

ID Image Characterization by Entropic Biometric Decomposition

Andreea Smoaca^{*,†}, Daniela Coltuc[†] and Thierry Fournel^{*}

** Université de Lyon / CNRS, UMR 5516, Laboratoire Hubert Curien
Université de Saint-Etienne, Jean Monnet, France*

*† Politehnica University of Bucharest
Faculty of Electronics, Telecommunications and Information Technology, Romania*

Abstract. The paper proposes a statistical-based biometric decomposition for ID image recognition robust to a series of non malicious attacks generated by print/scan operations. Our goal is to label the single face expression by a signature, which is almost invariant to low filtering, noise addition and geometric attacks. The method is based on Independent Component Analysis (ICA) in a configuration which will allow a decomposition into some face characteristics. In this configuration known in literature as Architecture I, the most important coefficients issued from ICA are selected by looking for the independent components with maximum local entropy. A biometric label of fixed length is associated to any ID image to be enrolled, after projection on the learned basis, uniform quantization of the obtained coefficients and binary encoding. Two parameters were tuned: the number of quantization levels and the number of face characteristics. The latter one was modified, either by discarding coefficients after Principal Component Analysis in the beginning of FastICA algorithm, or by selecting the most prominent biometric features by applying an entropic criterion. The suggested method inherits the robustness of a global approach.

Keywords: ID image recognition, Independent Component Analysis

PACS: 07.05.Pj ; 2.50.Fz

INTRODUCTION

With the wide deployment of digital technologies, a growing number of applications attempt to integrate face-recognition. Indeed in many of them, person's identity is helpful as in human behavior interpretation or even required as in biometry. The tolerated recognition performance depend on the targeted application domain. A particular case of face recognition is identity (ID) image recognition. This is a restrained case that takes into account only faces with neutral frontal expression. Here, the difficulties of recognition are caused by aging or by print/scan processes. The print/scan alter the images by introducing noise and geometrical distortions. A reliable ID image recognition method must reach a recognition rate of 100%.

Among various tools studied for face recognition, Principal Component Analysis (PCA) and Independent Component Analysis (ICA) allow a global approach including a preliminary learning step. PCA derives a basis of vectors, which are the eigenvectors of the covariance matrix of the serialized images in a training set. A new image, of the same class, may be represented by the series of coefficients obtained by projecting the image on the basis vectors.

Being based exclusively on the covariance matrix, PCA does not take into account the class statistics higher than two. A method that finds basis vectors sensitive to higher

– order statistics is Independent Components Analysis (ICA). The goal of ICA is to decompose a random signal, observed through its particular realizations collected in a matrix x , into a linear combination of unknown independent components (ICs), placed on the rows of s :

$$x = As \tag{1}$$

In order to maximize the independence among the rows of s , several algorithms that maximize/minimize a specific cost function were implemented. They iteratively optimize these functions whose global optima occur when independence is achieved. For instance, InfoMax algorithm [4] maximizes the entropy of the estimated independent vectors and FastICA [5] maximizes the non-Gaussianity. In [3], Draper and all show that the performance of face recognition is affected by the choice of the ICA algorithm.

Another factor that may affect the performance is the manner in which ICA is applied. As shown by Bartlett and all [2], two types of architecture can be exploited. They can lead either to global feature vectors (Architecture II) or to localized feature vectors (Architecture I) as in Fig. 1. In both cases, the faces in the training set are serialized and organized into the matrix x ; they are placed on rows in Architecture I or on columns in Architecture II. The first representation leads to statistically independent basis vectors (each image is represented as a linear combination of face characteristics) and the latter to statistically independent coefficients (each image is represented as a linear combination of face images).

The use of ICA algorithm has shown its relevance also in the case of ID image recognition [6], where Architecture II was used in order to obtain an image compressed representation. Our approach for ID image characterization being based on biometric features, implies the use of Architecture I. The use of FastICA algorithm, which maximizes the negentropy, justifies the natural choice of an entropic criterion to select the biometric characteristics.

The rest of the paper is organized as follows: the proposed method with different strategies for feature subspace selection are detailed in the next section. Another section is dedicated to the experimental results obtained by tuning two parameters: the number of quantization levels for the coefficients and the selection of basis vectors. In the last section, we discuss the performance of the method.

