

# Combining ICA Representations for Recognizing Faces

Ashraf Y. A. Maghari

Faculty of Information Technology  
Islamic University of Gaza  
Gaza, Palestine  
amaghari@iugaza.edu.ps

**Abstract**— Independent Component Analysis (ICA) is a generalization of Principal Component Analysis (PCA), and it looks for components that are both statistically independent and non-Gaussian. ICA is sensitive to high-order statistic and it expected to outperform PCA in finding better basis images. Moreover, with face recognition, high-order relationships among pixels may have more important information than those of pairwise relationships on which base images found by PCA depend. Two different representations can be applied by ICA; ICA architecture I and ICA architecture II. A new classifier that combines the two ICA architectures is proposed for face recognition. By the new classifier, the similarity measure vector was employed in which the similarity measure vectors for both ICA representations were resorted in descending order and then integrated by merging the corresponding values of each vector. The new classifier was performed on face images in the AR Face Database. Cumulative Match Characteristic was taken as a measure for evaluating the performance of the new classifier with illumination variation, expression, and Occlusion. The proposed classifier outperforms both ICA architectures in all cases especially in later ranks.

**Keywords**— ICA; PCA; face recognition; ICA representations

## I. INTRODUCTION

Human face recognition is a very important issue in the field of Computer Vision. It has gained a lot of attention during the last decades [1]. Current 2D face recognition systems can achieve good performance in constrained environments. However, they still encounter difficulties in handling large amounts of facial variations [2]. Extensive research is ongoing to make face recognition systems robust to typical operational environments where uncertainties such as occlusion, illumination and other variations are common [3]. In the last decades, a number of biometric face recognition algorithms have been proposed by computer scientist, neuroscientists and psychologist's efforts. The computer scientists seek to develop methods for face recognition whereas the psychologists and neuroscientists work on biological perception of human face recognition process i.e. face recognition is done holistically or feature analysis etc. [4]. There are two important matters in face recognition algorithms, feature representation and classification based on features. Based on feature representation; face recognition methods can be classified into two groups i.e. face and constituent. Face-based methods (appearance based technique) use raw information face pixel, whereas constituent based approaches use the relationships between face features i.e. nose, lips, and eyes. Compared to face-based methods, the constituent-based methods are more flexible but the performance depend on features [4].

Bio-inspired evolutionary search was presented in [4] where a constituent-based method was employed with fixed fiducial points extracted from the face image. GA was then used to search best feature combination that gives minimum training error. However, in this work, we used the appearance based technique so as the image is considered as 2D pattern. Among appearance based representation, PCA based method is one of the most powerful methods successfully applied in face recognition [4]. Principle Component Analysis (PCA) is a popular unsupervised statistical method used for dimension reduction and face recognition. The goal of PCA is to find a set of basis images so that the PCA coefficients are linearly independent. The performance of PCA depends on the task statement, the subspace distance metric, and the number of subspace dimensions retained. Independent Component Analysis (ICA) can be seen as a generational of PCA and its basic idea is to represent a set of random variables using basis functions, where the components are statistically independent or as independent as possible [5].

There are many algorithms used for performing ICA [6], [7]. In this study, the information maximization learning algorithm developed by [8], [9] is employed. The algorithm was developed from the principle of optimal information transfer in neural networks with sigmoidal transfer functions. It has been shown that this algorithm has proven successful for separating electroencephalogram (EEG) signals [10] and functional magnetic resonance imaging (fMRI) signals [11]. The non-locality of the learning algorithm is interesting when

biological significance of the learned filters is considered [9]. However, the biological plausibility of the algorithm is limited when the learning rule is nonlocal. The performance of ICA depends on the task, the algorithm used to approximate ICA, and the number of subspace dimensions retained. There are two different approaches of ICA to face recognition. ICA can be applied so as to treat images as random variables and pixels as observations, or to treat pixels as random variables and images as observations. In consistence with [12], [13], we refer to these two alternatives as ICA architecture I and architecture II, respectively.

