A COMPARATIVE STUDY OF PCA, ICA AND LDA

Kresimir Delac [1], Mislav Grgic [2] and Sonja Grgic [2]

[1] Croatian Telecom, Savska 32, Zagreb, Croatia, e-mail: *kdelac@ieee.org*
[2] University of Zagreb, FER, Unska 3/XII, Zagreb, Croatia

Abstract: *Different statistical methods for face recognition have been proposed in recent years and different research groups have reported contradictory results when comparing them. The goal of this paper is to present an independent, comparative study of three most popular appearance-based face recognition algorithms (PCA, ICA and LDA) in completely equal working conditions. The motivation was the lack of direct and detailed independent comparisons in all possible algorithm implementations (e.g. all algorithm-metric combinations). FERET data set will be used for consistency with other studies. It will be shown that no particular algorithm-metric combination is the optimal across all standard FERET tests and that choice of appropriate algorithm-metric combination can only be made for a specific task.*

Keywords: *Face Recognition, PCA, ICA, LDA, FERET, Subspace Analysis Methods*

1. INTRODUCTION

Face recognition has gained much attention in recent years and has become one of the most successful applications of image analysis and understanding. A general statement of the problem can be formulated as follows [1]: Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. Currently, image-based face recognition techniques can be divided into two groups based on the face representation which they use: 1) *appearance-based* which use holistic texture features and are applied to either whole-face or specific regions in a face image, and 2) *feature-based* which use geometric facial features (mouth, eyes, brows, cheeks etc.) and geometric relationships between them.

The goal of this paper is to present an independent, comparative study of three most popular appearance-based face recognition algorithms in completely equal working conditions. They are: Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA). PCA [2] finds a set of the most representative projection vectors such that the projected samples retain the most information about original samples. ICA [3], [4] captures both second and higher-order statistics and projects the input data onto the basis vectors that are as statistically independent as possible. LDA [5], [6] uses the class information and finds a set of vectors that maximize the between-class scatter while minimizing the within-class scatter. Comparison will be done using the FERET data set [7] for consistency with other studies. This study was motivated by the lack of direct and detailed independent comparisons of these three algorithms. Very rarely are they compared in the same paper and almost never are all possible implementations considered (e.g. all algorithm-metric combinations). Another reason is that the literature on this subject is contradictory. Bartlett et al. [3] and Liu [8] claim that ICA outperforms PCA, while Baek et al. [9] claim that

PCA is better. Moghaddam [10] states that there is no significant statistical difference. Beveridge et al. [11] claim that in their tests LDA performed uniformly worse than PCA, Martinez [12] states that LDA is better for some tasks, and Belhumeur et al. [5] and Navarrete et al. [13] claim that LDA outperforms PCA on all tasks in their tests (for more than two samples per class in training phase).

The rest of this paper is organized as follows: Section 2 gives a brief description of the algorithms to be compared, Section 3 reports the details of experimental design, Section 4 reports the results and compares it to other research groups and Section 5 concludes the paper.

2. ALGORITHMS

Even though algorithms and metrics used in this work are already well known, we will include a brief description for the sake of completeness (by *algorithm* we mean the subspace projection method and by *metric* we mean the distance measure). All three algorithms are so called *subspace analysis methods*. A two dimensional image Γ with m rows and n columns can be viewed as a vector (after concatenating its rows or columns) in N dimensional ($N = m \times n$) image space. Space derived this way is substantial and recognition algorithms therefore tend to derive lower dimensional spaces to do the actual recognition in. An illustration of general subspace face recognition system can be seen in Fig 1. This is an illustration of the recognition phase (which was used in our research) where a new image is normalized, mean-substracted, projected into a subspace and then its projection is compared to the stored projections of gallery images. All three algorithms used in this research were tested using a nearest neighbour classifier with four different distance metrics: 1) L1 (city block distance), 2) L2 (Euclidean distance), 3) cosine angle (COS in further text), and 4) Mahalanobis distance (MAH in further text). These four distance measures are currently common practice in face recognition and that is why they were chosen.

