Face Recognition Based on Locally Salient ICA Information

Jongsun Kim, Jongmoo Choi, Juneho Yi

School of Information & Communication Engineering
Sungkyunkwan University
300, Chunchun-dong, Jangan-gu Suwon 440-746, Korea
{jskim, jmchoi, jhyi}@ece.skku.ac.kr

Abstract. ICA (Independent Component Analysis) is contrasted with PCA (Principal Component Analysis) in that ICA basis images are spatially localized, highlighting salient feature regions corresponding to eyes, eye brows, nose and lips. However, ICA basis images do not display perfectly local characteristic in the sense that pixels that do not belong to locally salient feature regions still have some weight values. These pixels in the non-salient regions contribute to the degradation of the recognition performance. We have proposed a novel method based on ICA that only employ locally salient information. The new method effectively implements the idea of "recognition by parts" for the problem of face recognition. Experimental results using AT&T, Harvard, FERET and AR databases show that the recognition performance of the proposed method outperforms that of PCA and ICA methods especially in the cases of facial images that have partial occlusions and local distortions such as changes in facial expression and at low dimensions.

1 Introduction

Over the last ten years, canonical subspace projection techniques such as PCA and ICA are widely used in the face recognition research [2-4]. These techniques employ feature vectors consisting of coefficients that are obtained by projecting facial images onto their basis images. The basis images are computed offline from a set of training images. ICA is contrasted with PCA in that ICA basis images are more spatially local than PCA basis images. Fig. 1 (a) and (b) show facial image representation using PCA and ICA basis images, respectively, that are computed from a set of images randomly selected from the AR database. PCA basis images display global properties in the sense that they assign significant weights to the same pixels. It accords with the fact that PCA basis images are just scaled versions of global Fourier filters [21]. In contrast, ICA basis images are spatially more localized, highlighting salient feature regions corresponding to eyes, eye brows, nose and lips. This local property of ICA basis images makes the performance of ICA based recognition methods better than PCA methods in terms of robustness to partial occlusions and local distortions such as changes in facial expression. Thus, ICA techniques have popularly been applied to the problem of face recognition [3-6], especially for face recognition under variations of illumination, pose and facial expression. However, ICA basis images do not display

perfectly local characteristics in the sense that pixels that do not belong to locally salient feature regions still have some weight values. These pixels in the non-salient regions contribute to the degradation of the recognition performance.

We propose a novel method based on ICA, named LS-ICA (locally salient ICA) where the concept of "recognition by parts" [18-20] can be effectively realized for face recognition. The idea of "recognition by parts" has been a popular paradigm in object recognition research that can be successfully applied to the problem of object recognition with occlusion. Our method is characterized by two ideas: one is removal of non-salient regions in ICA basis images so that LS-ICA basis images only employ locally salient feature regions. The other is to use ICA basis images in the order of class separability so as to maximize the recognition performance. Experimental results show that LS-ICA performs better than PCA and ICA especially in the cases of partial occlusions and local distortions such as changes in facial expression. In addition, the performance improvement of LS-ICA over ICA based methods was much greater as we decrease the dimensionality (i. e. the number of basis images used).

The rest of this paper is organized as follows. Section 2 contrasts ICA with PCA in terms of locality of features. Section 3 describes the proposed LS-ICA method. Section 4 presents experimental results.
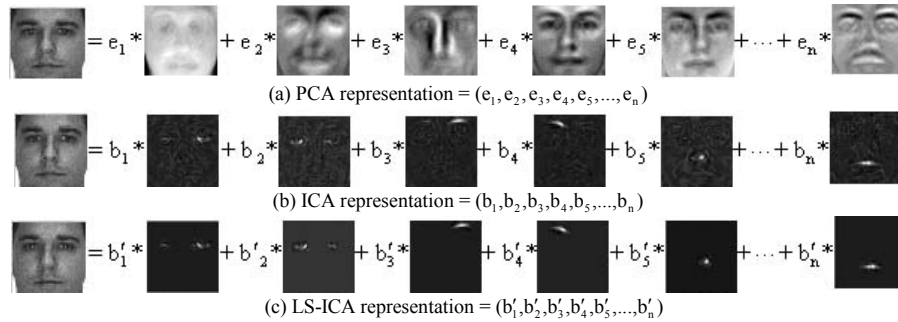

(a) PCA representation = $(e_1, e_2, e_3, e_4, e_5, ..., e_n)$

(b) ICA representation = $(b_1, b_2, b_3, b_4, b_5, ..., b_n)$

(c) LS-ICA representation = $(b'_1, b'_2, b'_3, b'_4, b'_5, ..., b'_n)$

Figure 1. Facial image representations using (a) PCA, (b) ICA and (c) LS-ICA basis images: A face is represented as a linear combination of basis images. The basis images were computed from a set of images randomly selected from the AR database. In the basis images of LS-ICA, non-salient regions of ICA basis images are removed. Using LS-ICA basis images, the concept of "recognition by parts" can be effectively implemented for face recognition.

