Effective Representation Using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion

Jongsun Kim, *Student Member*, *IEEE*,
Jongmoo Choi, *Member*, *IEEE*,
Juneho Yi, *Member*, *IEEE*, and
Matthew Turk, *Senior Member*, *IEEE*

Abstract—The performance of face recognition methods using subspace projection is directly related to the characteristics of their basis images, especially in the cases of local distortion or partial occlusion. In order for a subspace projection method to be robust to local distortion and partial occlusion, the basis images generated by the method should exhibit a part-based local representation. We propose an effective part-based local representation method named locally salient ICA (LS-ICA) method for face recognition that is robust to local distortion and partial occlusion. The LS-ICA method only employs locally salient information from important facial parts in order to maximize the benefit of applying the idea of "recognition by parts." It creates part-based local basis images by imposing additional localization constraint in the process of computing ICA architecture I basis images. We have contrasted the LS-ICA method with other part-based representations such as LNMF (Localized Nonnegative Matrix Factorization) and LFA (Local Feature Analysis). Experimental results show that the LS-ICA method performs better than PCA, ICA architecture I, ICA architecture II, LFA, and LNMF methods, especially in the cases of partial occlusions and local distortions.

Index Terms—Face recognition, part-based local representation, ICA, LS-ICA.

---

1 INTRODUCTION

OVER the past 10 years, canonical subspace projection techniques such as PCA, ICA, and FLD have been widely used in the face recognition research [1], [2], [3], [4], [5]. These techniques represent a face as a linear combination of low rank basis images. They employ feature vectors consisting of coefficients that are obtained by simply projecting facial images onto their basis images. In order for a subspace projection-based method to be robust to partial occlusions and local distortions, its basis images should effectively realize a part-based local representation. Local representation provides robustness to partial occlusions and local distortions because successful face recognition can be achieved by representing some important facial parts that correspond to feature regions such as eyes, eyebrows, nose, and lips. This "recognition by parts" paradigm [11] has been popular in the object recognition research because the approach can be successfully applied to the problem of object recognition with occlusion.

Facial image representations based on different basis images are illustrated in Fig. 1. ICA can be applied to face recognition in two different representations: ICA architecture I and II [2]. Please refer to Section 2.1 for more description about these two representations. PCA and ICA architecture II basis images, as shown in Figs. 1a and 1b, respectively, display global properties in the sense that they assign significant weights to potentially all the pixels. This accords with the fact that PCA basis images are just scaled versions of global Fourier filters [13]. In contrast, ICA architecture I basis images are

---

- *J. Kim, J.M. Choi, and J. Yi are with the School of Information & Communication Engineering, Biometrics Engineering Research Center, Sungkyunkwan University, Korea.*
  E-mail: {jskim, jmchoi, jhyi}@ece.skku.ac.kr.
- *M. Turk is with the Computer Science Department, University of California, Santa Barbara 93106. E-mail: mturk@cs.ucsb.edu.*

spatially more localized. This local property of ICA architecture I basis images makes the performance of ICA architecture I based recognition methods robust to partial occlusions and local distortions, such as local changes in facial expression, because spatially local features only influence small parts of facial images. However, ICA architecture I 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 nonzero weight values. These pixel values in nonsalient regions would appear as noise and contribute to the degradation of the recognition.

Among representative part-based local representations are Local Feature Analysis (LFA) [6] and Local Nonnegative Matrix Factorization (LNMF) [14] methods. The LFA method extracts local features based on second-order statistics. However, basis images from the LFA representation are not perfectly localized as shown in Fig. 1f. Thus, pixels in nonsalient regions degrade the recognition performance in the case of partial occlusions and local distortions. Recently, the LNMF method was reported in the literature, which led to an improved version of Nonnegative Matrix Factorization (NMF) [12]. In the LNMF method, locality constraints are imposed on the factorized matrices from NMF in terms of sparsity in matrix components. They successfully localized the components in basis images. However, the locality constraints do not guarantee that meaningful facial features should be localized in their basis images. As an example of the LNMF representation in Fig. 1e illustrates, some LNMF basis images represent regions such as cheek, forehead, and jaw that are not discriminant features for face recognition.

