KES'2002 proceedings - Knowledge-Based Intelligent Information and Engineering Systems, Crema (Italy),
16-18 September 2002, IOS Press, Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies
E. Damiani, R.J. Howlett, L.C. Jain, N. Ichalkaranje eds., ISBN 1-58603280-1, pp. 1207-1211

Independent Component Analysis for face authentication

C. Havran, L. Hupet, J. Czyz, J. Lee, L. Vandendorpe, M. Verleysen
Université catholique de Louvain, Electricity Dept.
3 place du Levant, B-1348 Louvain-la-Neuve, Belgium
{czyz, vandendorpe}@tele.ucl.ac.be; {lee, verleysen}@dice.ucl.ac.be

Abstract. In this paper, Independent Component Analysis (ICA) is presented as an alternative feature extraction algorithm to Principal Component Analysis (PCA) widely used in automatic face recognition/authentication tasks. We show that the promising ICA algorithm extracts from faces features that are relevant and efficient for authentication. This leads to improved success rates and a reduced client model size over a PCA based feature extraction.

1. Introduction

Face authentication has gained considerable attention these last years, through the increasing need for access verification systems using several modalities (voice, face image, fingerprints, pin codes, etc.). Such systems are used for the verification of a user's identity on the Net, when using a bank automaton, when entering a secured building, etc. Face authentication is different from face recognition (or classification): in authentication tasks, the system knows *a priori* the identity of the user (for example through its pin code), and has to *verify* this identity; in other words, the system has to decide whether the a priori user is an impostor or not. In face recognition, the a priori identity is not known: the system has to decide which of the images stored in a database resembles the most to the image to recognize; the decision is no more binary. Although ICA (Independent Component Analysis) could be beneficial both for face authentication and recognition, we will concentrate on the first in this paper.

In face authentication, as in most image processing problems, features are extracted from the images before processing. Working with rough images is not efficient: in face authentication, several images of a single person may be dramatically different, because of changes in viewpoint, in colour and illumination, or simply because the person's face looks different from day to day. Therefore extracting relevant features, or *discriminant* ones, is a must. Nevertheless, one hardly knows in advance which possible features will be discriminant or not. For this reason, one of the methods often used to extract features in face authentication is PCA (Principal Component Analysis) [1]. Another family of methods are the local feature-based methods such as [2], or those based on LDA (Linear Discriminant Analysis) as in [3]. In this paper, we show how the promising ICA (Independent Component Analysis) technique extracts features that are more closely related to our intuition of discriminant information, and that improve the success rate compared to an equivalent system using PCA. PCA, LDA and ICA belong to the family of subspace methods [4]. The remaining of this paper is organised as follows. Section 2 presents the problem of face authentication. Section 3 shows how to extract features from rough images, and presents the procedure based on ICA. Section 4 shows the experimental results.

KES'2002 proceedings - Knowledge-Based Intelligent Information and Engineering Systems, Crema (Italy),
16-18 September 2002, IOS Press, Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies
E. Damiani, R.J. Howlett, L.C. Jain, N. Ichalkaranje eds., ISBN 1-58603280-1, pp. 1207-1211

2. Face authentication

Face authentication systems typically compare a feature vector X extracted from the face image to verify with a client template, consisting in similar feature vectors Y_i extracted from images of the claimed person stored in a database ($1 \leq i \leq n$, where n is the number of images of this person in the learning set). The matching may be made in different ways, one being to take the Euclidean distance between vectors (this method will be taken as an example here). If the distance between X and Y_i is lower than a threshold, the face from which X is extracted will be deemed to correspond with the face from which Y_i is extracted. Choosing the best threshold is an important part of the problem: a too small threshold will lead to a high *False Rejection Rate* (FRR), while a too high one will lead to a high *False Acceptance Rate* (FAR); FRR and FAR are defined as the proportion of feature vectors extracted from images in a validation set being wrongly classified, respectively wrongly authentified and wrongly rejected. The validation and test sets must be independent (though with faces of the same people) from the learning set, in order to get objective results. One way of setting the threshold is to choose the one leading to equal FRR and FAR. If the a priori probabilities of having false acceptances (impostors) and false rejections are equal, this corresponds to minimizing the number of wrong decisions, as a result of Bayes' law. Other criteria could be considered, such as using individual thresholds for each person in the database; again, as our goal is to measure the advantages of ICA with respect to PCA feature extraction, we will not investigate other ways of fixing thresholds, and use the global threshold leading to FRR = FAR in the remaining of this paper.