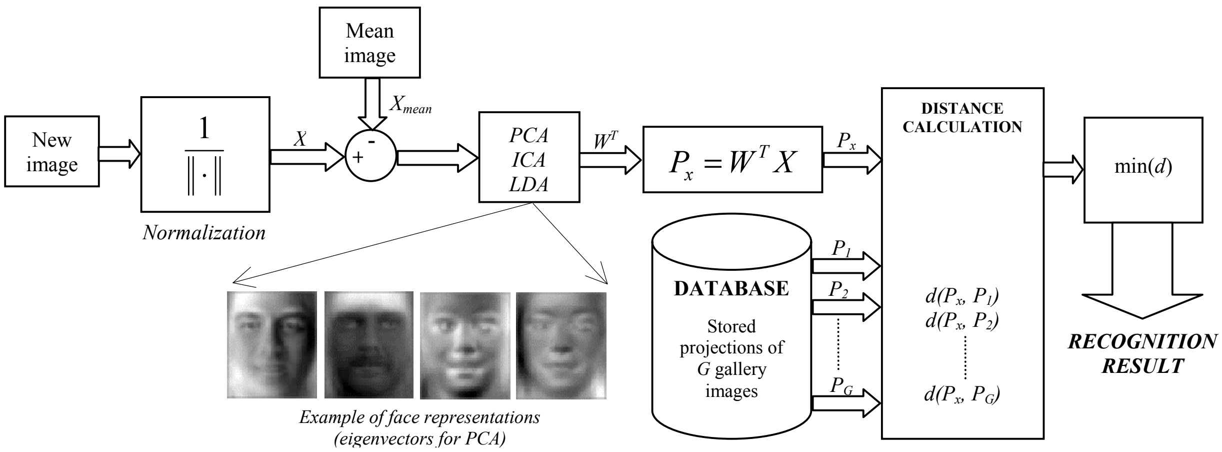

Fig. 1. A general subspace face recognition system

PCA. Given an s-dimensional vector representation of each face in a training set of M images, Principal Component Analysis (PCA) [2] tends to find a t-dimensional subspace whose basis vectors correspond to the maximum variance direction in the original image space. This new subspace is normally lower dimensional ($t << s$). New basis vectors define a subspace of face images called *face space*. All images of known faces are projected onto the face space to find a set of weights that describes the contribution of each vector. To identify an unknown image, that image is projected onto the face space to obtain its set of weights. By comparing a set of weights for the unknown face to sets of weights of known faces, the face can be identified. If the image elements are considered as random variables, the PCA basis vectors are defined as eigenvectors of the scatter matrix S_T defined as:

$$S_T = \sum_{i=1}^{M} (x_i - \mu) \cdot (x_i - \mu)^T \tag{1}$$

where μ is the mean of all images in the training set (the *mean face*) and x_i is the i-th image with its columns concatenated in a vector. The projection matrix W_{PCA} is composed of t eigenvectors corresponding to t largest eigenvalues, thus creating a t-dimensional face space. In our experiments we implemented PCA procedure as described in [2].

ICA. PCA considered image elements as random variables with Gaussian distribution and minimized second-order statistics. Clearly, for any non-Gaussian distribution, largest variances would not correspond to PCA basis vectors. Independent Component Analysis (ICA) [3], [4] minimizes both second-order and higher-order dependencies in the input data and attempts to find the basis along which the data (when projected onto them) are – *statistically independent*. Bartlett et al. [3] provided two architectures of ICA for face recognition task: *Architecture I* - statistically independent basis images (ICA1 in our experiments), and *Architecture II* - factorial code representation (ICA2 in our experiments). Also worth mentioning is the fact that our implementation of ICA uses the INFOMAX algorithm proposed by Bell and Sejnowski and used in [3]. PCA is used to reduce dimensionality prior to performing ICA. For details on ICA please refer to [3].

LDA. Linear Discriminant Analysis (LDA) [5], [6] finds the vectors in the underlying space that best discriminate among classes. For all samples of all classes the between-class scatter matrix S_B and the within-class scatter matrix S_W are defined by:

$$S_B = \sum_{i=1}^{c} M_i \cdot (x_i - \mu) \cdot (x_i - \mu)^T \tag{2}$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i) \cdot (x_k - \mu_i)^T \tag{3}$$

where M_i is the number of training samples in class i, c is the number of distinct classes, μ_i is the mean vector of samples belonging to class i and X_i represents the set of samples belonging to class i with x_k being the k-th image of that class. S_W represents the scatter of features around the mean of each face class and S_B represents the scatter of features around the overall mean for all face classes. The goal is to maximize S_B while minimizing S_W, in other words, maximize the ratio $\det|S_B| / \det|S_W|$. This ratio is maximized when the column vectors of the projection matrix (W_{LDA}) are the eigenvectors of $S_W^{-1} \cdot S_B$. In order to prevent S_W to become singular, PCA is used as a preprocessing step and the final transformation is $W_{opt}^T = W_{LDA}^T W_{PCA}^T$. For details on LDA please refer to [5].