The two architectures of ICA were performed by Bartlett et al. [13] on face images in the FERET database. They developed a classifier that combine the ICA representations to give better performance in comparison with the two architectures. The combined classifier was employed in which the similarity between the test image and the gallery image has been defined as the summation of the similarity measure of both ICA1 and ICA2 (ICA1+2).

In this study, we developed a new classifier that combine both ICA1 and ICA2 in different way. The classifier was adopted in which the similarity measure vectors for both ICA1 and ICA2 were resorted in descending order and then integrated by merging the corresponding values of the two vectors to reconstruct a new similarity measure vector. AR database has been used to train and test the new combined classifier.

This paper is organized as follows: Section II gives brief introduction on ICA. In Section III, the proposed classifier will be demonstrated, whereas in Section IV the Experiment results and discussion are reported. In the last section, the paper is summarized.

## II. INDEPENDENT COMPONENT ANALYSIS (ICA)

ICA is a generalized form of Principal Component Analysis (PCA), which is able to separate independent sources linearly mixed in several signals. The Information Maximization algorithm proposed by Bell and Sejnowski [8] has been used by [13] to perform ICA. The algorithm was derived from the principle of optimal information transfer in neurons with sigmoidal transfer functions. It finds the matrix that represents the statistically independent vectors of the face images. Given that  $X$  is an input matrix, in which each row represent an image, then  $U = WX$  is the output matrix of independent representation of the images, where  $W$  is an invertible weight matrix such that  $X' = W^{-1}U$ . The weight matrix,  $W$ , was found through an unsupervised learning algorithm that maximizes the mutual information between the input and the output of the nonlinear transformation [8]. The algorithm has proven successful for separating randomly mixed auditory signals and has been applied to natural scenes.

Regardless of which algorithm is used to compute ICA, there are two fundamentally different ways to apply ICA to face recognition. In Architecture I, the images are considered as random variables and the pixels as outcomes, while

Architecture II treaded the pixels as random variables and the images as outcomes.

### A. Architecture I(ICA1): Statistically Independent Basis Images

In this architecture, the face images are variables and the pixel values provide observations for the variables. The images are organized as a data matrix  $X^{m \times n}$ , where  $m$  is the number of training images and  $n$  is the number of pixels. The input face images in  $X$  are considered to be a linear mixture of statistically independent basis images  $S$  combined by an unknown mixing

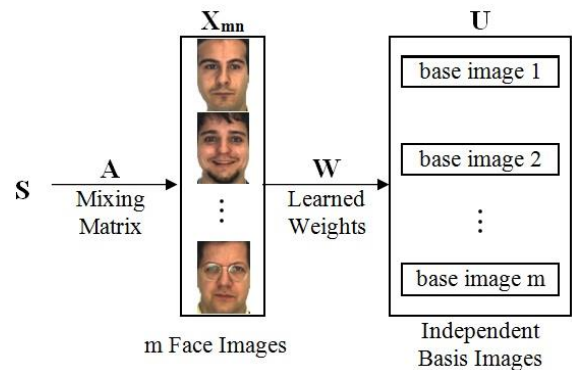


Fig. 1. Finding statistically independent basis images

matrix  $A$ . The weight matrix  $W$  is learnt by ICA algorithm such that the rows of  $U=WX$  are as statistically independent as possible. The source images estimated by the rows of  $U$  are then used as basis images to represent faces (Fig. 1). The source separation, therefore, is performed in face space. Projecting the input images onto the learned weight vectors produces the independent basis images. The compressed representation of a face image is a vector of coefficients used for linearly combining the independent basis images to generate the image. Bartlett et al. [13] first apply PCA to project the data into a subspace of dimension  $m$  to control the number of independent components produced by ICA. The Information Maximization algorithm [8] is then applied to the eigenvectors to minimize the statistical dependence among the resulting basis images. Using PCA as preprocessing allows ICA to create subspaces of size  $m$  for any  $m$ . Liu and Wechsler [14] argue that applying ICA on the projected data enhances the performance by

- Discarding small trailing eigenvalues before whitening.
- Reducing the computational complexity by minimizing pair-wise dependences.