ICA-ARCHITECTURE I DECOMPOSITION

According to this architecture, each face image is represented as a sum of face characteristics. As it can be observed in Fig. 1, behind human biometric features, more or less salient face images appear. Our purpose is to determine an optimal ID image representation via ICA in Architecture I, more precisely, to find the most relevant basis vectors in the sense of information content.

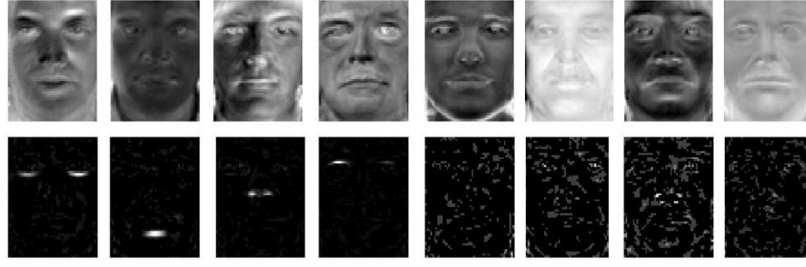


FIGURE 1. Eight feature vectors for the two architectures. The top row contains the eight non-localized ICA feature vectors obtained with ICA architecture II. The second row shows the eight (localized) feature vectors obtained with ICA architecture I.

The proposed methodology

Image registration is achieved by using a set of reference points during a preliminary preprocessing step. In this step every image is centered by using the coordinates of eyes and mouth, cropped and scaled. The images of a faces learning set are then serialized and stored as the different rows of a single matrix x . By applying an ICA algorithm, a vector basis is derived from x . The basis vectors \mathbf{s}_j consist in independent face features as shown in Fig. 1.

Any new face image I , registered and serialized in the same way as the training images, can be represented by the learned basis (Fig. 2) :

$$I = \sum_{j=1}^L b_j \mathbf{s}_j \quad (2)$$

The coefficients b_j are obtained by projection of I onto the basis vectors. After uniform quantization and binarization of coefficients b_j , a signature of I is formed by concatenating the bits.

FIGURE 2. New image represented as a linear combination of the independent basis images.

The proposed methodology can be summarized as the following steps:

- a** Image preprocessing and registration,
- b** Image serialization,
- c** Image basis vectors computation (from a learning set),
- d** Subspace selection,
- e** Subspace projection coefficients,
- f** Coefficient quantization and binarization.

Some strategies for subspace selection

The noise caused by print/scan processes alters the values of coefficients b_j so the image signature, decreasing thus the recognition performance with a certain risk of false authentication. In order to reduce such a risk, a subspace selection is recommended.

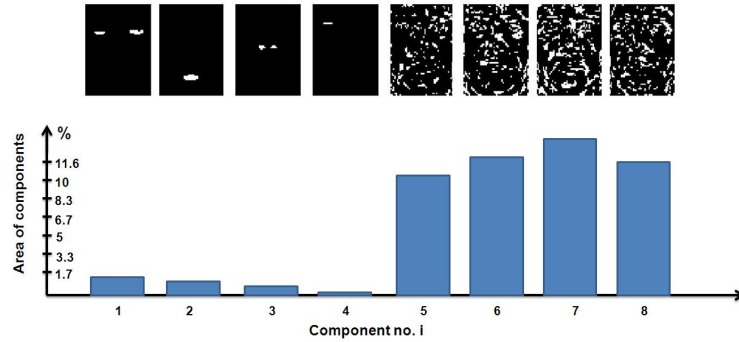


FIGURE 3. Selection of components using area-criterion. The top row contains the eight binarized feature vectors from Fig. 1 obtained with ICA architecture I. The second row represents the area of each of the above feature vector. Only the features with low area are of interest.

By ICs selection, only the most significant components are used for face representation (Fig. 3). The removal of insignificant components may be done by the following mechanisms:

- selection by PCA in the beginning of ICA.
- selection by ICA using:
 - a global entropic criterion.
 - a local entropic criterion.
- combination of the above two.

The selection by PCA consists in retaining only the highest variance coefficients while the selection of ICs is done by using an entropic criterion. In the case of global, respectively local entropy selection, the most significant features with minimal, respectively maximal entropy are retained.