3. EXPERIMENTAL DESIGN

Data. For consistency with other studies, we used the standard FERET data set including the data partitions (subsets) for recognition tests, as described in [7]. The *gallery* consists of 1,196 images and there are four sets of probe images that are compared to the *gallery* images in recognition stage. The *fb* probe set contains 1,195 images of subjects taken at the same time as *gallery* images with only difference being that the subjects were told to assume a different facial expression. The *fc* probe set contains 194 images of subjects under different illumination conditions. The *dup1* (duplicate I) set contains 722 images taken anywhere between one minute and 1,031 days after the *gallery* image was taken, and *dup2* (duplicate II) set is a subset of *dup1* containing 234 images taken at least 18 months after the *gallery* image was taken. All images in the data set are of size 384×256 pixels and grayscale.

Preprocessing. All three recognition algorithms and all image preprocessing steps were implemented in MATLAB®. Original FERET images were first spatially transformed (to get eyes at fixed points in imagery) based upon a ground truth file of eye coordinates supplied with the original FERET data. The standard *imrotate* MATLAB® function was used with bilinear interpolation parameter. After that, all images were cropped the same way (using the eyes coordinates) to eliminate as much background as possible. No masking was done since it turned out that cropping eliminated enough background. After cropping, images were additionally resized to be the size of 60×50 using the standard MATLAB® *imresize* function with bilinear interpolation. Finally, image pixel values were histogram equalized to the range of values from 0 to 255 using the standard *histeq* function.

Training. To train the PCA algorithm we used a subset of classes for which there were *exactly* three images per class. We found 225 such classes (different persons), so our training set consisted of $3 \times 225 = 675$ images ($M = 675$, $c = 225$). One important question worth answering at this stage is: *in what extent does the training set and gallery and probe sets overlap?* Out of 675 images in the training set, 224 were taken from the *gallery* (33%), another 224 (33%) were taken from the *fb* set and were of the same subject as the ones taken from the gallery, while 3 are in *dup1* set. The remaining 224 were not in any set used in recognition stage. We can therefore conclude that algorithms were trained roughly on 33% of subjects later used in the recognition stage. The effect that this percentage of overlap has on algorithm performance needs further exploration and will be part of our future work. PCA derived, in accordance with theory, $M - 1 = 674$ meaningful eigenvectors. We adopted the FERET recommendation and kept the top 40% of those, resulting in 270-dimensional PCA subspace (40% of $674 \approx 270$). It was calculated that 97.85% of energy was retained in those 270 eigenvectors. This subspace was used for recognition as PCA face space and as input to ICA and LDA (PCA was the preprocessing dimensionality reduction step). ICA yielded two representations (ICA1 & ICA2) using the input from PCA (as in [3]). Dimensionality of both ICA representations was also 270. However, LDA yielded only 224-dimensional space since it can, by theory, produce a maximum of $c - 1$ basis vectors. All of those were kept to stay close to the dimensionality of PCA and ICA spaces and thus make comparisons as fair as possible. After all the subspaces have been derived, all images from data sets were projected onto each subspace and recognition using nearest neighbour classification with various distance measures was performed.

4. RESULTS

Results of our experiment can be seen in Table 1 and Fig 2. We used two most popular ways to present the results: 1) table showing algorithm performance at rank 1 (recognition rate within the top one match), and 2) Cumulative Match Score (CMS) curve [7], showing the cumulative results for ranks 1 and higher. One interesting thing we noticed is the discrepancy in some cases between the rank 1 results and the CMS results when answering the question which algorithm performs better. It was noticed that the metric showing the best results at rank 1 did not always yield the best results at higher ranks. Five such cases were identified (most frequently for LDA) in this experiment. This can be seen in Table 1 by comparing the bolded best algorithm-metric combinations for rank 1 and the right two columns showing the best combinations at higher ranks. This brings to question any analyses done by comparing the CMS curves of those algorithm-metric combinations that yielded the best results at rank 1. This is why we decided to show the CMS curves for those metrics that produced best results at higher ranks for a specific algorithm.