We propose new basis images based on ICA architecture I, called LS-ICA (locally salient ICA) basis images, where only locally salient feature regions are retained. The idea of "recognition by parts" can be effectively realized for face recognition using LS-ICA basis images since each LS-ICA basis image represents only locally salient regions. These regions correspond to important facial feature regions such as eyes, eyebrows, nose, and lips. Note that ICA architecture I produces basis images that are localized edge filters [13], and they correspond to meaningful facial feature regions.

Our method for face recognition is characterized by two ideas: The first is the creation of the LS-ICA basis images using a modified version of Kurtosis maximization to remove residual nonlocal modulation in ICA architecture I basis images; these are used to represent faces. The second idea is to use LS-ICA basis images in the decreasing order of class separability so as to maximize the recognition performance. Experimental results show that LS-ICA performs better than PCA, ICA architecture I, ICA architecture II, LFA, and LNMF, especially in the cases of partial occlusions and local distortions such as local changes in facial expression.

The rest of this paper is organized as follows: Section 2 briefly describes the ICA, LFA, and LNMF methods that are most relevant to our research. We present the proposed LS-ICA method in Section 3. Section 4 gives experimental results.

2 RELATED WORK

2.1 ICA (Independent Component Analysis)

ICA is a widely used subspace projection technique that projects data from a high-dimensional space to a lower-dimensional space [2], [3], [4]. This technique is a generalization of PCA that decorrelates the high-order statistics in addition to the second-order moments. In this research, we compute ICA basis images using the FastICA algorithm [3] while other methods such as InfoMax [2] or Maximum likelihood [4] can also be employed.

The FastICA method computes independent components by maximizing non-Gaussianity of whitened data distribution using a kurtosis maximization process. The kurtosis measures the non-Gaussianity and the sparseness of the face representations [13]. The FastICA algorithm is briefly described as follows: Let $\mathbf{S}$ be the vectors of unknown source signals and $\mathbf{x}$ be vectors of observed mixtures. If $\mathbf{A}$ is an unknown mixing matrix, then the mixing model can be written as $\mathbf{X} = \mathbf{A}\mathbf{S}$. The task is to estimate the independent
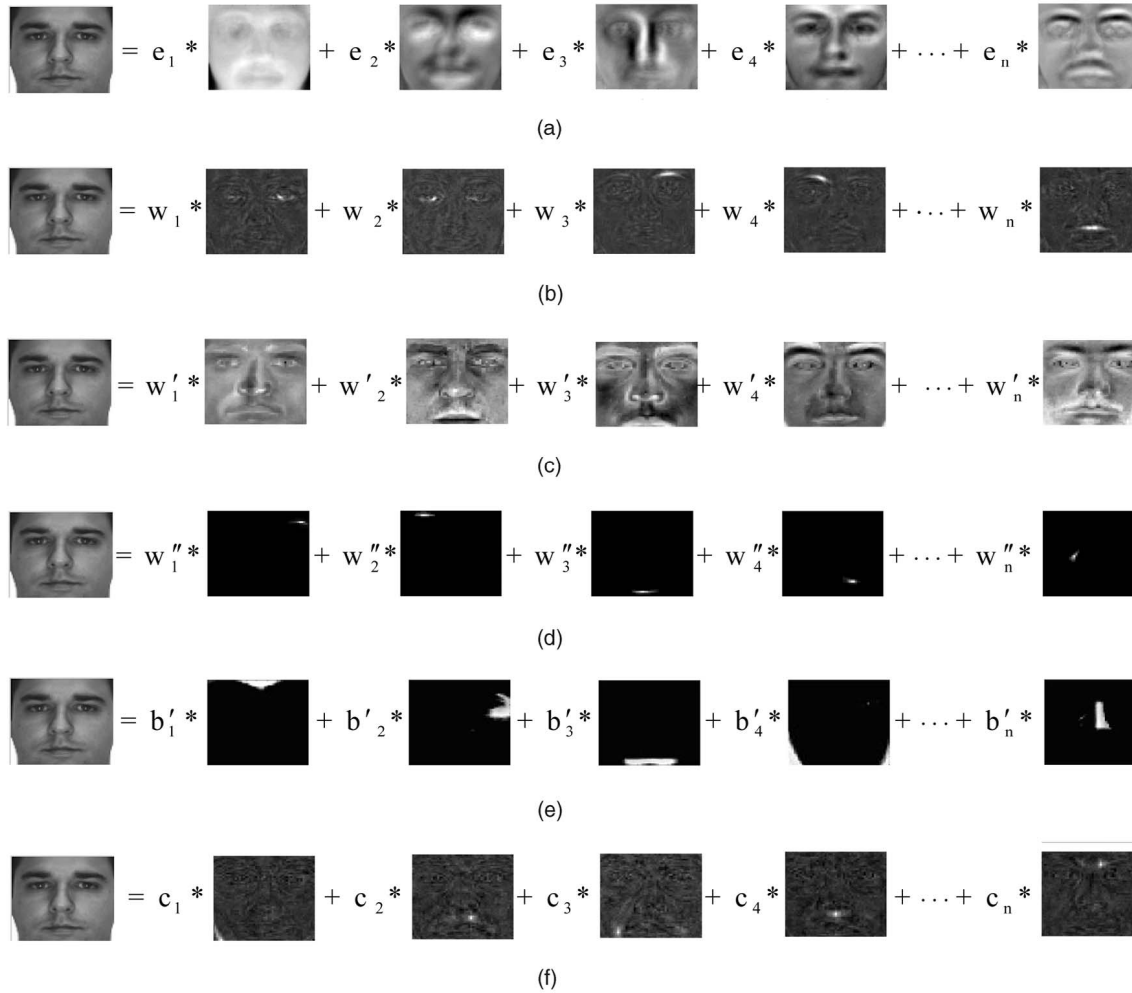
(a)