For entropic criterion we do not need an accurate source model for face characteristics. Our goal is rather to have an efficient and fast selection. It is the reason why we can consider face characteristics binarization and zero-order entropy estimate. It is known that for a binary memory-less source of information, the entropy is maximum for $p = 0.5$ and symmetrical about this point (p is the probability of one of the symbols).

Entropy estimate can be advantageously replaced by area measurements, which provide permissive thresholding Fig. 3.

For local entropy criterion, the ICs images were split in 3 strips : the upper third containing the eyes and eyebrows, the middle third containing the nose and the lower one containing the mouth. The retained ICs have maximum entropy (bigger area) on a single strip and (quasi) null entropy on the other two.



FIGURE 4. Samples with frontal view and neutral expression extracted from FERET database.

EXPERIMENTAL RESULTS

Our approach was tested on a subset of FERET database of facial images([8], [9]). Only the images of frontal view and neutral expression subjects were retained, in order to stay close to ID image recognition. The training set used for learning the basis vectors has consisted of 300 gray level images of 384x256 pixels (Fig. 4). For tests, another set of 210 different subjects face images with the same characteristics were considered. All the images were normalized in order to have zero mean and variance equal to one, registered, cropped and resized to 60x50 pixels. The registration was done by using the eyes and mouth coordinates provided by FERET databases. By averaging the training images coordinates, a reference set of coordinates has been obtained. All the faces, for training or tests, were centered to the reference set. An example is shown in Fig. 5.

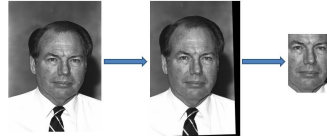


FIGURE 5. Image preprocessing step.

By projecting the test images on the learned ICs and by uniformly quantizing the obtained coefficients on 8 levels and binary encoding, a binary signature is obtained for each face.

The signatures were tested against a series of stirmark attacks [10], [11] that we have considered appropriate for simulating print/scan noise (affine transforms, additive Gaussian white noise and median filter). The Hamming distance was used to compare the binary signatures before and after attack. We were also interested in evaluating the gap between the distributions of Hamming distances obtained for similar faces (originals and attacked) and dissimilar ones. The gap has been estimated by the following ratio:

$$r = \frac{|\mu_s - \mu_d|}{\sigma_s + \sigma_d} \quad (3)$$

where μ_s , σ_s are the mean and standard deviation of the similar faces, respectively μ_d , σ_d the mean and standard deviation of the dissimilar faces. The choice of such measure is justified by our intention to convert these signatures into hash values that need to be stable against print/scan attacks.

TABLE 1. Comparison between strong features and weak features.

	Attack	Similar faces		Dissimilar faces		r
		mean	σ	mean	σ	
Case I	AFFINE _1	5.6	2.3	37.7	6.5	3.7
	AFFINE _2	18	4.3	38.5	6.3	1.9
	AFFINE _3	6.6	2.4	37.8	6.4	3.5
	AFFINE _4	28.7	5.7	42	5.6	1.2
	AFFINE _5	8.2	2.7	37.8	6.4	3.3
	AFFINE _6	9.9	3.2	38	6.3	3
	WN $\sigma = 0.01$	17.5	5.5	34.53	7.1	1.3
	Median 5×5	13.6	4.4	38.1	6.4	2.3
Case II	AFFINE _1	4	1.9	31.7	6.7	3.2
	AFFINE _2	18.7	4	33.4	5.6	1.5
	AFFINE _3	5	2	31.9	6.5	3.1
	AFFINE _4	21.5	4.5	34.9	5.2	1.3
	AFFINE _5	7.3	2.6	31.8	6.5	2.7
	AFFINE _6	9.1	2.9	32.1	6.4	2.5
	WN $\sigma = 0.01$	10.2	3.7	24.3	7	1.3
	Median 5×5	9.2	3.1	32.7	6.8	2.3