Table 1. Algorithm performance across four metrics. Left part contains the results for rank 1 and the best algorithm-metric combinations are bolded. Right part contains are the best CMS results obtained by determining which metric gives the highest curve for a specific algorithm at a specific probe set.

Results at rank 1					**CMS results**	
Metric:	**L1**	**L2**	**MAH**	**COS**	Highest curve	Same as rank 1
Algorithm:			***fb***			
PCA	**82,26%**	*82,18%*	*64,94%*	*81,00%*	PCA+COS	N
ICA1	*81,00%*	**81,51%**	*64,94%*	*80,92%*	ICA1+L2	Y
ICA2	*64,94%*	*74,31%*	*64,94%*	**83,85%**	ICA2+COS	Y
LDA	*78,08%*	**82,76%**	*70,88%*	*81,51%*	LDA+COS	N
			fc			
PCA	**55,67%**	*25,26%*	*32,99%*	*18,56%*	PCA+L1	Y
ICA1	*18,04%*	*17,53%*	**32,99%**	*12,89%*	ICA1+L1	N
ICA2	*15,98%*	*44,85%*	*32,99%*	**64,95%**	ICA2+COS	Y
LDA	*26,80%*	*26,80%*	**41,24%**	*20,62%*	LDA+L2	N
			dup1			
PCA	**36,29%**	*33,52%*	*25,62%*	*33,52%*	PCA+L1	Y
ICA1	**32,55%**	*31,86%*	*25,62%*	*32,27%*	ICA1+L1	Y
ICA2	*28,81%*	*31,99%*	*25,62%*	**42,66%**	ICA2+COS	Y
LDA	**34,76%**	*32,96%*	*27,70%*	*33,38%*	LDA+L1	Y
			dup2			
PCA	**17,09%**	*10,68%*	*14,53%*	*11,11%*	PCA+L1	Y
ICA1	*8,97%*	*7,69%*	**14,53%**	*8,97%*	ICA1+MAH	Y
ICA2	*16,24%*	*19,66%*	*14,53%*	**28,21%**	ICA2+COS	Y
LDA	*16,24%*	*10,26%*	**16,67%**	*10,68%*	LDA+L1	N

Let us now try to draw some conclusions based on a specific task:

fb (the *different expression* task). Even though ICA2+COS combination produces the best results at rank 1 (Table 1), LDA+COS outperforms it from rank 6 further on (Fig 2). Actually, even PCA+COS outperforms it and ICA2+COS performs uniformly worse at higher ranks. But, it can be stated that all four best algorithm-metric combinations produce similar results and no straightforward conclusion can be drawn regarding which is best for this specific task (it stays unclear whether the differences in performance are statistically significant or not).

fc (the *different illumination* task). ICA2+COS wins here at rank 1 (Table 1) but PCA+L1 is much better from rank 17 on (Fig 2). ICA1 is the worst choice for this task. This is not surprising since ICA1 tends to isolate the face parts and therefore should be better at recognizing facial actions than anything else.

dup1 & dup2 (the *temporal changes* tasks). ICA2+COS is the best here at every rank (as clearly illustrated in Fig 2 for *dup1* & *dup2* and ICA1 the worst. L1 norm seems to produce the best results for almost all other algorithms and it is surprising that it is so rarely used in comparisons.

