

Face Recognition in Different Subspaces: A Comparative Study

Borut Batagelj and Franc Solina

University of Ljubljana, Faculty of Computer and Information Science,
Tržaška 25, SI-1000 Ljubljana, Slovenia
{Borut.Batagelj, Franc.Solina}@fri.uni-lj.si

Abstract. Face recognition is one of the most successful applications of image analysis and understanding and has gained much attention in recent years. Among many approaches to the problem of face recognition, appearance-based subspace analysis still gives the most promising results. In this paper we study the three most popular appearance-based face recognition projection methods (PCA, LDA and ICA). All methods are tested in equal working conditions regarding pre-processing and algorithm implementation on the FERET data set with its standard tests. We also compare the ICA method with its whitening preprocess and find out that there is no significant difference between them. When we compare different projection with different metrics we found out that the LDA+COS combination is the most promising for all tasks. The L1 metric gives the best results in combination with PCA and ICA1, and COS is superior to any other metric when used with LDA and ICA2. Our results are compared to other studies and some discrepancies are pointed out.

1 Introduction

As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past few years. The problem of machine recognition of human faces continues to attract researchers from disciplines such as image processing, pattern recognition, neural networks, computer vision, computer graphics, computer art [2], and psychology. The strong need for user-friendly systems that can secure our assets and protect our privacy without losing our identity in dozens of passwords and PINs is obvious. One of the advantages of the personal identification system based on analysis of frontal images of the face regard on other biometric analysis is that it is effective without the participant's cooperation or knowledge. A general statement of the problem of machine recognition of faces can be formulated as follows: given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. A survey of face recognition techniques is given in [1].

In general we can divide the face recognition techniques into two groups: geometric feature-based approach and appearance-based approach. The geometric feature-based approach uses properties of facial features such as eyes, nose, mouth, chin and there relations for face recognition descriptors. Advantages of this approach include economy

and efficiency when achieving data reduction and insensitivity to variations in illumination and viewpoint. However, facial feature detection and measurements techniques developed to date are not reliable enough for geometric feature-based recognition. Such geometric properties alone are inadequate for face recognition because rich information contained in the facial texture or appearance is discarded. This problem tries to achieve local appearance-based feature approaches.

On the other hand, the appearance-based approach, such as PCA, LDA and ICA based methods, has significantly advanced face recognition techniques. Such an approach generally operates directly on an image-based representation. It extracts features into a subspace derived from training images. In addition those linear methods can be extended using nonlinear kernel techniques to deal with nonlinearity in face recognition. Although the kernel methods may achieve good performance on the training data, it may not be so for unseen data owing this to their higher flexibility than linear methods and a possibility of overfitting therefore.

Subspace analysis is done by projecting an image into a lower dimensional space and after that recognition is performed by measuring the distances between known images and the image to be recognized. The most challenging part of such a system is finding an adequate subspace. In the paper three most popular appearance-based subspace projection methods will be presented: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA). Using PCA [3], a face subspace is constructed to represent “optimally” only the face object. Using LDA [4], a discriminant subspace is constructed to distinguish “optimally” faces of different persons. In comparison with PCA which takes into account only second order statistics to find a subspace, ICA [5] captures both second and higher-order statistics and projects the input data onto the basis vectors that are as statistically independent as possible. We made a comparison of those three methods with three different distance metrics: City block (L1), Euclidean (L2) and Cosine (COS) distance.

For consistency with other studies we used the FERET data set [9], with its standard gallery images and probe sets for testing. Even though a lot of studies were done with some of those methods it is very difficult to compare the results with each other because of different preprocessing, normalization, different metrics and even databases. Although the researcher used the same database they chose different training sets. We also noticed that the results of other research groups are often contradictory. In most cases the results are given only for one or two projection-metric combinations for a specific projection method, and in some cases researchers are using nonstandard databases or some hybrid test sets derived from standard database. Bartlett et al. [5] and Liu et al. [10] claim that ICA outperforms PCA, while Beak et al. [11] claim that PCA is better. Moghaddam [12] states that there is no significant difference. Beveridge et al. [13] claim that in their test LDA performed uniformly worse than PCA, Martinez [14] states that LDA is better for some tasks, and Navarrete et al. [15] claim that LDA outperforms PCA on all tasks in their tests.

The rest of the paper is organized as follows: Section 2 gives brief description of the algorithm to be compared, Section 3 reports the details of methodology, Section 4 presents the results and compares our results to results of other research groups and Section 5 concludes the paper.

