# A Comparative Evaluation of Global Representation-Based Schemes for Face Verification

*S. Cruz-Llanas, J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez*

cruzll@atvs.diac.upm.es, {jfierrez,jortega,jgonzalez}@diac.upm.es
Biometrics Research Lab. – Universidad Politecnica de Madrid

**ABSTRACT**

This paper is focused on algorithmic issues for biometric face verification (i.e., given an image of the face and an identity claim, decide whether they correspond to each other or not). Several alternatives for geometric normalization of images, photometric normalization, dimensionality reduction and similarity measures are proposed and compared using the XM2VTS database and the associated *Lausanne* protocol [10], [11]. Experiments under this particular framework show that best verification results are obtained when holistic approaches for face recognition (such as eigenfaces or fisherfaces) are combined with techniques traditionally associated to local feature-based approaches, such as Gabor decompositions.

## 1. INTRODUCTION

Biometric signals and traits (fingerprints, speech, face images, etc.) contain identity information about the subject they belong to. Automatic extraction of these cues has given raise to a particular branch of pattern recognition (*biometrics*) where the goal is to infer identity of people from biometric data. The increasing interest on biometrics is related to the important number of applications where a correct assessment of identity is a crucial point.

One of the drawbacks of face recognition, compared to other biometric modalities as fingerprint, is that it is very sensitive to the problem of variability or mismatch between training and testing conditions. Some sources of this variability are changes in scale, location or rotation due to the point of view of the camera, changes in expression, illumination, ageing, make-up, etc. (see Fig. 1).

Fig. 1: Example of variability of face images

Representation schemes for face recognition can be divided into local and global, depending on whether the face is represented as a whole, or as a series of small regions [1]. Face recognition and verification systems based on both approaches have already been compared in published competitions [2],[3]. On the one hand, dimensionality reduction is a key step in the case of global strategies for face recognition [4], [5]. On the other hand, Gabor features (commonly used in local face appearance representation schemes) constitute also a successful option among face recognition practitioners [6]. Experiments reported here show that face verification performance can be improved considering jointly these two important ideas.

## 2. DESCRIPTION OF THE SYSTEM

All the results provided here have been obtained following the same basic processing scheme in order to obtain a final similarity measure for verification. It consists of the following sequence of steps: geometric normalization, photometric normalization, dimensionality reduction and final computation of the similarity measure.

### 2.1. Geometric normalization

The first task to be performed in a fully automatic face recognition practical system should be the detection and segmentation of the face [7]. As the emphasis in this contribution is put on recognition, automatic segmentation will be skipped over here, and manually located positions of 6 relevant facial features (center of the iris, nostrils and mouth corners, see Fig. 2) will be used.

Geometric normalization is performed in order to overcome two drawbacks. First non-face pixels (background) are not meaningful for identity purposes. Second, subsequent processing techniques assume that the images are geometrically aligned, i.e. an association between pixel position and facial feature represented by this pixel can be roughly established. The operations comprised in this geometric normalization stage in our tests are:

Fig. 2: From left to right: example images after steps (1), (2) and (3) of the geometric normalization stage.

(1) The original 720×576 X2MVTS images are first converted to 128×128 images, simply by cropping and decimation.
(2) The 128×128 images obtained in (1) are warped so that the 6 reference points fall on fixed positions using thin-plate splines as in [8]. Occasionally, only 2 points (centers of the

iris) are used in the experiments. In this case an affine geometric transformation is used.
(3) A binary mask consisting of an elliptic patch is applied to the result of (2), so that only interior parts of the face meaningful for recognition are kept.

The result of the geometric normalization stage (see Fig. 2) consists of a vector comprising the 2347 grayscale pixels inside the elliptic patch.

### 2.2. Photometric normalization

The following options are proposed and tested here concerning what we call photometric normalization (i.e. techniques applied in order to reduce the influence of factors of variability affecting globally the values of the pixels, such as illumination):

(1) No photometric normalization.
(2) Application of two affine transformations.
(3) Histogram equalization.
(4) Gabor transform.

In option (2), first, for each 2347 vector a transformed vector is obtained so that the mean value is 0 and its standard deviation is 1. Afterwards, a similar transformation is applied, but now the purpose is to have 0 as mean and 1 as standard deviation for each of the components of all the images belonging to the training set. The same transformation is applied to test images (not belonging to the training set).

