Image	Total	DB	Person	Image	Total
E1	5 images per person exhibiting illumination and small view variations	<b>1-1</b> [train:test=1:1]	<i>E1</i>	337	5	1685	<i>E1</i>	298	5	1490
A2	15 images per person: 5views*3illuminations	<b>2-1</b> [train:test=1:1.5]	<i>E1</i>	40	5	200	<i>E1</i>	595	5	2975
M3	10 images per person: 5 views*2 different times (session 1&4)	<b>2-2</b> [train:test=1:3]	<i>E1</i>	160	5	800	<i>E1</i>	475	5	2375
F4	875 persons 4000 images selected for the "background" at testing stage	<b>3-1</b> [train:test=1:2]	<i>A2</i>	40	15	600	<i>A2</i>	40	15	600
B5	10 images per person: 4 * office, 4*outdoor,2*ideal; each image taken at different time		<i>B5</i>	-	-	-	<i>B5</i>	52	10	520
			<i>E1</i>	317	5	1,585	<i>E1</i>	318	5	1,590
			<i>M3</i>	147	10	1,470	<i>M3</i>	148	10	1,480
			<i>F4</i>	-	-	-	<i>F4</i>	-	-	4,000
		<b>Total</b>	504		3,655	<b>Total</b>	558		8,190	

Table 3. Retrieval Performance		Table 4. Comparison of Sub-space methods		Table 5. Generalization Test			
*[Ex. No. 1-1]		*[Ex. No. 1-1]		Unit : ANMRR			
Components	ANMRR	Methods	ANMRR	Holistic LDA	Ex. 2-1	Ex. 2-2	Ex. 1-1
Forehead	0.437	PCA w/o first 8	0.499	<b>Component LDA</b>	0.159	0.104	0.107
Left Eye	0.620	PCA-ICA w/o first 8	0.367	<b>Combined LDA</b>	N/A	N/A	0.067
Right Eye	0.633	Component-based					
Left Check	0.670	PCA-ICA w/o first 8	0.252				
Right Check	0.678						

Table 6. Retrieval Results for Pose Compensation & Database							Table 7. Cascade LDA and Recursive Retrieval				
*[Ex. No. 3-1]							*[Ex. No.3-1]				
Unit: ANMRR		Total	A2	B5	E1	M3	Combining	Matching	Dimension	ANMRR	
W/O Pose	Holistic LDA	0.473	0.468	0.700	0.207	0.681	With P. C.	Weighted sum		240	0.394
	Compensation	Combined LDA	0.438	0.421	0.556	0.157		0.705	Cascaded GDA [6]	Simple matching	50
<b>With Pose Compensation</b>	Holistic LDA	0.440	0.435	0.655	0.181	0.645	Combi ned LDA	Cascaded LDA	Simple matching	50	0.403
	Compensation	Combined LDA	0.394	0.388	0.541	0.145		0.611	Recursive matching	150	0.387
										50	0.377
									150	0.359	

matching was applied to a face descriptor for more accurate image retrieval. The optimal number of the recursive matching steps and the size of the buffers were examined through the repeated retrieval experiments. It is noted that the recursive retrieval yields an additional accuracy enhancement with negligible computational costs.

## 6. CONCLUDING REMARKS

In this paper, we have proposed a face description based on face image decomposition and the projection of each component by LDA. The component LDA augmented by the holistic LDA is then transformed by another LDA. The method gives impressive retrieval accuracy compared with the conventional PCA/ICA/LDA techniques. The dimensionality of the descriptor and therefore its computational complexity are very low. The experimental results showed that the proposed description yields better generalization performance and that it is feasible application to large data sets.

## ACKNOWLEDGEMENT

The authors would like to thank Jiří Matas, Vojtech Franc

in Center for Machine Perceptron, and Toshio Kamei in NEC for their helpful discussion.

## 7. REFERENCES

- [1] T.Kamei, A.Yamada, "Face retrieval by an adaptive Mahalanobis distance using a confidence factor", *ICIP*, N.Y. 2002.
- [2] H.C.Kim, D.Kim and S.Y.Bang "Face retrieval using 1<sup>st</sup>- and 2<sup>nd</sup>-order PCA mixture model", *ICIP*, N.Y. 2002.
- [3] A.Nefian and M.Hayes, "An embedded hmm-based approach for face detection and recognition," *ICASSP*, Vol. 6, 1999.
- [4] B.Heisele, P.Ho and T.Poggio, "Face Recognition with Support Vector Machines: Global versus Component-based Approach," *ICCV*, 2001.
- [5] M.S.Bartlett, "Face Image Analysis by Unsupervised Learning," Kluwer Academic Publishers, 2001.
- [6] V.Franc, J.Matas, "An extension of the component-based LDA descriptor by the Generalized Discriminant Analysis" ISO/IEC JTC1/SC21/WG11 M8727, Klagenfurt, AT, July 2002.
- [7] T.-K. Kim, H. Kim, W. Hwang and S.-C. Kee, "Component-based LDA Face Descriptor for Image Retrieval", *British Machine Vision Conference(BMVC)*, Cardiff, UK, 2002.
- [8] B.Moghaddam, A.Pentland, "Face Recognition using View-based and Modular Eigenspaces," *SPIE Automatic Systems for the Identification and Inspection of Humans*, July 1994.
- [9] B.S.Manjunath, P.Salembier, and T.Sikora, "Introduction to MPEG-7: Multimedia Content Description Interface", John Wiley & Sons Ltd., 2002.
- [10] T.-K. Kim, H. Kim, W. Hwang, S.-C. Kee and J. Kittler, "Independent component analysis in a facial local residue space", *IEEE CVPR*, Madison, 2003.