2 Algorithms

For face recognition and comparison we used well known appearance-based methods: PCA, LDA and ICA. All three methods reduce the high dimension image space to smaller dimension subspace which is more appropriate for presentation of the face images. A two dimensional image X with m rows and n columns can be viewed as a vector in $\Re^{N=m \times n}$ dimensional space. Image comparison is very difficult in such high dimension space. Therefore, the methods try to reduce the dimension to lower one while retaining as much information from the original images as possible. In our case, where the normalized image of the face has $N = 60 \times 50$ pixels, the image space dimensionality is $\Re^{50 \times 60 = 3000}$. With subspace analysis method we reduce this image space to $\Re^{m=403}$.

Reduced image space is much lower than original image space ($m \ll N$), in spite of that we retained 98.54% of original information.

Figure 1a presents a general appearance-based system for face recognition. In the left part of the figure is the training of the subspace system and in the right is the procedure for projecting gallery images onto subspace with the projection matrix W^T . Matrix X containing the images as vectors in its columns, vector x_{mean} presenting mean image, matrix \tilde{X} containing mean-subtracted images in its columns, vector x_g presenting image from gallery. During the training phase, the projection matrix W^T is calculated which contains the basis vectors of the subspaces. Then the gallery images of known persons are projected onto subspace. At the end, such presented images are stored in the database. Later, in the matching phase (Fig. 1b), normalized and mean-subtracted probe image is projected onto the same subspace as the gallery image was and its projection is then compared to stored gallery projection. For comparison the nearest neighbor is determined by calculating the distance from a probe image projection to all gallery images projections and then choosing the minimal distance as similarity measure. The most similar gallery image is then chosen to be the result of the recognition and the unknown probe image is identified.

2.1 Principal Component Analysis (PCA)

The PCA method [3] tends to find such subspace whose basis vectors correspond to the maximum variance direction in the original image space. 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 sets of weights that describe the contribution of each vector. For identification an unknown person, the normalized image of person is first projected onto face space to obtain its set of weights. Then we compare these weights to sets of weights of known people from gallery. If the image elements are considered as random variables, the PCA basis vectors are defined as eigenvectors of scatter matrix S_T :

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

where μ is the mean of all images in the training set (the mean face, Fig. 1), x_i is the i -th image with its columns concatenated in a vector and M is the number of all

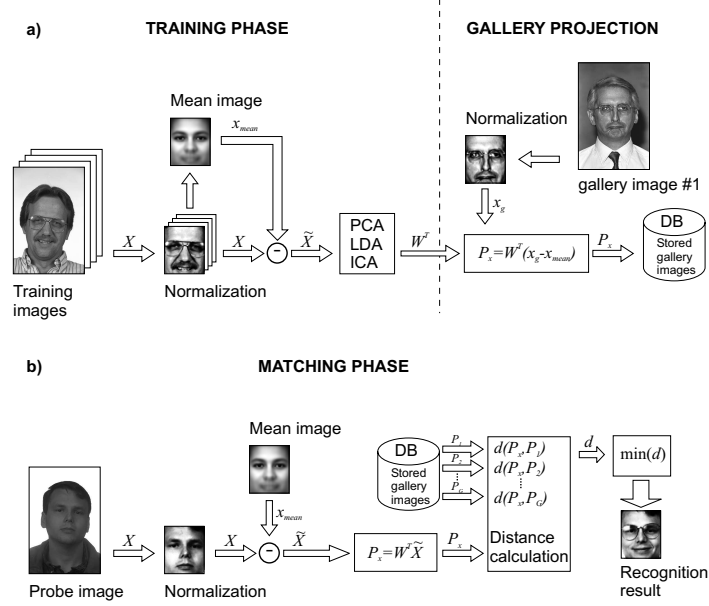


Fig. 1. A general subspace appearance-based face recognition system. a) Training images is determined a subspace and gallery images are projected and stored as prototypes. b) Probe images are projected to the known subspace and the identification is determined based on minimal distance.

training images. The projection matrix W_{PCA} is composed of m eigenvectors corresponding to m eigenvalues of scatter matrix S_T , thus creating a m -dimensional face space. Since these eigenvectors (PCA basis vectors) look like some ghostly faces they were conveniently named eigenfaces.