What we call Gabor transform (4) consists of the application of a set of Gabor filters and posterior subsampling. The Gabor filters are:

$$\psi_j(\mathbf{x}) = \frac{k_j^2}{\sigma^2}\exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\left[\exp(i\mathbf{k}_j\mathbf{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right]$$

where $j$ is the index for the filter used and $\mathbf{k}_j$ its central frequency. 12 Gabor filters were applied to the full 128×128 images (without elliptic patch), corresponding to 4 orientations (0, $\pi/4$, $\pi/2$ and $3\pi/4$) and 3 scales ($\pi/8$, $\pi/4$, $\pi/2$) for the central frequency and $\sigma$=2. The final vector representation had a dimension of 3108 and was obtained retaining only the magnitude of pixel values inside the elliptic patch containing the face in the subsampled images.

Fig. 3: From left to right: example images after photometric normalization schemes (1) and (3). Right image is a reconstruction from Gabor units.

### 2.3. Dimensionality reduction

Two traditional techniques in face recognition, Principal Component Analysis (PCA, [4]) and Linear Discriminant Analysis (LDA) ([5], [9]), have been applied in order to reduce the dimension of the vector. PCA provides an orthonormal reduced basis of vectors (eigenfaces) so that, when training examples are projected onto it, the projections have minimun squared error with respect to the original vectors. LDA provides a linear transformation (by projecting to fisherfaces) that reduces the dimensionality trying to maximize some criteria of separability between classes.

From these techniques, the following options are considered for reducing de dimensionality:

(1) No reduction.
(2) PCA features: reducing just by projecting onto PCA-space (150 components).
(3) White features: reducing by projecting onto PCA-space (150 components) and scaling each component in the PCA space so that all of them have unit variance.
(4) LDA features: reducing just by applying the LDA transformation (180 components).

### 2.4. Similarity measures

For verification purposes, a measure of similarity (a *score* for brief) is needed in order to compare the reference data **m** (in our case, the mean of the vectors with reduced dimensionality corresponding to the claimed identity within the training set) with the data to be tested **t**. In face recognition systems, the measure of similarity is usually fairly simple thanks to the complexity of the feature extraction process [9]. In particular, the following scoring formulae will be tested:

(1) Minus Euclidean distance: $s_1(\mathbf{m},\mathbf{t}) = -\sqrt{\sum_i (\mathbf{m}[i]-\mathbf{t}[i])^2}$
(2) Dot product: $s_2(\mathbf{m},\mathbf{t}) = \sum_i \mathbf{m}[i]\mathbf{t}[i]$
(3) Norm. dot product: $s_3(\mathbf{m},\mathbf{t}) = s_2(\mathbf{m},\mathbf{t}) \big/ \sqrt{s_2(\mathbf{m},\mathbf{m})s_2(\mathbf{t},\mathbf{t})}$

## 3. EXPERIMENTS

### 3.1. Verification performance assessment

Two kinds of measures of the performance of verification systems are provided here: numerical and graphical. First, numerical values correspond to False Alarm (FA) and False Rejection (FR) rates at some working points. A false alarm event arises when an impostor accesses the system claiming the identity of a user. False reject happens when a user, claiming his identity, is not accepted. Graphical representations for the verification performance are provided here in terms of DET curves (which are FA vs. FR rate plots with particular axis scaling) [5].

### 3.2. Database description and protocol

The XM2VTS [10] multimodal database consists of face images, video sequences and speech information of 295 subjects.

Experiments reported here use 8 frontal face images from 4 different sessions for each subject. Lausanne protocol [11] splits the subjects into 3 groups: 200 users, 25 evaluation impostors and 70 test impostors. In this context, verification performance is studied comprising 3 stages: training (the systems learn from particular data of each user how to compute a similarity value as it has been seen in section 2), evaluation (data from users and impostors are used in order to fix threshold values on the similarity score so that FA or FR meet certain requirement on the evaluation set) and testing (final values of FA and FR are measured in a set different from the one used for training or for