3. Feature extraction

Taking decisions on rough images has been shown [3] to be dramatically sensitive to illumination conditions, viewpoints, expression and day-to-day differences in a face of the same person, to the point that two very similar (to the human eye) images could be extremely different if compared pixel by pixel. It is therefore necessary to extract relevant, discriminant features from the images and to compare the features instead of the rough images. Of course, the more discriminant are the features, the easier will be the subsequent authentication.

3.1. Principal Component Analysis (PCA)

A traditional way of extracting features is to use PCA. From a set of N d-dimensional images in the learning set (d being the number of pixels in each image), the PCA method extracts so-called *principal components* as the eigenvectors of the covariance matrix of the data. Geometrically, the principal components are the directions in the data space maximizing the variance of the projection of the original vectors on these axes. The principal components are ranked by the associated eigenvalues of the same matrix, the largest eigenvalue corresponding to the axis maximizing the variance of the projections, and so on. As principal components have the same dimension d as the original images, and can be represented as such, they are often referred to as *eigenfaces*.

The set of eigenfaces is built on the whole learning set. What is important to notice is the fact that, once the eigenfaces are known, each face in the learning and in the validation set may be coded, or *reconstructed*, as a linear combination of eigenfaces; the sets of linear coefficients form the feature vectors. If there are at least d uncorrelated faces in the learning set, $N \geq d$, then PCA will extract d eigenfaces and the reconstruction is error-free. However, if only P eigenfaces are kept ($P < d$), the coding minimizes the mean square error between the original faces and the coded ones.

KES'2002 proceedings - Knowledge-Based Intelligent Information and Engineering Systems, Crema (Italy), 16-18 September 2002, IOS Press, Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies
E. Damiani, R.J. Howlett, L.C. Jain, N. Ichalkaranje eds., ISBN 1-58603280-1, pp. 1207-1211

While projecting the original faces on the principal components is mathematically justified by the fact that it maximizes the variance after projection, there is no rationale under the fact that the directions in which the variance is maximum are the most discriminant directions for taking the right decision. There is even some intuitive evidence that it is not the case: the first eigenvectors will correspond to the general shapes of faces, which are common to all of them (therefore not discriminant). The idea is then to replace PCA by a method able to extract more perceptive features from faces. By perceptive, we mean features that could be used by a human being to discriminate between or to describe faces. However, as it will be justified below, ICA will not replace PCA, but will be used as a supplementary step after PCA. This will allow keeping a larger number of eigenvectors after PCA (keeping thus more information in the feature vectors), this number being further reduced by ICA. PCA will act as a first whitening filter, while the discriminant reduction will be achieved by ICA.

3.2. Independent Component Analysis (ICA)

ICA is a data analysis tool derived from the "source separation" signal processing techniques. The aim of source separation is to recover original signals S_i, from known observations X_j, where each observation is an (unknown) mixture of the original signals. Under the assumption that the original signals S_i are statistically independent, and under mild conditions on the mixture, it is possible to recover the original signals from the observations. The algorithmic techniques making this task possible are often called ICA, as they factorise the observations as a combination of original sources. If the mixing is linear, ICA estimates the inverse of the mixing matrix. The number of observations N ($1 \le j \le N$) must be at least equal to the number of original signals M ($1 \le i \le M$); often it is assumed that $N = M$. It is not necessary to have signals X_j to consider using ICA: X_j may also be multi-dimensional data (vectors). Assuming that each X_j is an unknown, different combination of original "source vectors" S_i, ICA will expand each signal X_j into a weighted sum of source vectors S_i (ICA estimates both the source vectors S_i and the coefficients of the weighted sum). This view is not far from the PCA expansion: the eigenvectors of PCA are replaced by the independent source vectors in ICA. For a review of ICA techniques and properties, see for example [5].

In our case, we assume that the faces in the learning set, viewed as high-dimensional vectors, are linear combination of unknown independent source vectors. This may not be strictly true, depending on the respective number of images in the database and size of an image in pixels, but in any case ICA will find estimates of independent source vectors that are optimal to reconstruct the original images (observations) in the least-square sense. The idea is then to substitute PCA with ICA, and to use the coefficients of the ICA expansion (instead of those from PCA) as feature vectors for the faces. It is expected that, ICA source vectors being independent (instead of PCA eigenvectors being uncorrelated only), they will be closer to natural features of images, and thus more able to represent differences between faces.

