ON TESTING METHODS FOR BIOMETRIC AUTHENTICATION

Enrico Grosso and Massimo Tistarelli

University of Sassari
Computer Vision Laboratory
Palazzo del Pousalid, p.zza Duomo, 07041 Alghero (SS) – Italy
e-mail:tista@uniss.it – grosso@uniss.it

ABSTRACT

The use of biometric data for user authentication and/or recognition is now a reality. On the other hand, there is still a strong need for new technologies to overpass intrinsic limitations of already "established" techniques. This not only requires to devise new algorithms but to determine the real potential and limitations of existing techniques. This is possible only devising standard testing and assessment procedures based on statistical observations of the outputs of the system. In order to define better a standard evaluation process, a system based on space-variant iconic image matching is described and the validation procedure defined. I turns out that all methods based on the same biometric measurements have the same intrinsic limitations, which can be only overcome by the adoption of a multi-modal or multi-algorithmic approach.

1. INTRODUCTION

Many efforts have been devoted to the study of computer systems for automatic verification of person's identity (either by means of recognition from a database of known individuals or as authentication of one's identity).
This relatively new technology has an indubitable potential: surveillance, secure access control and e-commerce are just few of the possible envisaged applications.
In principle (particularly for social acceptability) the analysis of face images seems to be the best way to accomplish the task of determining the personal identity. Many difficulties arise from the enormous dimensionality of the search space when dealing with natural images (both for the number of elements in a typical data set and for the number of samples for each data). These and other issues related to the definition of "best" similarity measurements for complex shapes like face images, make face recognition and visual authentication a still open and challenging problem in computer vision [1,2,3,4,5,6,7,8].
On the other hand, the real assessment of a new technology is generally determined more by its use than by empirical figures and theoretical projections. This is possible, not just benchmarking an algorithm with respect to the more favorable working conditions, but rather with the most probable working conditions which also constitute the minimal impact or invasiveness for the users [9,10].
Towards this end this paper tries to develop a framework to assess the real performances of a given face recognition system, regardless of the matching engine and face representations used to compare different subject's descriptions.

2. BIOMETRIC AUTHENTICATION

2.1. Analysis of matching techniques

A first technique has been tested where a collection of fixations from the face image is used to represent a subject. The matching is performed by computing the correlation between the representation of the reference subject and the one requesting the access. The algorithm is based on the following steps:

1. Given the position of selected facial features (the eyes and the mouth), three log-polar fixations are extracted from the acquired image of the subject.
2. The log-polar images are warped to simulate views which are as close as possible to the pose and orientation of the reference subject's face (generally parallel to the image plane).
3. Corresponding fixations are compared by computing the sum of the absolute value of gray level differences[1] and the normalized correlation. Two matching scores are obtained from each fixation independently.

[1] To compute the difference, the gray levels of each log-polar fixation are first normalized to the range of intensity values of the corresponding facial feature of the reference subject.

4. The scores obtained by the log-polar fixations are combined to form a 6 components vector representing the similarity between the subject and the model.

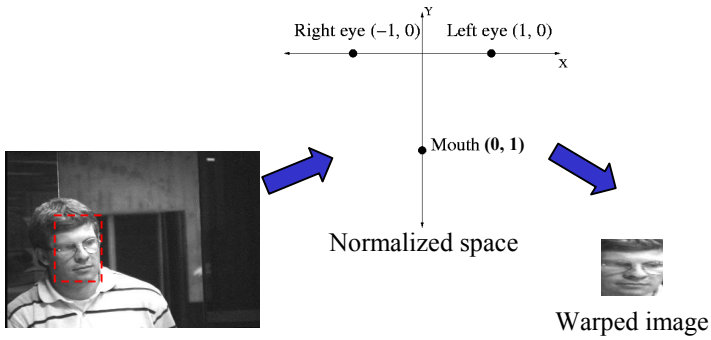

Figure 1 Extraction of the face window and warping.

A simpler technique performs the matching on just a single window containing the whole face in Cartesian coordinates. As a major problem with template matching is the registration of the two images, the window is warped according to a feature space determined by the position of the facial features. Therefore, in this case, the facial features are not used to extract sub-windows out of the subject's face but rather to align and scale the face with the model image.

2.2. Extraction of facial features

The position of the facial features is determined in two steps:

- by first computing the cumulative values of the filtered image along the rows. The eyes correspond to the area with higher cumulative values;
- the same process is performed along the columns in the area corresponding to the eyes, determined at the previous step. Again the two maxima correspond to the horizontal position of the two eyes.