The learning set was analyzed by FastICA software [7]. For 300 images, in Architecture I, FastICA can extract at maximum 300 ICs. Since by quantizing and binarizing, the signature would be 900 bits long, we decided to shorten it by discarding less significant ICs. Since FastICA includes PCA, a first attempt to reduce the ICs number was by means of PCA. Thus, by discarding coefficients carrying 6% of signal energy, the number of ICs has been reduced at 60. Further, in order to select ICs with salient features, we have used the criterion of global entropy. By evaluating the white area on the binarized ICs, obtained by automatic thresholding (Otsu's method), the basis was split into a subset of 33 ICs with low entropy (Case I) and 27 with high entropy (Case II). A threshold value equal to 50 was used for this selection. Table 1 shows the usefulness of this criterion. Because of the few number of ICs (≈ 30 for each situation) the recognition rate is not always 100 %, but even so the performance is higher in Case I. For AFFINE _1 attack the recognition rate is 100 % in both cases, but in Case I the ratio r is higher (3.7 vs. 3.2).

We have tested in the same way other reducing strategies : only by PCA (PCA180, PCA120 and PCA66), by PCA and local entropy selection (120LE) and by simultaneously using local and global entropy selection criteria (66LE-GE). The results are presented in table 2. Both recognition rate and ratio increase when the number of selected components is higher. For example, for geometric attack (AFFINE _1) when using the first 180 PC, r is 5 compared with 4.7 obtained for the first 120 PC. If the number of quantization levels is tuned, for a higher level the recognition rate and ratio are higher (95.2 % vs. 94.7 % for AFFINE _4 with 8, respectively 4 levels). When the number of components and the number of levels are constant (120, respectively 8), the recognition

TABLE 2. Results for different attacks and selection criteria : PC, LE (Local Entropy), GE (Global Entropy)

	Attack	Rate (%)	r	Attack	Rate (%)	r
180 PCA $L = 8$	AFFINE_1	100	5	AFFINE_2	100	2.4
	AFFINE_3	100	4.7	AFFINE_4	98.6	2.1
	AFFINE_5	100	4.2	AFFINE_6	100	3.8
	AFFINE_7	100	4.1	AFFINE_8	100	4
	WN $\sigma = 0.01$	99.1	2.5	WN $\sigma = 0.02$	96	1.9
	Median 3×3	100	4.4	Median 5×5	100	3.1
120 PCA $L = 8$	AFFINE_1	100	4.7	AFFINE_2	96.7	2.1
	AFFINE_3	100	4.3	AFFINE_4	95.2	1.8
	AFFINE_5	100	3.8	AFFINE_6	100	3.5
	AFFINE_7	100	3.8	AFFINE_8	100	3.7
	WN $\sigma = 0.01$	97.6	2.3	WN $\sigma = 0.02$	91	2.3
	Median 3×3	100	4.2	Median 5×5	100	3
120 PCA $L = 4$	AFFINE_1	100	3.7	AFFINE_2	95.2	2.1
	AFFINE_3	100	3.5	AFFINE_4	94.7	1.7
	AFFINE_5	100	3.2	AFFINE_6	100	3
	AFFINE_7	100	3.1	AFFINE_8	100	3.2
	WN $\sigma = 0.01$	96.3	2	WN $\sigma = 0.02$	88.9	1.7
	Median 3×3	100	2.5	Median 5×5	100	2.7
120 LE $L = 8$	AFFINE_1	100	4.7	AFFINE_2	99	2.4
	AFFINE_3	100	4.3	AFFINE_4	99	2.2
	AFFINE_5	100	4	AFFINE_6	100	3.7
	AFFINE_7	100	3.8	AFFINE_8	100	3.9
	WN $\sigma = 0.01$	99	2.5	WN $\sigma = 0.02$	94.7	1.9
	Median 3×3	100	4.3	Median 5×5	100	3
66 PCA $L = 8$	AFFINE_1	100	4.8	AFFINE_2	84.3	1.7
	AFFINE_3	100	4.3	AFFINE_4	77.7	1.5
	AFFINE_5	100	3.7	AFFINE_6	100	3.4
	AFFINE_7	100	3.5	AFFINE_8	100	3.6
	WN $\sigma = 0.01$	98.5	2.3	WN $\sigma = 0.02$	84.8	1.7
	Median 3×3	100	4.1	Median 5×5	100	2.8
66 LE – GE $L = 8$	AFFINE_1	100	4.4	AFFINE_2	91.9	1.9
	AFFINE_3	100	4.1	AFFINE_4	95.2	1.9
	AFFINE_5	100	3.5	AFFINE_6	100	3.3
	AFFINE_7	100	3.3	AFFINE_8	100	3.3
	WN $\sigma = 0.01$	98.6	2.2	WN $\sigma = 0.02$	86.2	1.6
	Median 3×3	100	3.9	Median 5×5	100	2.6