2 ICA Versus PCA

PCA and ICA are the most widely used subspace projection techniques that project data from a high-dimensional space to a lower-dimensional space [2, 4]. PCA addresses only second-order moments of the input. It is optimal for finding a reduced representation that minimizes the reconstruction error, but it is not optimal for classification. ICA is a generalization of PCA that decorrelates the high-order statistics in addition to the second-order moments. Much of information about

characteristic local structure of facial images is contained in the higher-order statistics of the images. Thus ICA, where the high-order statistics are decorrelated, may provide a more powerful representational basis for face recognition than PCA, where only the second-order statistics are correlated. Figure 2 illustrates PCA and ICA axes for the same 2D distribution. PCA finds an orthogonal set of axes pointing in the directions of maximum covariance in the data, while ICA attempts to place axes pointing in the directions of spatially localized and statistically independent basis vectors [17].

As previously described, global properties of faces may be more easily captured by PCA than ICA. As shown in Figure 1, ICA basis images are more spatially localized and never overlap unlike their PCA counterpart [1]. Since spatially localized features only influence small parts of facial images, ICA based recognition methods are less susceptible to occlusions and local distortions than are global feature based methods such as PCA. We can compute ICA basis images using various algorithms such as InfoMax [3, 10], FastICA [5, 8] and Maximum likelihood [6, 7].
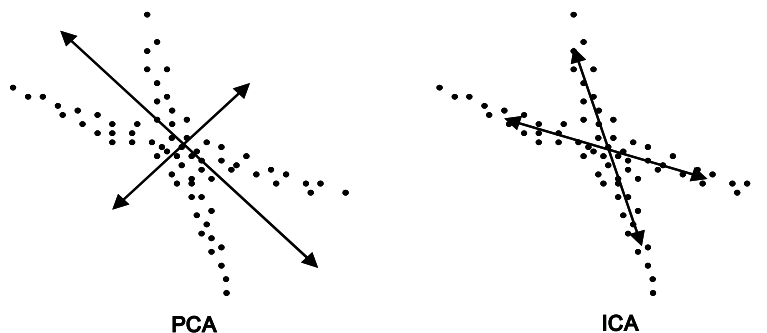

Figure 2. PCA and ICA axes for an identical 2D data distribution [17]

3 The LS-ICA (Locally Salient ICA) Method

The LS-ICA method features the use of new basis images made from ICA basis images that are selected in the decreasing order of class separability. Only salient feature regions are contained in the LS-ICA basis images. As in most algorithms that employ subspace projection, the LS-ICA method computes a projection matrix, off-line from a set of training images. Let W_{ls-ica} denote the projection matrix. The columns of W_{ls-ica} are LS-ICA basis images. During recognition, given an input face image $\mathbf{x}$, it is projected to $\Omega' = W_{ls-sica}^{T}\mathbf{x}$ and classified by comparison with the vectors Ω_T 's that were computed off-line from a set of training images.

Figure 3 shows a block diagram of the method. First, we preprocess training images by applying histogram equalization and scale normalization, where the size of images is adjusted so that they have the same distance between two eyes. Second, we com-
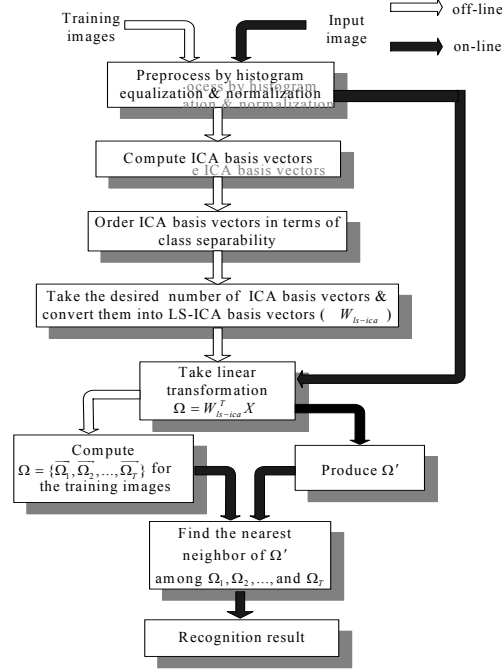
Figure 3. Algorithm overview