(b)

(c)

(d)

(e)

(f)

Fig. 1. Facial image representations using (a) PCA, (b) ICA architecture I, (c) ICA architecture II, (d) proposed LS-ICA, (e) LNMF, and (f) LFA 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. Using LS-ICA basis images, the concept of "recognition by parts" can be effectively implemented for face recognition.

source signals $\mathbf{U}$ by computing the separating matrix $\mathbf{W}$ that corresponds to the mixing matrix $\mathbf{A}$ using the following relation:

$$\mathbf{U} = \mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{A}\mathbf{S}. \qquad (1)$$

First, the observed samples are whitened. Let us denote the whitened samples by $\mathbf{Z}$. Then, we search for the matrix such that the linear projection of the whitened samples by the matrix $\mathbf{W}$ has maximum non-Gaussianity of data distribution. The kurtosis of $\mathbf{U}_i = \mathbf{W}_i^T\mathbf{Z}$ is computed as in (2) and the separating vector W_i is obtained by maximizing the kurtosis [3].

$$kurt(\mathbf{U}_i) = \left| E\left\{ (\mathbf{U}_i)^4 \right\} - 3\left(E\left\{ (\mathbf{U}_i)^2 \right\} \right)^2 \right|. \qquad (2)$$

ICA can be applied to face recognition in two different architectures [2]. The ICA architecture I considers the input face images, $\mathbf{X}$, as a linear combination of statistically independent basis images, $\mathbf{S}$, combined by an unknown matrix, $\mathbf{A}$. The coefficients obtained by projecting input images onto the statistically independent basis images are not statistically independent. On the other hand, the ICA architecture II finds statistically independent coefficients that represent input images. The ICA architecture II basis images display global properties as shown in Fig. 1c. Since the kurtosis maximization yields sparseness of basis images, the ICA architecture I basis images are spatially localized edge filters [13]. However, they do not display perfectly local characteristics in the sense that pixels that do not belong to locally salient feature regions

still have some nonzero weight values. These pixel values would contribute to the degradation of the recognition performance in the case of local distortion and partial occlusion.