3.3. Use of ICA in face authentication

However ICA does not have advantages only. ICA algorithms are iterative, and sometimes converge difficultly. Moreover, ICA methods show difficulties to handle large number of signals (or high-dimensional vectors in our case). The FastICA package [6] has been used for its good performances in our simulations. To overcome the difficulties related to the high dimensionality of vectors, their dimensionality has first been reduced by PCA. This might sound odd given the above arguments; however, the experiments in section 4 make clear that the dimension P of the vectors after PCA reduction will be chosen much larger than in "PCA only" experiments. Furthermore, PCA whitening helps ICA to converge.

KES'2002 proceedings - Knowledge-Based Intelligent Information and Engineering Systems, Crema (Italy),
16-18 September 2002, IOS Press, Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies
E. Damiani, R.J. Howlett, L.C. Jain, N. Ichalkaranje eds., ISBN 1-58603280-1, pp. 1207-1211

A further difficulty of ICA compared to PCA is the ordering of source vectors. In PCA, the corresponding eigenvalues are used to rank the eigenvectors (according to their contribution to the total variance of data), the first ranked (corresponding to the largest eigenvalues) being kept in case of dimensionality reduction. ICA does not offer an ordering of the source vectors. For this reason, as suggested by [7], we rank them according to a class separability criterion estimated over the learning set $r_i = \sigma_{bi} / \sigma_{wi}$ where σ_{bi} and σ_{wi} are the within and between class variance of the source component S_i (the classes correspond to the different identities in the database). A high value of r_i corresponds to a discriminant source vector S_i, thus only the Q source vectors associated with the highest values of r_i are conserved. Note that ordering PCA eigenvectors according to a class separability criterion (instead of eigenvalues) would lead to a method similar to LDA (see introduction).

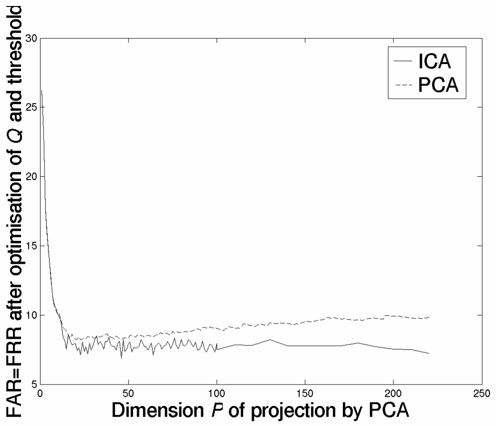
The following procedure is thus suggested to expand the faces into feature vectors and to authentify a face accordingly: (i) the d-dimensional images in the learning set are reduced by PCA to P-dimensional vectors; the last are formed by the coefficients of the PCA expansion corresponding to the largest eigenvalues; (ii) the P-dimensional vectors are further expanded into Q-dimensional feature ones by ICA; the choice of ICA source vectors is made according to criterion (1); (iii) projection matrices from steps 1 and 2 (sizes Pxd and QxP respectively) are used to transform the images in the validation set into Q-dimensional feature vectors; (iv) these feature vectors are authentified according to the procedure described in section 2. Three parameters must be determined in the method: P, Q, and the threshold used for the authentication procedure. For each value of P and Q, the threshold is fixed to have FAR=FRR; P and Q are chosen to minimize this error rate. Finally, (v) the performances of the method (including the threshold value) are measured on an independent test set (on this set, FAR will not be necessarily equal to FRR).

4. Experimental results

Our experiments were performed on frontal face images from the XM2VTS database [8]. XM2VTS is a publicly available multimodal database recorded specifically for assessing the performances of biometric approaches to identity verification. It contains 8 face images of 295 persons. The subjects were recorded in four separate sessions distributed over a period of 5 months. The standard experimental protocol associated with the database divides the database into 200 clients and 95 impostors. The protocol specifies a partitioning of the database into disjoint sets for training, validation and testing.