pute ICA basis images, using the FastICA algorithm [5, 8]. The FastICA method computes the independent components that become uncorrelated by a whitening process and then maximizes non-Gaussianity of data distribution by using kurtosis maximization [5]. We then compute a measure of class separability, r, for each ICA basis vector and sort the ICA basis vectors in the decreasing order of class separability [3]. To computer r for each ICA basis vector, the between-class variability $\sigma_{between}$ and within-class variability σ_{within} of its corresponding projection coefficients of training images are obtained as follows.

$$\sigma_{between} = \Sigma_i \left(M_i - M \right)^2 \qquad (1)$$

$$\sigma_{within} = \Sigma_i \Sigma_j \left(b_{ij} - M_i \right)^2 \qquad (2)$$

M and M_i are the total mean and the mean of each class, and b_{ij} is the coefficient of the jth training image in class i. The class separability, r, is then defined as the ratio

$$r = \frac{\sigma_{between}}{\sigma_{within}} . \qquad (3)$$

Third, we create LS-ICA basis images from the ICA basis images selected in the decreasing order of the class separability. This way, we can achieve both dimensional reduction and good recognition performance. To create an LS-ICA basis image, we apply a series of operations to its corresponding ICA basis image as shown in Figure 4. In order to detect locally salient regions, we simply find extreme values by thresholding a histogram of pixel values (Figure 4 (b)), followed by the application of morphological operations to find a blob region (Figure 4 (d)). As a result, we get an LS-ICA basis image (Figure 4 (e)) where only pixels in the blob regions have grey values copied from the corresponding pixels in the original ICA image. The values of the rest of the pixels in the image are set to zero. These LS-ICA basis images are used to represent facial images as shown in Figure 1 (c).
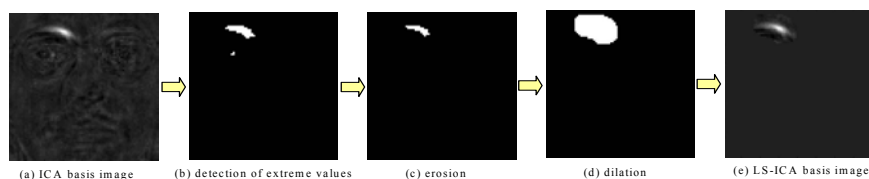

(a) ICA basis image   (b) detection of extreme values   (c) erosion   (d) dilation   (e) LS-ICA basis image

Figure 4. Illustration of creating an LS-ICA basis image

4 Experimental Results

We have used several facial image databases such as AT&T [13], Harvard [14], FERET [15] and AR [16] databases in order to compare the recognition performance of LS-ICA with that of PCA and ICA methods. For fair comparisons with PCA and ICA based methods, PCA and ICA basis images were also used in the decreasing order of class separability, r. In the case of the ICA method, the recognition performance was greater when the basis images were ordered in terms of class separability. However, the PCA method did not show any noticeable performance difference between the ordering in the class separability and the orginal ordering in terms of eigenvalues.

Table 1 lists the number of training and testing images used in each facial image databases for the experiment. Figure 5 shows example images from these databases. In the AT&T database, all the images are taken against a dark homogeneous background and the subjects are in an up-right, frontal position with tolerance for some side movement. In Harvard database, a subject held his/her head steady while being illuminated by a dominant light source. In the FERET database, we have used a subset of the images of subjects under significantly different lighting and facial expression. The AR database contains local distortions and occlusions such as changes in facial expression and sunglasses worn.

All images were converted to 256 gray-level images and background regions were removed. We have also applied histogram equalization to both training and testing images in order to minimize variations of illumination. We have experimented using

thirty different sets of training and testing images for each database. We have computed recognition performances for three different distance measures (L1, L2, cosine) since we are concerned with performance variations independent of the distance measure used [1].