2.2 LFA (Local Feature Analysis)

LFA defines a set of local topographic kernels that are derived from the principal component eigenvectors $\mathbf{E}$ and coefficients $\mathbf{D}$ according to covariance matrix $\mathbf{S}$ using the following equation:

$$\mathbf{K} = \mathbf{E}\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T, \text{ where } \mathbf{D}^{-\frac{1}{2}} = diag\left(\frac{1}{\sqrt{\lambda_i}}\right) \ i = 1, \ldots, p. \qquad (3)$$

λ_is are eigenvalues of the covariance matrix, $\mathbf{S}$. The rows of $\mathbf{K}$ contain the kernels. The kernels are topographic in that they are indexed spatially. The number of kernels corresponds to the number of pixels in an image. The LFA uses a sparsification algorithm in order to reduce the dimensionality of the representation. The algorithm iteratively selects kernels that have the largest mean reconstruction error. We are concerned with basis images from the LFA method. They are not perfectly localized and pixels in nonsalient regions contribute to the degradation of the recognition performance.

2.3 LNMF (Local Nonnegative Matrix Factorization)

The LNMF method [14] is a technique that improves the standard NMF method [12]. It is aimed at learning spatially localized, parts-based subspace representation of basis images by imposing additional constraints on the NMF basis. The following objective function is used to compute LNMF basis images:

Fig. 2. Example images from AT&T (left), AR (middle), and FERET (right) facial databases.

TABLE 1
FERET Database Used in the Experiment

Two letter code	Description	The number of images	The number of subjects	Experiment using 'ba'-'bj' sets	Experiment using 'ba'-'bk' sets
fa	frontal images with neutral facial expression	1758	1009	Training images (1758)	
ba	frontal images which is entirely analogous to 'fa'	200	200	Gallery (100)	Gallery (100)
bj	frontal images with expressions different from 'ba'	200	200	Probe (100)	
bk	illumination different from 'ba'	194	194		Probe (100)

$$D(X||BH) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(X_{ij} \log \frac{X_{ij}}{[BH]_{ij}} - X_{ij} + [BH]_{ij} + \alpha U_{ij} \right) - \beta \sum_{i} V_{ii}, \tag{4}$$

where $\alpha, \beta > 0$ are some constants, $U = B^T B$, $V = HH^T$, and $B, H \geq 0$ means that all entries of basis images B and coefficients H are nonnegative. The minimization of $U = B^T B$ imposes both maximum sparsity in H and minimum redundancy between different bases. On the other hand, by maximizing $\sum_i V_{ii}$, basis components that carry much information about the training images are retained. Refer to [14] for further justification of the objective function. The LNMF update rule for H uses square root as in (5) to satisfy the additional constraints. The update for B is identical to that of NMF [15].

$$H_{aj}^{(t+1)} = \sqrt{H_{aj}^{(t)} \sum_{i} B_{ai}^{T(t)} \frac{X_{ij}}{(B^{(t)} H^{(t)})_{ij}}}. \tag{5}$$

Fig. 1e shows an example of the LNMF representation. As described earlier, the additional constraints only focus on locality and it is not necessarily guaranteed that meaningful facial features are localized in their basis images.

3 THE LS-ICA (LOCALLY SALIENT ICA) METHOD

The computation of LS-ICA basis images consists of two steps: The first step is concerned with the creation of part-based local basis images based on ICA architecture I. The second is to order the basis images obtained in the order of class separability for good recognition performance.