rate and ratio r are better when using an entropic criterion (r is 4.9 in this case). This is valid also for the geometric attacks with recognition rate smaller than 100 %. Even when performing a second step of selection, on the 120 LE components, by choosing the components with low area, the rate of recognition is better (for the cases where full recognition is not achieved) than when only using PC selection. The performance of the combined strategy proved to be closer to PCA strategy with a double number of components. For example, for AFFINE_4 attack, the recognition rate is 95.2 % for 120PCA and 66LE–GE and only 77.7 % for 66PCA.

DISCUSSION

Biometric features are selected after architecture I ICA to obtain global decomposition of ID images. The performed tests show that the recognition rate is high, 100 % in almost all the studied cases, whatever the chosen strategy. The performance i.e. recognition rate and gap between similar and dissimilar distributions, varied when using different numbers of ICs and quantization levels. The more ICs are used the more high performance is. The same tendency was observed when increasing the number of quantization levels. However the discrimination power given by the gap is higher for an entropic selection of a given number of face characteristics.

In fact, architecture I reveals two categories of face characteristics, depending on whether salient face image appears or not. A local entropy maximization which allows their automatic clustering corresponds to the most suitable strategy for subspace selection.

In order to further enforce the discrimination power, the method could be combined with a local approach detecting some biometric characteristic points as in photocomparison of skulls [12].

ACKNOWLEDGMENTS

The first author acknowledges for the financial support from the ESF, Contract no. POSDRU/6/1.5/S/19 and UEFISCU Romania, Grant 610/2008.

REFERENCES

1. A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, 2001, John Wiley & Sons, Inc.
2. M. S. Bartlett, J. R. Movellan and T. J. Sejnowsky, *Face Recognition by Independent Component Analysis*, IEEE Transaction on Neural Networks, Vol.13, 2002, pp. 1450–1464.
3. B. A Draper, K. Baek, M. S. Bartlett and J. R. Beveridge, *Recognizing faces with PCA and ICA*, Computer vision and image understanding 91: Special issue on Face Recognition, 2003, pp. 115-137.
4. J. F. Cardoso, *Infomax and maximum likelihood for source separation*, IEEE Letters on Signal Processing, Vol. 4, 1997, pp. 112–114.
5. A. Hyvärinen, *The fixed-point algorithm and maximum likelihood estimation for independent component analysis*, Neural Processing Letters, Vol. 10, 1999, pp. 1–5.
6. T. Fournel and D. Coltuc, *Robust visual hashing via ICA*, J. Phys.: Conf. Ser. 206 012035, 2010.
7. FastICA package for Matlab. Available at: <http://www.cis.hut.fi/projects/ica/fastical/>
8. P. J. Phillips, H. Wechsler, J. Huang and P. Rauss, *The FERET database and evaluation procedure for face recognition algorithms*, Image and Vision Computing J, Vol. 16, No. 5, 1998, pp. 295–306.
9. P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss, *The FERET Evaluation Methodology for Face Recognition Algorithms*, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 12, 2000, pp. 1090–1104.
10. F. A.P. Petitcolas, R. J. Anderson, M. G. Debois, *Attacks on copyright marking systems*, in David Aucsmith, Information Hiding, Second International Workshop, IH'98, Portland, Oregon, U.S.A., April 15–17, 1998, Proceedings, LNCS 1525, Springer-Verlag, ISBN 3-540-65386-4, pp. 219–239.
11. F. A.P. Petitcolas *Watermarking schemes evaluation*, I.E.E.E. Signal Processing, vol. 17, no.5, pp. 58–64, September 2000.
12. Y. Desbois, R. Perrot, C. Debois, *Incidence de l'occlusion dentaire lors d'une craniophotocomparaison : a propod'un cas. Paleobios*, Vol. 13, 2004.