Table 1. Facial databases used in the experiment

Database	The number of total images	The number of persons	The number of training images	The number of testing images
AT&T	400	40	200	200
Harvard	165	5	82(83)	83(82)
FERET	605	127	127	478
AR	800	100	200	600

Figure 5. Example images from AT&T (top left), Harvard (top right), FERET (bottom left) and AR (bottom right) facial databases

Figure 6 shows the recognition performances of PCA, ICA and LS-ICA methods for the four facial databases. The recognition rate of the LS-ICA method was consistently better than that of PCA and ICA methods regardless of distance measures used. ICA also consistently outperformed PCA except the case where the L1 measure was used for the FERET database. What is more interesting is that LS-ICA method performed better than the other methods especially at low dimensions. This property is very important when we need to store facial feature data in a low capacity storing devices such as smart cards and barcodes. To clearly show this, we displayed in Figure 7 the performance improvement of the LS-ICA method over the ICA method. The performance improvement was the greatest in the case of the AR database, as we expected. The AR database contains local distortions and occlusions such as sunglasses worn. The LS-ICA method that only makes use of locally salient information can achieve a higher recognition rate than ordinary ICA methods that are influenced by pixels not belonging to salient regions. The experimental results show that, especially at low dimensions, LS-ICA basis images better represent facial images than ICA basis images.
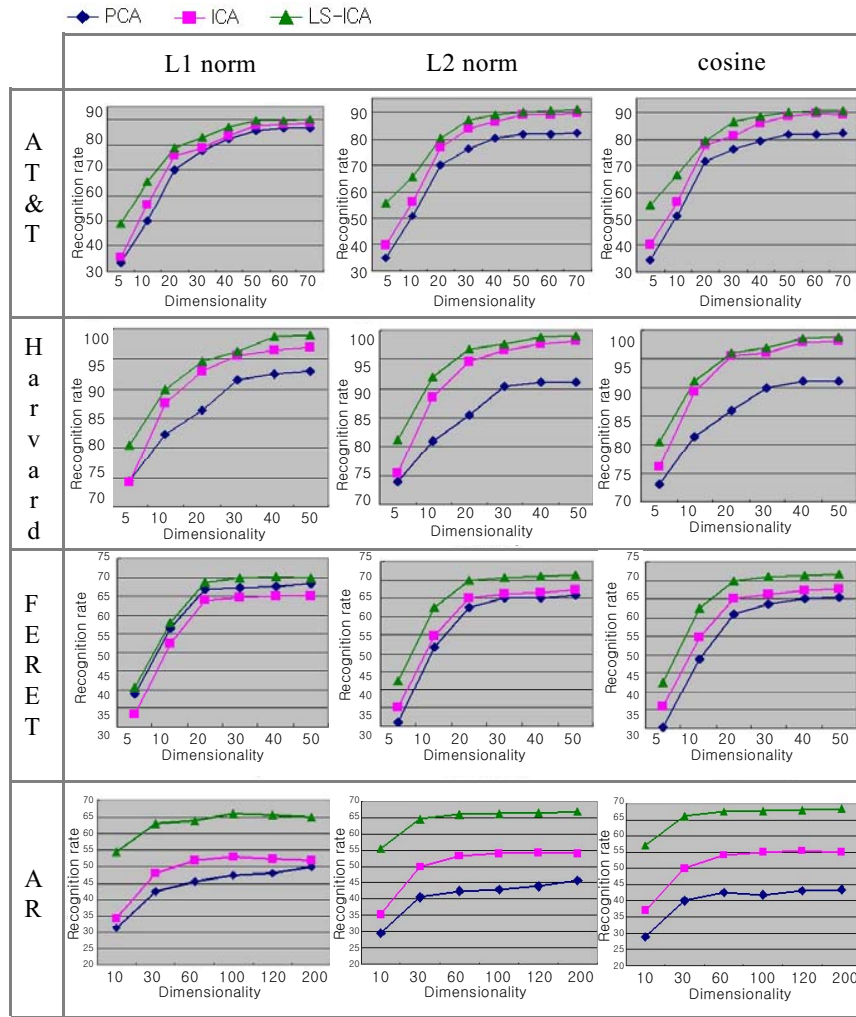
Figure 6. The recognition performance of PCA, ICA and LS-ICA methods for the four facial databases. The recognition rates represent the average performance of the experiment using thirty different sets of training and testing images.