2.2 Linear Discriminant Analysis (LDA)

LDA method [4] finds the vectors in the underlying space that best discriminate among classes. For all samples of all classes it defined two matrix: between-class scatter matrix S_B and the within-class scatter matrix S_W . S_B represents the scatter of features around the overall mean μ for all face classes and S_W represents the scatter of features around the mean of each face class:

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

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

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. The goal is to maximize S_B while minimizing S_W , in other word, maximize the ratio $\det|S_B|/\det|S_W|$. This ratio is maximized when the column vectors of projection matrix (W_{LDA}) are the eigenvectors of $S_W^{-1} \cdot S_B$.

To prevent singularity of the matrix S_W , PCA is used as preprocessing step and the final transformation is $W_{opt} = W_{PCA}W_{LDA}$.

2.3 Independent Component Analysis (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. ICA [5] minimizes both second-order and higher-order dependencies in the input data and attempts to find the basis along which the projected data are statistically independent. For the face recognition task were proposed two different architectures: Architecture I - has statistically independent basis images (ICA I) and Architecture II assumes that the sources are independent coefficients (ICA II). These coefficients give the factorial code representation. A number of algorithm exist; most notable are Jade, InfoMax, and FastICA. Our implementation of ICA uses the FastICA package [7] for its good performances.

The Architecture I provides a more localized representation for faces, while ICA Architecture II, like PCA in a sense, provides a more holistic representation (Fig. 2). ICA I produces spatially localized features that are only influenced by small parts of an image, thus isolating particular parts of faces. For this reason ICA I is optimal for recognizing facial actions and suboptimal for recognizing temporal changes in faces or images taken under different conditions. Preprocessing steps of the methods ICA involves a PCA process by vertically centering (for ICA I), and whitened PCA process by horizontally centering (for ICA II). So, it is reasonable to use these two PCA algorithms to reevaluate the ICA-based methods [8].

ICA Architecture I includes a PCA by vertically centering (PCA I):

$$P_v = X_v V^T \Lambda^{-1/2} \quad (4)$$

where X_v is the vertically-centered training image column data matrix. Symbols Λ and V correspond to largest eigenvalues and eigenvectors of S_T matrix respectively:

$$S_T = \sum_{i=1}^M (x_i - \mu_v) \cdot (x_i - \mu_v)^T, \quad \mu_v = \frac{1}{N} \sum_{j=1}^N x_j \quad (5)$$

In contrast to standard PCA, PCA I removes the mean of each image while standard PCA removes the mean image of all training samples.

ICA Architecture II includes a whitened PCA by horizontally centering (PCA II):

$$P_w = P_h \cdot \left(\frac{1}{M} \Lambda\right)^{-1/2} = \sqrt{M} X_h V \Lambda^{-1}, \quad (6)$$

where P_h is the projection matrix of standard PCA method:

$$P_h = X_h V \Lambda^{-1/2}. \quad (7)$$

Matrix X_h contains in rows horizontally-centered training images. PCA II is actually the whitened version of standard PCA.

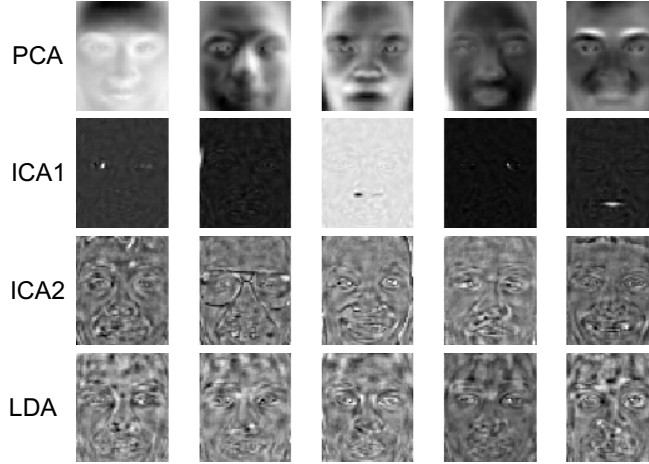


Fig. 2. Face representations found by PCA, ICA I, ICA II and LDA methods.

Figure 2 shows first five eigenfaces of PCA, ICA and LDA methods. These images look like ghostly faces are basis vectors produced by projection methods, reshaped to a matrix form of the same size as original image.

2.4 Distance measures

To measure the distance between unknown probe image and gallery images stored in database (Fig. 1b) three different distance measures will be used. Manhattan (L1), Euclidean (L2) and Cosine (COS) distance. Generally, for two vectors, x and y distance measures are defined as:

$$d_{L1}(x, y) = |x - y| \quad (8)$$

$$d_{L2}(x, y) = \|x - y\| \quad (9)$$

$$d_{COS}(x, y) = 1 - \frac{x^T \cdot y}{\|x\| \cdot \|y\|}, \quad (10)$$

where the L2-norm of the vector is denoted as $\|\cdot\|$ and the L1-norm as $|\cdot|$.