The LS-ICA method creates part-based local basis images by imposing additional localization constraint in the process of the kurtosis maximization. The solution at each iteration step is weighted so that it becomes sparser by only emphasizing large pixel values. Localization emerges from this sparsification. Let $\mathbf{u}$ be a solution vector at an iteration step, we can define a weighted solution as $\mathbf{b}$, where $b_i = |u_i|^{\alpha} u_i$ and $\mathbf{b} = \mathbf{b}/\|\mathbf{b}\|$. $\alpha > 1$ is a small constant. The kurtosis is maximized in terms of $\mathbf{b}$ instead of $\mathbf{u}$ as in (6).

$$kurt(\mathbf{b}) = \left| E\left\{ (\mathbf{b})^4 \right\} - 3 \left(E\left\{ (\mathbf{b})^2 \right\} \right)^2 \right|. \tag{6}$$

A solution to the above function can be found by using the following update rules:

$$\mathbf{w}^{(t+1)} = E\left\{ |u_i|^{\alpha} \mathbf{Z} \left(|u_i|^{\alpha} \mathbf{w}^{T(t)} \mathbf{Z} \right)^3 \right\}, \tag{7}$$

where $\mathbf{w}$ is a separating vector and $\mathbf{Z}$ contains whitened image samples. The resulting basis image is $b_i = |u_i|^{\alpha} (\mathbf{w}^T \mathbf{Z})_i$. As an alternative method, we would like to point out that other simple operations such as morphological operations can be employed to detect salient regions from ICA architecture I basis images.

We then compute a measure of class separability, r, for each LS-ICA basis vector and sort the LS-ICA basis vectors in the decreasing order of class separability [2]. To compute r for each LS-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:

$$\sigma_{between} = \Sigma_i (M_i - M)^2, \tag{8}$$

$$\sigma_{within} = \Sigma_i \Sigma_j (h_{ij} - M_i)^2. \tag{9}$$

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

$$r = \frac{\sigma_{between}}{\sigma_{within}}. \tag{10}$$

We then create new LS-ICA basis images from the LS-ICA basis images selected in the decreasing order of the class separability. This way, we can achieve both dimensionality reduction and good recognition performance. The LS-ICA representation is based on the idea of localized edge filters that come from ICA basis images. The resulting basis images contain localized facial features that are discriminant for face recognition.

We have calculated pairwise mutual information for the basis images from (2) versus (6) using the following joint entropy used in [16]:

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \tag{11}$$

where $H(X, Y)$ is the joint entropy of X and Y. The mean values of the pairwise mutual information for (2) versus (6) are 0.0227 and 0.0013, respectively. This experimentally shows that the sparsification also enhances the independence solution.

4 EXPERIMENTAL RESULTS

We have used several facial image databases such as the FERET [8], AR [9], and AT&T [10] databases in order to compare the recognition

Fig. 3. Example AT&T images having random occluding patches of sizes (from left to right) 10 x 10, 20 x 20, and 30 x 30.

performance of LS-ICA with that of PCA, ICA architecture I, ICA architecture II, LFA, and LNMF methods. Fig. 2 shows example images from these databases. For fair comparisons with the above methods, their basis images were also used in the decreasing order of class separability, r. We have computed recognition performances for three different distance measures (L1, L2, cosine) to see if there is any performance variation depending on the distance measure used [7].

We have used subsets of the FERET images under significantly different lighting and facial expression. The whole set of images, U, used in the experiment, consists of four subsets called "fa," "ba," "bj," and "bk," as summarized in Table 1. For the experiment, we have divided the whole set U into training set (T), gallery set (G), and probe set (P). No one within the training set (T) is included in the gallery or the probe sets. The AR database contains 800 frontal facial images from 100 subjects. The number of images used for training and testing are 200 and 600 images, respectively. Test images contain local distortions and occlusions such as changes in facial expression and sunglasses worn.

In order to show the performance comparisons under occlusion, we have used the AT&T database. It consists of 400 images of 40 persons, 10 images per person. The images are taken against a dark homogeneous background and the subjects are in an upright, frontal position with tolerance for some side movement. A set of 10 images for each person is randomly partitioned into five training images and five testing images. The occlusion is simulated in an image by using a white patch of size $s \times s$ with $s \in \{10, 20, 30\}$ at a random location as shown in Fig. 3.