Figure 1 shows the FAR=FRR obtained on the validation set after optimization of the threshold and Q, with respect to P. The two curves show respectively (plain line) the result of the algorithm described in section 3, and (dashed line) the result of a PCA only applied to the original data. It clearly shows the improvement obtained by the use of ICA for a wide range of the dimension P. It also shows that, contrarily to the use of PCA only, applying the procedure from section 3 gives comparable results in a wide range of the parameter P, making its choice less critical. Figure 2 shows a detail from Figure 1, with two supplementary curves corresponding to error bars around the ICA curve (the experiments have been conducted a large number of times to assess the reliability of the ICA step). Even the top curve shows improved results compared to PCA-only feature extraction. Both Figures 1 and 2 have been obtained by using the Euclidean distance between feature vectors for their matching (see section 2). It has been found experimentally that using the angle between these vectors instead of the Euclidean distance further improves the results. Table 1 summarizes some results obtained, both with the use of the Euclidean distance and of the angle. The values shown for dimensions P and Q are those found after optimization.

KES'2002 proceedings - Knowledge-Based Intelligent Information and Engineering Systems, Crema (Italy),
16-18 September 2002, IOS Press, Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies
E. Damiani, R.J. Howlett, L.C. Jain, N. Ichalkaranje eds., ISBN 1-58603280-1, pp. 1207-1211



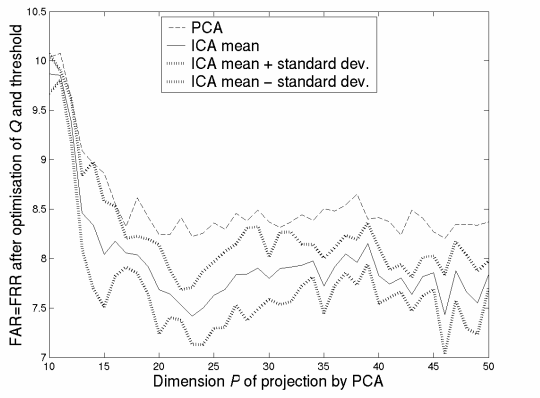


Figure 1: performance of ICA and PCA (see text for details). Figure 2: detail of Fig.1, + standard deviations.

Table 1: results of the method, and of PCA-only feature extraction for comparison.

Type of distance	Dimension *P* after PCA	Dimension *Q* after ICA	FAR=FRR on learning set	(FAR+FRR)/2 on validation set
Euclidean	46	no ICA	8.2	8.06
Euclidean	23	18	7	7.34
Angle	100	no ICA	6.44	5.44
Angle	81	71	5.62	5.21

5. Conclusion

This papers describes a procedure for using ICA as feature extractor in the context of face authentication. Results on experiments performed on a standard database show increased performances with respect to the use of PCA only as feature extractor. Moreover, the results also show a lower sensitivity to the choice of the projection dimension after PCA. Further work may consist in replacing the simple decision system authentifying the faces through simple distance comparisons between feature vectors, by a multi-dimensional classifier (artificial neural network) on the components of these vectors.

6. Acknowledgements

The authors would like to thank Christian Jutten (I.N.P. Grenoble) for fruitful discussions. Michel Verleysen is a senior research associate of the Belgian FNRS. Part of this work was realized with the support of the 'Ministère de la Région wallonne', under the 'Programme de Formation et d'Impulsion à la Recherche Scientifique et Technologique' and by the European Union IST projects BANCA and OPTIVIP.

References

[1] M. Turk and A. Pentland, Eigenfaces for recognition. Journal of cognitive neuroscience, **3**(1), 1991.
[2] L.Wiskott, J.-M. Fellous, N.Kruger and C.von der Malsburg "Face recognition by elastic bunch graph matching" Trans. on Pattern Analysis and Machine Intelligence, Vol.19, pp.775-779 , 1997
[3] P. Belhumeur, J.P. Hespanha, D.J. Kriegman , Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. on Pattern Analysis and Machine Intelligence, **19**(7), 1997, pp. 711-720.
[4] E. Oja, Subspace methods of Pattern Recognition, Research Studies Press Ltd., 1983.
[5] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New-York, 2001.
[6] The FastICA package, http://www.cis.hut.fi/projects/ica/fastica/.
[7] M. Bartlett, H. Lades, T. Sejnowski, Independent component representations for face recognition". In Proceedings of the SPIE Symposium of Eletronic Imaging, **3299**, 1998, pp. 528-539.
[8] K.Messer, J.Matas, J. Kittler, J. Luettin, G. Maître, XM2VTSBD: The Extended M2VTS database. Int'l Conf. on Audio- & Video-based Biometric Authentication (AVBPA 99), Washington D.C.,1999.