Let  $X$  be the matrix containing the zero-mean images and  $V$  be the matrix containing (in its columns) the first  $m$  eigenfaces that have the  $m$  highest eigenvalues. Then the PC representation of  $X$  is defined as  $R_m = XV$ . ICA is performed on  $V$  to produce the independent basis images matrix  $U$ , weight matrix  $W$  and sphering matrix  $W_z$ . The IC representation of the face images based on the set of  $m$  statistically independent basis images  $U$  is given by the rows of the matrix

$$F = R_m W_I^{-1}, \quad (1)$$

where  $W_I = W W_z$  such that  $W_I V^T = U$ . The representation for test images is obtained by the following equation

$$F_{test} = R_{test} W_I^{-1}, \quad (2)$$

where  $R_{test} = X_{test} V$ .

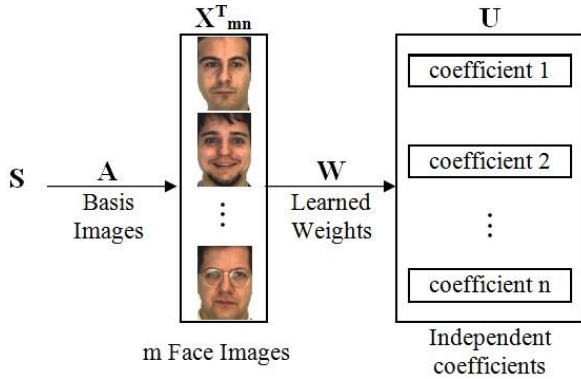


Fig. 2. Finding statistically independent coefficients

### B. Architecture II(ICA2): A Factorial Face Code

In Architecture II the coefficients that represent the images are represented by ICA such that they are statistically independent. The data matrix  $X$  is organized so that rows represent different pixels and columns represent different images. The inverse matrix  $A = W_I^{-1}$  contains the basis images in its columns. The statistically independent source coefficients (ICA representations) in  $S$  that comprise the input images are recovered in the columns of  $U = W_I X$  (Fig. 2). Each column of  $U$  contains the coefficients of basis images in  $A$  for reconstructing each image in  $X$  such that each face  $x = u_1 * a_1 + u_2 * a_2 + \dots + u_n * a_n$  where the column of  $U$  ( $u_1, u_2, \dots, u_n$ ) is the ICA factorial representation. The representational code for test images is obtained by  $W_I X_{test} = U_{test}$ , where  $X_{test}$  is the zero-mean matrix of test images, and  $W_I$  is the weight matrix found by performing ICA on the training images.

### C. Combined ICA Reconstruction System

A combined classifier "ICA1+2" was employed by [13] in which the summation of the similarity measures  $c_1$  and  $c_2$  of ICA1 and ICA2 respectively has been used as the similarity between test images and gallery (training) images. The similarity measure was evaluated by the cosine of the angle between the coefficient vector of the test image and the coefficient vectors of the training images

$$c = \frac{b_{test} \cdot b_{train}}{\|b_{test}\| \cdot \|b_{train}\|} \quad (3)$$

The cosines was used as a similarity measure because ICA performs significantly better than when using Euclidean distance [13]. The combined classifier was defined as  $c_1 + c_2$ , where  $c_1$  and  $c_2$  correspond to the similarity measure  $c$  in (3). According to [13] the classifier that combined the two ICA representations improved the performance of both ICA1 and ICA2 and outperformed PCA on all test sets. However, and according to extensive experiments, we found that many faces can be recognized using ICA1 while it cannot be recognized accurately using ICA2 and vice versa. Accordingly, we proposed a new classifier (Merging Distances) to utilize the advantages of both ICA architectures. It also utilizes the improvement of ICA1+2 in which the smallest similarity measure value resulted by ICA1+2 is integrated in the first rank of the new classifier.