5 Conclusion and Future Research Directions

We have proposed the LS-ICA method that only employs locally salient information in order to maximize the benefit of applying the idea of "recognition by parts" to the problem of face recognition under partial occlusion and local distortion. The performance of the LS-ICA method was consistently better than the ICA method regardless of the distance measures used. As expected, the effect was the greatest in the cases of facial images that have partial occlusions and local distortions such as
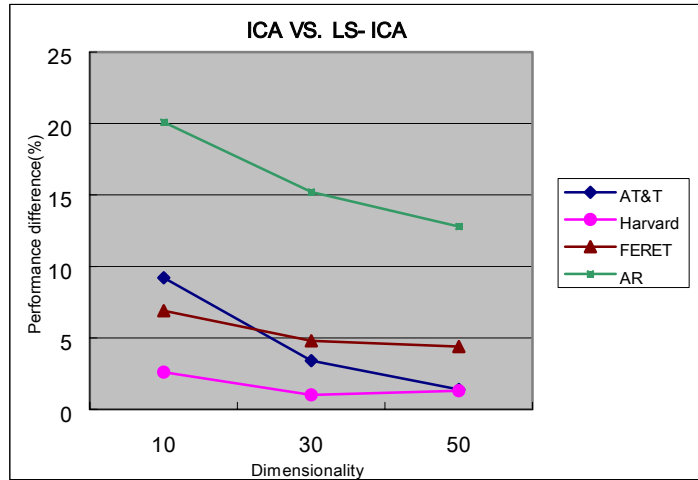
Figure 7. The performance improvement of LS-ICA method over ICA method for the four facial databases. The performance improvement was the greatest in the case of AR database that contains local distortions and occlusions such as sunglasses worn.

changes in facial expression. However, we expect that a combination of the proposed LS-ICA method with a global feature based method such as PCA will yield better recognition rates since face recognition is a holistic process [19]. Further research efforts will be made to develop an optimal method that best combines two methods.

Acknowledgements

References

[1] K. Back, B. A. Draper, J. R. Beveridge, and K. She, "PCA vs ICA: A comparison on the FERET data set," Joint Conference on Information Sciences, Durham, N. C., 2002.
[2] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," Cognitive Neuroscience, vol.3, no.1, pp. 71-86, 1991.
[3] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representations for face recognition," Proceedings of the SPIE, vol. 3299: Conference on Human Vision and Electronic Imaging III, pp. 528-539, 1998.
[4] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," IEEE Transaction on Neural Networks, Vol 13, pp. 1450-1464, 2002.
[5] Aapo Hyvarinen and Erki Oja, "Independent component analysis: a tutorial," http://www.cis.hut.fi/~aapo/papers/IJCNN99_tutorialweb/, 1999.
[6] J. F. Cardoso, "Infomax and Maximum Likelihood for Source Separation," IEEE Letters on Signal Processing, vol. 4, pp. 112-114, 1997.

[7] A. Hyvärinen, "The Fixed-point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis," Neural Processing Letters, vol. 10, pp. 1-5, 1999.

[8] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," Neural Networks, pp. 411-430, 2000.

[9] C. Liu and H. Wechsler, "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition," International Conference on Audio and Video Based Biometric Person Authentication, Washington, D.C., 1999.

[10] A. J. Bell and T. J. Sejnowski, "An information-maximization Approach to Blind Separation and Blind Deconvolution," Neural Computation, vol. 7, pp. 1129-1159, 1995.

[11] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.

[12] K. Fukunaga, "Introduction to statistical pattern recognition," Academic Press, second edition, 1991.

[13] http://www.uk.research.att.com/facedatabase.html.

[14] http://cvc.yale.edu/people/faculty/belhumeur.htm.

[15] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," Pattern Analysis and Machine Intelligence, vol. 22, pp. 1090-1104, 2000.

[16] A. M. Martinez and R. Benavente, "The AR face database," CVC Tech. Report #24, 1998.

[17] M. S. Bartlett, "Face Image Analysis by Unsupervised Learning," Foreword by T. J. Sejnowski, Kluwer International Series on Engineering and Computer Science, Boston: Kluwer Academic Publishers, 2001.

[18] A. P. Pentland, "Recognition by parts," IEEE Proceedings of the First International Conference on Computer Vision, pp. 612—620, 1987.

[19] W. Zhao, R. Chellappa and A. Rosenfeld, "Face Recognition: A Literature Survey," To appear in ACM Computing Surveys.

[20] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, pp.788-791, 1999.

[21] A. J. Bell and T. J. Sejnowski, "The independent components of natural scenes are edge filters," Advance in Neural Information Processing Systems 9, 1997.