3 Methodology

3.1 Face Database

For consistency with other studies, we used the standard FERET data set. The FERET database includes the data partitions (subsets) for recognition tests, as described in [9].

The gallery consists of 1196 images, one image per subject and there are four sets of probe images (fb, fc, dup1 and dup2) that are compared to the gallery images in recognition stage. The fb probe set contains 1195 images of subjects taken at the same time as gallery images with different facial expression. The fc probe set contains 194 images of subjects under different illumination conditions. The dup1 set contains 722 images taken anywhere between one minute and 1031 days after the gallery image was taken, and dup2 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 size 384×256 and grayscale.

3.2 Normalization

All algorithms and all image preprocessing were done with Matlab. The standard *imrotate* function was used with bilinear interpolation parameter to get the eyes at fixed points. Transformation is based upon a ground truth file of eye coordinates supplied with the original FERET data. All images were then cropped the same way 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 standard *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.

3.3 Training

To train the PCA algorithm we used $M=1007$ FERET images of $c=504$ classes (different persons). Each class contains a different number of persons. These numbers vary from 1 to 10. Out of 1007 images in training set, 396 of images are taken from the gallery (39% of all training images) and 99 images are taken from dup1 probe set (10% of all training images). The remaining 512 are not in any set used for recognition. The training set and gallery overlap on about 33% and with dup1 probe set on about 14%.

PCA derived, in accordance with theory, $M - 1 = 1006$ meaningful eigenvectors. We adopted the FERET recommendation and kept the top 40% of those, resulting in 403-dimensional PCA subspace. In such way 98.54% of original information (energy) was retained in those 403 eigenvectors. This subspace was used for recognition as PCA face space and as input to LDA and ICA (PCA was the preprocessing dimensionality reduction step). For ICA representation we also try to use more eigenvectors but the performance was worse. We also confirm the findings in [8] that recognition performance is not different if we use only preprocessing step of ICA method. In our case where the dimensionality of ICA representation is the same as the dimensionality of PCA the performance is the same for L2 and COS metrics and for the L1 metrics the performance is not much different. Besides of using time consuming ICA methods we can use only preprocessing whitening step (PCA I instead of ICA I and PCA II instead of ICA II). Although LDA can produce a maximum of $c - 1$ basis vectors we kept only 403 to make fair comparisons with PCA and ICA methods. After all the subspaces have been derived, all images from data sets were projected onto subspace and recognition using neighbor classification with various distance measures was conducted.

4 Results

Results of our experiment can be seen in Table 1. We test all the projection-metric combinations. Since we implemented four projection methods (PCA, LDA, ICA1 and ICA2) and three distance measures (L1, L2 and COS) it can be concluded that we compared 12 different algorithms. The best performance on each data set for each method is bolded.

Table 1. Performance across four projection methods and three metrics. The best projection-metric combinations are in bold.

	L1	L2	COS	L1	L2	COS
	<i>fb probe set</i>			<i>fc probe set</i>		
PCA	88.87%	87.70%	86.78%	54.64%	14.95%	16.49%
LDA	81.42%	83.35%	91.46%	53.61%	54.12%	79.38%
ICA1	91.97%	87.70%	87.36%	23.20%	14.95%	14.43%
ICA2	71.30%	79.00%	89.04%	34.02%	51.03%	78.87%
	<i>dup1 probe set</i>			<i>dup2 probe set</i>		
PCA	42.52%	37.26%	37.95%	20.51%	13.68%	14.10%
LDA	43.21%	47.09%	64.13%	27.35%	35.04%	47.01%
ICA1	41.55%	37.26%	37.53%	15.81%	13.68%	13.68%
ICA2	20.50%	33.52%	47.92%	10.26%	22.65%	30.77%

On the fb (the different expression task) probe set the best combination is ICA1+L1, but it can be stated that the remaining three projection-metric combinations (LDA+COS, ICA2+COS and PCA+L1) produce similar results and no straightforward conclusion can be drawn regarding which is the best for specific task. ICA1 performance was comparable to LDA and this confirms the theoretical property of ICA1 that it is optimal for recognizing facial actions.

On the fc (the different illumination task) probe set LDA+COS and ICA2+COS win. ICA1 is the worst choice, which is not surprising since ICA1 tends to isolate the face parts and is therefore not appropriate for recognizing images taken under different illumination conditions.