Figs. 4 and 5 show the recognition performances of PCA, ICA architecture I, ICA architecture II, LNMF, LFA, and LS-ICA methods for the three facial databases. The recognition rate of the LS-ICA method was consistently better than that of PCA, ICA architecture I, ICA architecture II, LNMF, and LFA methods regardless of the distance measures used. In Fig. 4, ICA architecture I also consistently outperformed PCA except the case where the L1 measure was used for the FERET database. This also accords with the experimental results of Draper et al. [7]. Fig. 5 compares the six representations under varying degrees of occlusion, in terms of the recognition accuracies versus the size $s \times s$ of occluding patch for $s \in \{10, 20, 30\}$. The LS-ICA and LNMF methods performed better than the other methods under partial occlusion, especially as the patch size increases. The LFA and ICA architecture I methods appear influenced by pixels not belonging to salient regions. The ICA architecture I method showed better performance than the LFA method. Bartlett [16] have also reported that the ICA representation performs better than the LFA representation in cases of facial expression analysis and face recognition. This experimentally shows
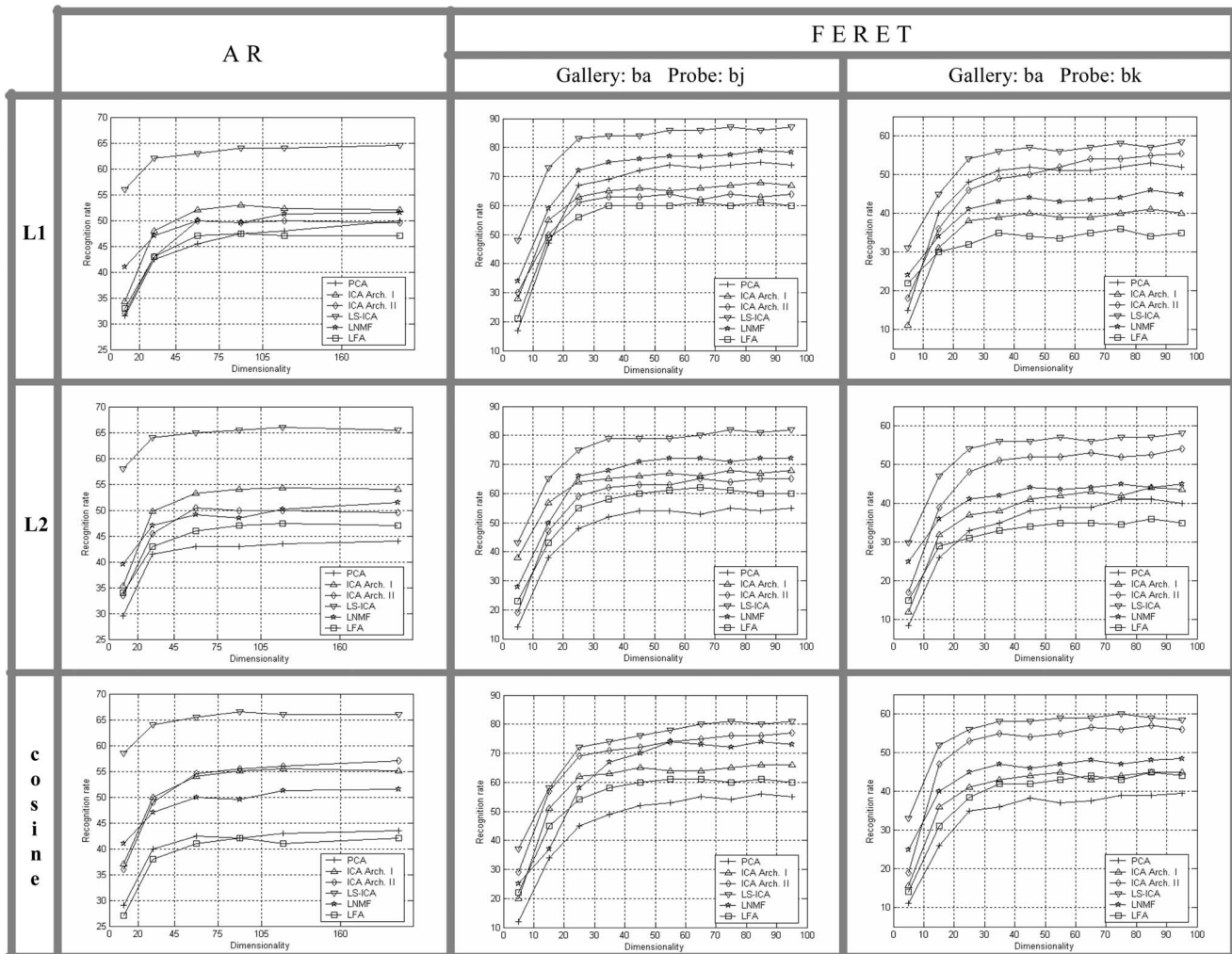