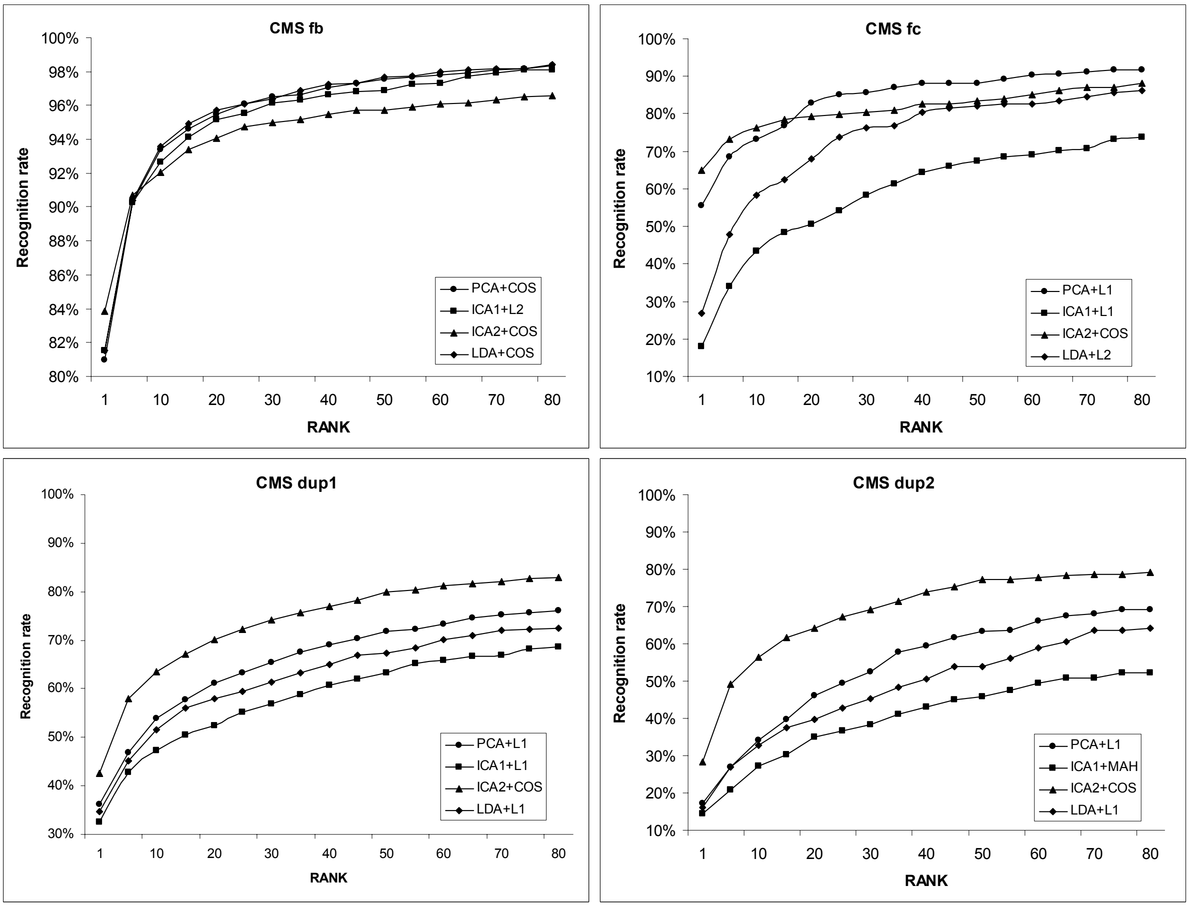

Fig. 2. Cumulative Match Score (CMS) plots of best algorithm-metric combinations (the ones that yielded the highest curve when all metrics were compared for a specific algorithms) for a given probe set.

Metrics comparison. L1 gives the best results in combination with PCA across all four probe sets so it can be concluded that L2 distance metric is suboptimal for PCA (one exception being that COS outperforms L1 for the *fb* set, but statistical significance remains questionable). Following the same line of thought, it can be concluded that COS is superior to any other metric when used with ICA2. Actually, L2 is the best metric in only two combinations across all probe sets and algorithms. We found this result surprising since this was the most frequently used measure in the past. No clear conclusion can be drawn as to which metric works best with ICA1 and LDA and, at best, it can be stated that it depends on the nature of the task. This tells us that no combination of algorithm-metric for ICA1 and LDA are robust enough across all tasks. MAH turned out to be the most disappointing metric in our tests and this could be due to the way MATLAB® calculates it (this needs more investigation). If we analyse the best results given by the CMS, the metrics ranking looks something like this: L1 – 7 best results, COS – 6, L2 – 2, MAH – 1.

Comparison to previous work. First of all, we can state that our results are consistent to [7] regarding the relative ranking of probe sets. *fb* was found to be the easiest (highest recognition rates) and *dup2* the hardest (lowest recognition rates). This is in clear contradiction with Beak et al. [9] who stated that *fc* is the hardest probe set. Also consistent with [7] is that LDA+COS outperforms all others for the *fb* set. Both [8] and [4], when comparing PCA and ICA, claim that ICA2 outperforms PCA+L2 and this is also the case in our results. However, our detailed research also produced some new conclusions: PCA+COS outperforms ICA2+COS for *fb* probe set and PCA+L1 outperforms ICA2+COS for *fc* probe set at higher ranks. Bartlett et al. [3] favour ICA2 over PCA, mostly on difficult time-separated images and our results confirm that at all ranks. We found that, for the *fb* set, ICA2+COS gives better results only at rank 1 and perform worse than PCA at higher ranks. As stated in [3], we also found that ICA2 gives best results when combined with COS. Navarrete et al. [13] claim that LDA+COS works better than PCA, which is certainly not the case here at rank 1 and is questionably true for higher ranks. We agree with Moghaddam et al. [10] who stated that there is no significant difference between PCA and ICA at rank 1, but we think that ICA is significantly worse at higher ranks.