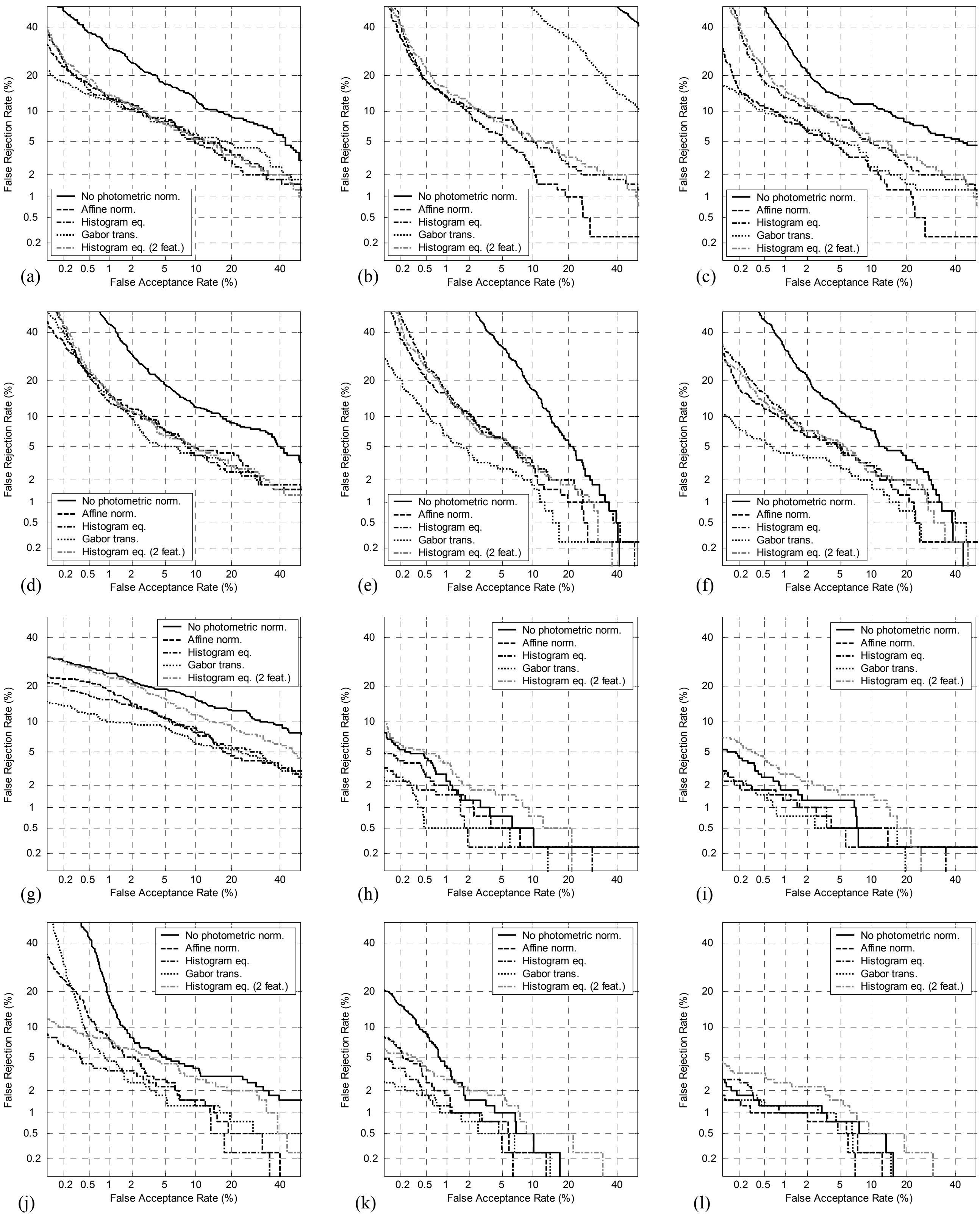


Fig. 4: Performance on the evaluation set (a posteriori thresholds) for different methods of dimensionality reduction (rows), similarity measure (columns) and photometric normalization (see legends in each plot). From upper to lower row: no dimensionality reduction, PCA, white features and LDA. From left to right column: euclidean distance, dot product and normalized dot product.

fixing the thresholds). Configuration II from this protocol employs the following statistics: 4 images from each user (200×4) are used for training, corresponding to sessions 1 and 2. For evaluation, 2 images per user (session 3, 200×2=400) are used to compute FR and all the available images from evaluations impostors (25×8=200) to compute FA. *A posteriori* thresholds on the similarity values can be computed then to meet specific requirements over FA and FR in the evaluation set. For testing, 2 images per user (session 4, 200×2=400) are used to compute FR and all the available images from test impostors (70×8=560) to compute FA. Final values of FR and FA are computed with the thresholds established *a priori* in the evaluation set.

### 3.3. Results

Detailed results for *a posteriori* decision threshold selection (on evaluation data) are provided in Fig. 4 as a grid of DET plots, where the wide range of performances attainable when combining in different ways the available options described so far is remarkable.