Fig. 4. The recognition performance of PCA, ICA1, ICA2, LNMF, LFA, and LS-ICA methods for the AR and FERET ("ba"-"bj" set and "ba"-"bk" set) facial databases.
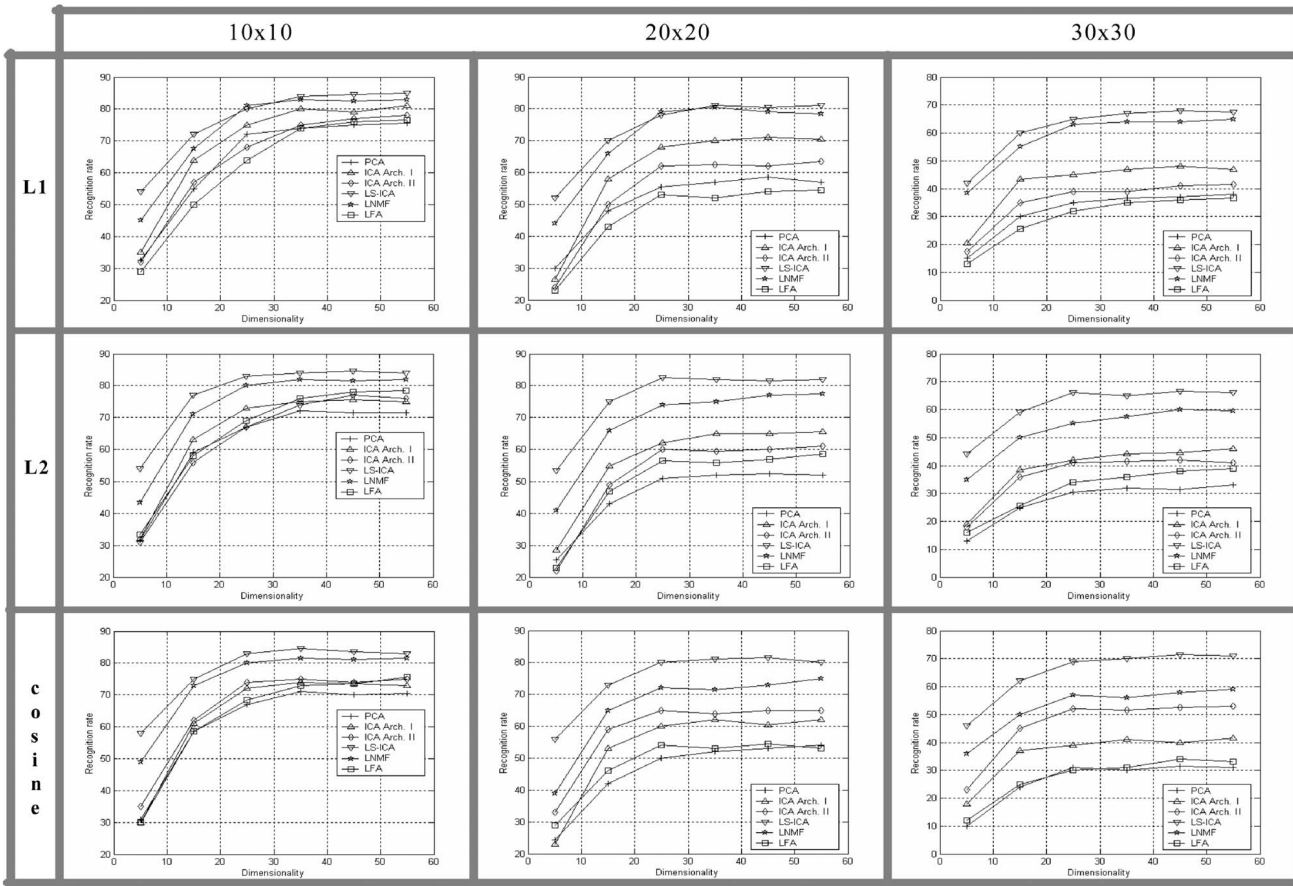
Fig. 5. Recognition performance versus the size (in 10 x 10, 20 x 20, and 30 x 30) of occluding patches for PCA, ICA1, ICA2, LNMF, LFA, and the proposed LS-ICA method for the AT&T database.