5. CONCLUSION AND FURTHER WORK

This paper presented an independent, comparative study of three most popular appearance-based face recognition algorithms (PCA, ICA and LDA) in completely equal working conditions and across all implementations (all algorithm-metric combinations). It was found that no algorithm-metric combination is the state-of-the-art at this time, and the space of algorithm comparisons needs further research. Such a research would produce deeper understanding of individual algorithm performance, across various tasks, and yield some unified frameworks using algorithm combinations, as in [14]. From the results obtained in our experiment we can draw a few conclusions: 1) ICA2+COS combination turned out to be the best choice for *temporal changes* task, 2) COS seems to be the best choice of metric for ICA2 and gives good (but not always the best) results for all probe sets, 3) PCA+L1 outperformed all others with *illumination changes* task, 4) no claim can be made regarding which is the best combination for the *different expression* task since the differences do not seem to be statistically significant (although LDA+COS seems to be promising), 5) L1 and COS metrics produced best overall results across all algorithms and should be further investigated, 6) in most cases L2 produced lower results than L1 or COS and it is surprising that L2 was used so often in the past. Finally, it can be stated that no algorithm-metric combination is the state-of-the-art and the choice of appropriate algorithm-metric combination can only be made for a specific task.

Our future work shall be focused on determining the generalization abilities of these three algorithms. Also, we shall change the experimental design (we shall not use the standard FERET subsets) and perform some permutation tests to be able to use the hypothesis testing in order to determine statistical significance of performance differences (following the lead of [11] perhaps). We would also like to study the effect of the exact choice of images in a gallery or in a probe set has on face recognition performance. Some variations or better implementations of Mahalanobis metric will be investigated and other metrics (distance measures) will be implemented as well.

ACKNOWLEDGMENT

Portions of the research in this paper use the Color FERET database of facial images collected under the FERET program.

REFERENCES

[1] W. Zhao, R. Chellappa, J. Phillips, A. Rosenfeld, "Face Recognition in Still and Video Images: A Literature Survey", *ACM Computing Surveys*, Vol. 35, Dec. 2003, pp. 399-458

[2] M. Turk, A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86

[3] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, "Face Recognition by Independent Component Analysis", *IEEE Trans. on Neural Networks*, Vol. 13, No. 6, November 2002, pp. 1450-1464

[4] B. Draper, K. Baek, M.S. Bartlett, J.R. Beveridge, "Recognizing Faces with PCA and ICA", *Computer Vision and Image Understanding (Special Issue on Face Recognition)*, Vol. 91, Issues 1-2, July-August 2003, pp. 115-137

[5] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *Proc. of the Fourth European Conference on Computer Vision*, Vol. 1, 14-18 April 1996, Cambridge, UK, pp. 45-58

[6] W. Zhao, R. Chellappa, A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition", *Proc. of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, 14-16 April 1998, Nara, Japan, pp. 336-341

[7] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, October 2000, pp. 1090-1104

[8] C. Liu, H. Wechsler, "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition", *Second International Conference on Audio- and Video-based Biometric Person Authentication*, Washington D.C., USA, 22-23 March 1999

[9] K. Baek, B. Draper, J.R. Beveridge, K. She, "PCA vs. ICA: A Comparison on the FERET Data Set", *Proc. of the Fourth International Conference on Computer Vision, Pattern Recognition and Image Processing*, Durham, NC, USA, 8-14 March 2002, pp. 824-827

[10] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for Visual Recognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, October 2002, pp. 780-788

[11] J.R. Beveridge, K. She, B. Draper, G.H. Givens, "A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition", *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, December 2001, Kaui, HI, USA, pp. 535- 542

[12] A. Martinez, A. Kak, "PCA versus LDA", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, February 2001, pp. 228-233

[13] P. Navarrete, J. Ruiz-del-Solar, "Analysis and Comparison of Eigenspace-Based Face Recognition Approaches", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 16, No. 7, November 2002, pp. 817-830

[14] X. Wang, X. Tang, "A Unified Framework for Subspace Face Recognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, September 2004, pp. 1222-1228