### III. THE PROPOSED CLASSIFIER (MERGING DISTANCES)

The proposed classifier was adopted in which the similarity measure vectors for both ICA1 and ICA2 were resorted in descending order and then integrated by merging the corresponding values of the two vectors to reconstruct a new similarity measure vector. The two similarity measure vectors  $v_{c1}$  and  $v_{c2}$  were combined as follows

$$v_c(i \times 2 - 1) = v_{c1}(i), \quad v_c(i \times 2) = v_{c2}(i) \quad (4)$$

where  $v_{c1}$  and  $v_{c2}$  is the similarity measure vectors for ICA1 and ICA2, respectively,  $i = 2, \dots, n/2$ ,  $n$  is the number of ranks for both ICA1 and ICA2. By this way we ensure that the smallest similarity measure values of  $v_{c1}$  and  $v_{c2}$  will be in the early ranks of recognition which means that improvement in later ranks is guaranteed. In our experiments  $n$  has been chosen to be 30 ranks. To further utilizes the improvement of the combined classifier ICA1+2 [13], the smallest similarity measure value resulted by ICA1+2 is added in the first rank of the new similarity measure vector  $v_c$ . Consequently, the proposed classifier has the advantages of ICA1, ICA2, and ICA1+2. The combined classifier (Merging Distances) is deployed on  $v_c$  as a similarity measure.

### IV. EXPERIMENTS AND DISCUSSION

This paper proposed a new classifier that combines the similarity measures of both ICA architectures ICA1 and ICA2 for face recognition across facial variation. We evaluated the performance of the proposed classifier compared with that of ICA1, ICA2, and ICA1+2. The AR Database has been used in the evaluation since it contains faces with different expression, illumination condition, and occlusions. 100 preprocessed subjects are used for training and testing. The first three images of each subject were used for training while 6 classes out of the 23 remaining images for testing. The 6 classes were chosen from the two sessions and contain expressions, illumination, and occlusions.

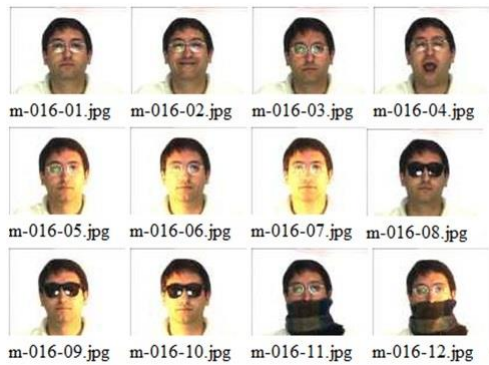


Fig. 3. Face Images for one subject from the AR face database: Examples of 12 expression recorded in the first session.

A. AR Face Database

A case study is conducted using AR Face database [15]. This database includes more than 4000 color images for 126 subject (70 men and 56 women). All images are frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. Every subject has 26 face images recorded in two separate sessions. Fig. 4 shows 12 expressions out of 13 recorded in the first session for one individual from the AR face database. The data set was divided into training and testing groups, ICA was trained from the first session on first three images per subject, which represent neutral expression, smiling, and anger. While the algorithm is tested with 6 images per subject including "4: scream", "7: all side lights on", and "9: wearing sun glasses and left light on" from the first section, and "17: scream", "20: all side lights on", and "22: wearing sun glasses and left light on" from the second session.

B. Face recognition performance

The different Architectures of ICA (ICA1 and ICA2), and the combined classifier ICA1+2 were evaluated using AR