that the ICA architecture I better represents some important facial parts than the LFA method. The LS-ICA method that only makes use of locally salient information from important facial parts achieved higher recognition rates than the LNMF method. We can see that the LS-ICA representation is an effective part-based local representation for face recognition robust to local distortion and partial occlusion.

5   CONCLUSION

We have proposed the LS-ICA method that only employs locally salient information from important facial parts 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 other representative local representation based methods 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 changes in facial expression.

REFERENCES

[1]   M.A. Turk and A.P. Pentland, "Eigenfaces for Recognition," *Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.
[2]   M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Trans. Neural Networks,* vol. 13, no. 6, pp. 1450-1464, 2002.
[3]   A. Hyvarinen and E. Oja, "Independent Component Analysis: A Tutorial," http://www.cis.hut.fi/~aapo/papers/IJCNN99_tutorialweb/, 1999.
[4]   A. Hyvärinen, "The Fixed-Point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis," *Neural Processing Letters,* vol. 10, pp. 1-5, 1999.
[5]   P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, 1999.
[6]   P. Penev and J. Atick, "Local Feature Analysis: A General Statistical Theory for Object Representation," *Network: Computation in Neural Systems,* vol. 7, no. 3, pp. 477-500, 1996.
[7]   B.A. Draper, K. Baek, M.S. Bartlett, and J.R. Beveridge, "Recognizing Faces with PCA and ICA," *Computer Vision and Image Understanding,* vol. 91, no. 1, pp. 115-137, 2003.
[8]   P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, 2000.
[9]   A.M. Martinez and R. Benavente, "The AR Face Database," *CVC Tech,* 1998.
[10]   http://www.uk.research.att.com/facedatabase.html, 1994.
[11]   A.P. Pentland, "Recognition by Parts," *IEEE Proc. First Int'l Conf. Computer Vision,* pp. 612-620, 1987.
[12]   D.D. Lee and H.S. Seung, "Learning the Parts of Objects by Nonnegative Matrix Factorization," *Nature,* vol. 401, pp. 788-791, 1999.
[13]   A.J. Bell and T.J. Sejnowski, "The Independent Components of Natural Scenes Are Edge Filters," *Vision Research,* vol. 37, no. 23, pp. 3327-3338, 1997.
[14]   S.Z. Li, X.W. Hou, and H.J. Zhang, "Learning Spatially Localized, Parts-Based Representation," *Computer Vision and Pattern Recognition,* vol. 1, pp. 207-212, 2001.
[15]   S. Wild, J. Curry, and A. Dougherty, "Motivating Non-Negative Matrix Factorizations," *Proc. Eighth SIAM Conf. Applied Linear Algebra,* July 2003.
[16]   M.S. Bartlett, *Face Image Analysis by Unsupervised Learning.* Kluwer Academic, 2001.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.