On the dup1 and dup2 (the temporal change tasks) probe sets, again LDA+COS wins and ICA1 is the worst, especially for the dup2 data set. ICA2+COS also did very good on such difficult tasks.

If we compare the metrics the L1 gives the best results in combination with PCA and ICA1. It can be concluded that COS is superior to any other metric when used with LDA and ICA2. We found it surprising that L2 is not the best choice in any of the combinations, but in the past research it was the most frequently used metric.

Fb probe set was found to be the easiest (highest recognition rates) and dup2 the most demanding (lower recognition rates), which is consistent with [9], but in contradiction with Beak et al. [11] who stated that fc is the most demanding probe set. Also consistent with [9] is that LDA+COS outperforms all others. Both [9] and [6], when comparing PCA and ICA, claim that ICA2 outperforms PCA+L2 and this is what we

also found. As stated in [5], we also found that ICA2 gives best result when combined with COS. We also agree with Navarrete et al. [15] that LDA+COS works better than PCA. We agree with Moghaddam et al. [12] and with Yang et al. [8] who stated that there is no significant difference between PCA and ICA. We also confirm the result in [8] that there is no significant performance difference between ICA and preprocessing whitening PCA step.

5 Conclusion

This paper presented an independent, comparative study of three most popular appearance based face recognition projection methods (PCA, LDA and LDA) and their accompanied three distance metrics (City block, Euclidean and Cosine) in equal working conditions. This experimental setup yielded 12 different algorithms to be compared. From our independent comparative research we can derive that the L2 metric is the most promising combination for all tasks. Although ICA1+L1 seems to be promising, except for the illumination changes task where LDA+COS and ICA2+COS outperforms PCA and ICA1. For all probe sets the COS seems to be the best choice of metric for LDA and ICA2 and L1 for PCA and ICA1. LDA+COS combination turned out to be the best choice for temporal changes task. In spite of the fact that L2 metric produced lower results it is surprising that it was used so often in the past. We also tested only whitened PCA preprocessing step of ICA method and it confirms that there is no performance difference between ICA and preprocessing PCA.

References

1. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: A Literature Survey, *ACM Computing Surveys*, (2003) 399–458
2. Solina, F., Peer, P., Batagelj, B., Juvan, S., Kovač, J.: Color-Based Face Detection in the "15 Seconds of Fame" Art Installation, *International Conference on Computer Vision / Computer Graphics Collaboration for Model-based Imaging, Rendering, image Analysis and Graphical special Effects MIRAGE'03*, (2003) 38–47
3. Turk, M., Pentland, A.: Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, 3(1), 1991, 71–86
4. Zhao, W., Chellappa, R., Krishnaswamy, A.: Discriminant Analysis of Principal Components for Face Recognition, *Proc. of the 3rd IEEE International Conference on Face and Gesture Recognition, FG'98*, (1998) 336
5. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face Recognition by Independent Component Analysis, *IEEE Trans. on Neural Networks*, 13(6), (2002) 1450–1464
6. Draper, B., Baek, K., Bartlett, M.S., Beveridge, J.R.: Recognizing Faces with PCA and ICA, *Computer Vision and Image Understanding (Special Issue on Face Recognition)*, 91(1-2), (2003) 115–137
7. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. *Neural Networks*. 13(4-5) (2000) 411–430
8. Yang, J., Zhang, D., Yang, J.Y.: Is ICA Significantly Better than PCA for Face Recognition? *ICCV*. (2005) 198–203

9. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET Evaluation Methodology for Face-Recognition Algorithms, *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 22(10), (2000) 1090–1104
10. Liu, C., Wechsler, H.: Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition, *Second International Conference on Audio- and Video-based Biometric Person Authentication*, (1999) 22–23
11. Baek, K., Draper, B., Beveridge, J.R., She, K.: PCA vs. ICA: A Comparison on the FERET Data Set, *Proc. of the Fourth International Conference on Computer Vision, Pattern Recognition and Image Processing*, (8-14), (2002) 824–827
12. Moghaddam, B.: Principal Manifolds and Probabilistic Subspaces for Visual Recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(6), 2002 780–788
13. Beveridge, J.R., She, K., Draper, B., Givens, G.H.: 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*, (2001) 535–542
14. Martinez, A., Kak, A.: PCA versus LDA, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2), 2001 228–233
15. Navarete, P., Ruiz-del-Solar, J.: Analysis and Comparison of Eigenspace-Based Face Recognition Approaches, *International Journal of Pattern Recognition and Artificial Intelligence*, 16(7), 2002 817–830