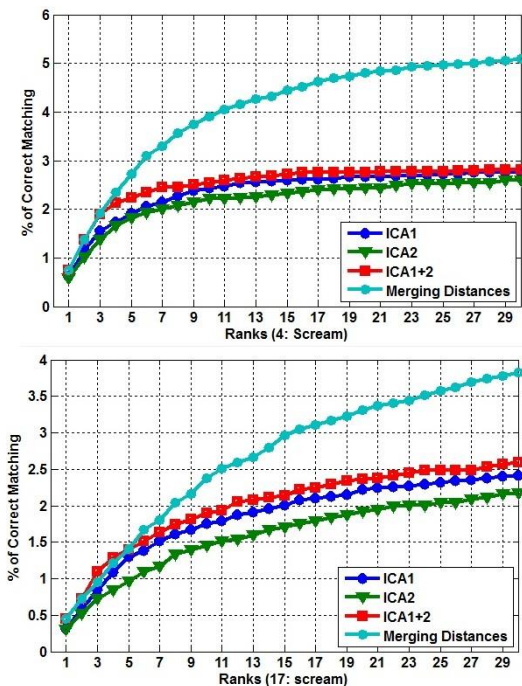


Fig. 4. Face recognition performance of (Merging Distances) compared to the individual classifiers ICA1 and ICA2 and the combined classifier ICA1+2 using images with scream expression

database. The recognition performance of the two architectures was evaluated by the nearest neighbor procedure using cosines as the similarity measure. The combined classifiers were deployed in which the similarity between a test image and a gallery image depends on both  $c_1$  of ICA1 and  $c_2$  of ICA2, where  $c_1$  and  $c_1$  correspond to the similarity measure. To analyze the performance of the two architectures of ICA algorithm we used Cumulative Mach Characteristic CMC.

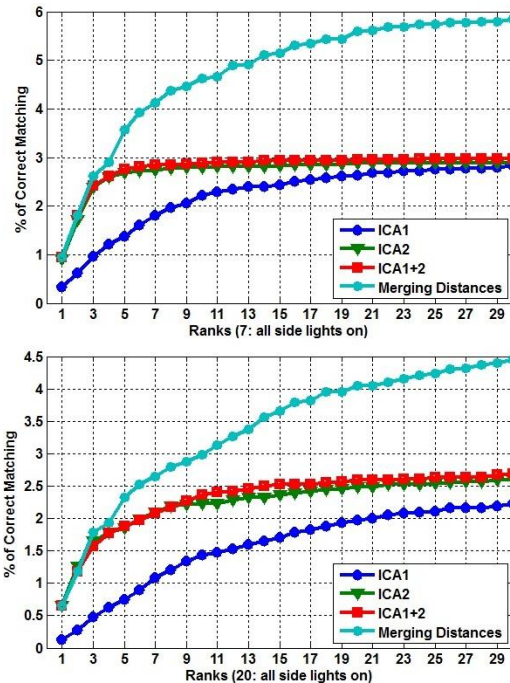


Fig. 5. Face recognition performance of (Merging Distances) compared to the individual classifiers ICA1 and ICA2 and the combined classifier ICA1+2 using images with illumination

1) Face images with scream expression and illumination: Fig. 4 and Fig. 5 show the Cumulative Mach Performance (CMC) of the performance of the ICA1, ICA2, ICA1+2, and the proposed classifier (Merging Distances) using face images with scream expression and illumination (e.g. all side lights on) taken in the same day and different day. As shown in the two figures, the proposed classifier consistently outperformed ICA1, ICA2, and ICA1+2 especially after rank 3.

2) Face images with occlusion: In case of face images with sun glasses, the proposed classifier has been tested using 2 testing sets including "wearing sun glasses and left light on" taken in the same day and different day. Fig. 6 shows the CMC using face images with sun glasses and left light on. The performance of the proposed classifier outperforms ICA1, ICA2, and ICA1+2 in later ranks. While in earlier ranks the performance of ICA1+2 is comparable with the proposed classifier. For the all testing faces taken in two different days,

it was noted that ICA1+2 performed better than both ICA1 and ICA2 in all ranks. However, it was comparable with "Merging Distances" in earlier ranks. By higher ranking, however, "Merging Distances" outperformed ICA1, ICA2 and ICA1+2. The overall results show that the proposed classifier has a consistence performance through different classes of the images in the database. While, the performance of ICA1 and ICA2 can be differ by changing the testing sets (See Fig. 4 and Fig. 5)