Results on test data with thresholds established *a priori* over the evaluation set according to Lausanne protocol for selected options are given in Table I, where the combination with best verification performance have been highlighted.

TABLE I
PERFORMANCE ON THE TEST SET FOR A PRIORI THRESHOLDS

| | | FRE=0 | | FRE=FAE | | FAE=0 | |
|---|---|---|---|---|---|---|---|
| System (Fig. 4) | Phot. Norm. | FR (%) | FA (%) | FR (%) | FA (%) | FR (%) | FA (%) |
| (a) | No Phot. | 0.0 | 90.0 | 11.6 | 12.5 | 95.0 | 0.0 |
| (e) | Affine | 0.5 | 55.7 | 5.8 | 5.8 | 91.7 | 0.0 |
| (h) | Hist. eq. | 0.3 | 28.7 | 0.5 | 1.9 | 20.9 | 0.0 |
| (h) | Hist. eq. (2 feat.) | 0.5 | 20.9 | 2.3 | 2.4 | 43.7 | 0.0 |
| **(h)** | **Gabor** | **0.3** | **14.3** | **1.0** | **0.9** | **11.8** | **0.1** |
| (l) | Gabor | 0.3 | 15.8 | 1.0 | 1.9 | 9.8 | 0.0 |
| (l) | Hist. eq. (2 feat.) | 0.3 | 30.6 | 0.8 | 3.0 | 12.3 | 0.0 |

## 4. CONCLUSIONS

Our performance results (see highlighted row in Table I) show competitiveness with respect to other referenced works [3] on same database with same experimental protocol. Two fundamental differences from referenced systems [3] are our geometric normalization stage (based on 6, instead of 2, manually marked reference points) and the Gabor photometric normalization scheme.

With regard to geometric normalization, we have included results in Fig. 4 and Table I using only 2 reference features (eyes) in order to test the performance worsening. This result encourages the research in full automatic geometric normalization schemes for face recognition when global appearance representation strategies are used.

In relation to photometric normalization, we have shown the verification performance improvement of several techniques, being the Gabor-based method the best of them in almost every case. Dimensionality reduction techniques have also been shown to increase verification performance, yielding PCA+whitening and LDA similar results for best performing similarity measures, which are dot product-based.

## 5. ACKNOWLEDGEMENTS

This work has been supported by the Spanish Ministry of Science and Technology under project TIC2000-1669-C04-01.

## 6. REFERENCES

[1] J. Ortega-García, *et al.*, "From Biometrics Technology to Applications regarding Face, Voice, Signature and Fingerprint Recognition Systems", in *Biometrics Solutions for Authentication in an E-World*, (D. Zhang, ed.), pp. 289-337, Kluwer Academic Pub., 2002.

[2] P.J. Phillips, *et al.*, "The FERET Evaluation Methodology for Face-Recognition Algorithms", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, nº 10, pp. 1090-1103, October 2000.

[3] J. Matas, *et al.*, "Comparison of Face Verification Results on the X2MVTS Database", Proc. of IEEE 15th Intl. Conf. on Pattern Recognition, vol. 4, pp. 858-863, Los Alamitos, USA, September 2000.

[4] M. Turk, A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, vol. 3, pp. 71-86, 1991.

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

[6] L. Wiskott, J.M. Fellous, N. Krüger, C.v.d Malsburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, nº 7, pp. 775-779, 1997.

[7] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", Proc. of the IEEE, pp. 705-740, 1995.

[8] A. Lanitis, C.J. Taylor and T.F. Cootes, "Automatic Interpretation and Coding of Face Images Using Flexible Models", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, nº 7, pp. 743-756, July 1997.

[9] J. Kittler, Y.P. Li, J. Matas: "On matching scores for LDA-based face verification", Proc. of British Machine Vision Conf. 2000, M. Mirmehdi and B. Thomas (eds.), pp. 42-51, 2000.

[10] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre, "XM2VTSDB: The extended M2VTS database". Audio- and Video-based Biometric Person Authentication, AVBPA'99, pp. 72-77, 1999.

[11] J. Luettin, G. Maitre, "Evaluation protocol for the extended M2VTS database (Lausanne protocol)", Technical Report 98-05 from IDIAP, 1998.
ftp.idiap.ch/pub/reports/1998/com98-05.ps.gz.

[12] A. Martin, *et al.*, "The DET Curve in Assessment of Decision Task Performance", Proc. of ESCA 5th Eur. Conf. on Speech Comm. and Tech., Eurospeech '97, pp. 1895-1898, 1997.