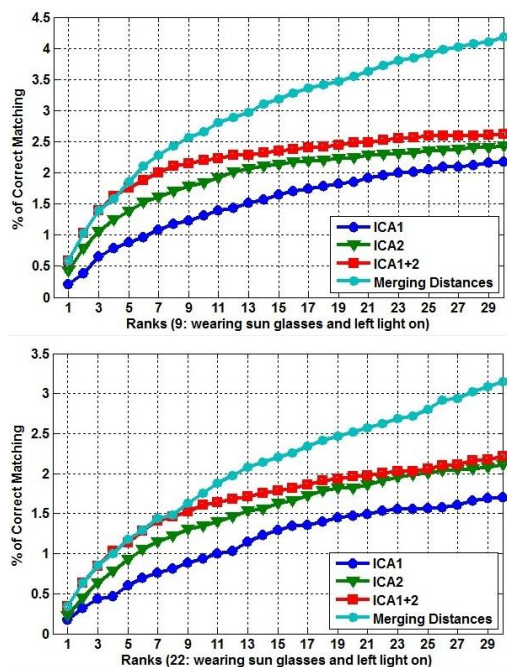


Fig. 6. Face recognition performance of the proposed classifier (Merging Distances) compared to the individual classifiers ICA1 and ICA2 and the combined classifier ICA1+2 using images with sun glasses and illumination

## V. CONCLUSION

We have proposed a new classifier that merges the two ICA representations ICA1 and ICA2 for face recognition. The proposed classifier (Merging Distances) was employed in which the similarity measure vectors for both ICA representations were integrated by joining the corresponding values of the two vectors to reconstruct a new similarity measure vector. The two architectures of ICA were performed on face images in the AR Face Database. The new classifier that combined the two ICA architectures were tested using faces with occlusion, illumination and different expressions. Cumulative Match Characteristics was taken as a measure for evaluating the performance of the new classifiers. In the early ranking, the proposed classifier was comparable with ICA1,

ICA2 and the combined classifier ICA1+2 developed by [13], while it outperformed them in higher ranking (After rank 3). The new combined classifier achieves reasonable results in recognizing faces in all test cases. Future work will consider recognizing the face by using any non-occluded facial area.

## REFERENCES

- [1] L. Xiaoguang, K. J. Anil, and C. Dirk, "Matching 2.5D Face Scans to 3D Models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 31-43, 2006.
- [2] X. Luan, B. Fang, L. Liu, W. Yang, and J. Qian, "Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion," *Pattern Recognition*, vol. 47, pp. 495-508, 2// 2014.
- [3] I. Venkat, A. T. Khader, K. G. Subramanian, and P. De Wilde, "Recognizing occluded faces by exploiting psychophysically inspired similarity maps," *Pattern Recognition Letters*, vol. 34, pp. 903-911, 6/1/ 2013.
- [4] M. I. Razzak, M. K. Khan, and K. Alghathbar, "Bio-inspired Hybrid Face Recognition System for Small Sample Size and Large Dataset," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, 2010, pp. 384-388.
- [5] J. Lu and E. Zhang, "Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion," *Pattern Recognition Letters*, vol. 28, pp. 2401-2411, 12/1/ 2007.
- [6] M. Girolami, *Advances in independent component analysis*. London [u.a.]: Springer, 2000.
- [7] T.-W. Lee, *Independent component analysis : theory and applications*. Boston: Kluwer Academic Publishers, 1998.
- [8] 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.
- [9] A. J. Bell and T. J. Sejnowski, "The "independent components" of natural scenes are edge filters," *Vision research*, vol. 37, pp. 3327-3338, 1997.
- [10] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems*, pp. 145-151, 1996.
- [11] M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, et al. (1997). *Analysis of fMRI Data by Blind Separation into Independent Spatial Components*.
- [12] M. S. Bartlett and T. J. Sejnowski, "Viewpoint invariant face recognition using independent component analysis and attractor networks," *Advances in neural information processing systems*, pp. 817-823, 1997.
- [13] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 13, pp. 1450-1464, 2002.
- [14] C. Liu and H. Wechsler, "Comparative assessment of independent component analysis (ICA) for face recognition," in *International conference on audio and video based biometric person authentication*, 1999.
- [15] A. M. Martinez, "The AR face database," *CVC Technical Report*, vol. 24, 1998.