In order to avoid false matches a geometrical constraint is enforced to the position of the eyes and mouth, which is to lie at the vertexes of a triangle. The values assumed by the angles of the triangle are bounded by values determined experimentally ($44° < \alpha_i < 84°$).
The exact position of the mouth is finally determined by computing the cross-correlation between the image and a feature template, within a 10x10 pixels window centered on the previously determined position. The template is obtained by just cutting the eyes and mouth out of a sample image of an unknown subject outside the test database, but with the facial features clearly evident. From an extensive test it has been this choice demonstrated to gives more accurate results than computing an average template. This is due to the fact that the averaging process deforms considerably the feature's shape degrading the matching results.
The three correlation values stemming from the eyes and mouth are averaged to obtain a score between –1 and 1. If the geometric constraint is satisfied and the matching score is higher than a given threshold the fixations are considered as valid ones. In order to determine the discriminant value for the correlation score, a validation test has been performed on a set of 2019 images completely uncorrelated from the recognition database. These images have been divided into two classes:

- all images (1609) where the facial features are partially occluded or not visible, plus all the images where the mean difference between the estimated and the manually determined feature positions is greater than a given threshold[2];
- all remaining images in the set (410).

The FAR and FRR test values were computed from the feature correlation scores of the two image sets. These statistical measures represent the capability of separating the two classes, or to determine whether the features can be accurately localized or not. The score value corresponding to equal FAR and FRR is taken do decide if the estimated features positions are reliable and can be used to proceed in the face matching process. Otherwise the face image is discarded

2.3. Comparison of matching techniques

Once the similarity scores are computed a statistical classifier can be used to determine the similarity between a given model and the subject. Even though this is still an open and crucial problem in biometric authentication in this paper the task of optimal classification is not addressed. On the other hand, a framework is devised to determine, from the raw output data, the real potential and performances of any iconic-based face recognition system. In principle it may be extended to address any face matching system. Two different schemes are considered:

1. image matching performed on three independent fixations and two independent similarity measurements;
2. image matching performed on a single window centered on the subject's face and warped

[2] This threshold is determined statistically by computing the probability of locating the facial features correctly in more than 50% in a given image ensemble.

according to a common reference frame.

The system's performances are greatly influenced by the accuracy in the estimation of the position of facial features. For this reason two approaches were tested for feature detection: the former based on matching a generic template of the facial features, the latter applying a specific template extracted from the image of the model face, which correspond to the features of the subject to be recognized. The second approach effectively maximizes the probability of correct feature localization for the subject to be recognized.

3. A GENERAL TESTING PROTOCOL

In order to define a common test bed, a complete matching is performed over all the images in the data set (all subjects versus all images) Given *M* images for each of *N* subjects, the results obtained can be divided into two classes[3]:

- matching scores obtained comparing all different images of the same subject, equal to *N x M x (M - 1)* comparisons (*client tests*);
- matching scores obtained comparing all different images of different subjects, equal to $N \times M^2 - N \times (N + M - 1)$ comparisons (*impostor tests*);

it is assumed that more than a single score is available for each image comparison (i.e. more than one measurement is performed, for example matching the whole face and also small windows from the same image). A covariance matrix is defined describing each of the two classes:

$$\Sigma_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(x_j^i - m_i\right)\left(x_j^i - m_i\right)^t$$

where N_i represents the number of elements of class "*i*" and m_i is the mean value of the same class. If the measurements are independent the rank of the covariance matrix corresponds to the number of measurements used.
From the two classes it is possible to define the inter-class and intra-class discrimination capability of the matching algorithm. Given the entire ensemble of matching scores for the two classes (each score can be regarded as a vector within the class), the discrimination power can be defined through three statistical indexes:

- The intraset and interset distances (class separability indexes).
- The Bayesian error probability.
- The false acceptance, false rejection and the equal error rate (FAR, FRR, EER).

The first two indexes define the distances among the elements of the same class and between the two classes. By comparing the two it is possible to define the separability between the two classes, e.g. to discriminate the set of clients from all the impostors. Given the intraset distances R_1 and R_2 , computed as the mean distances between all matching vector pairs[4] in the two classes, and the interset distance *H*, computed as the mean distance among all vectors in the two classes, for a good separation between the two classes the intraset distances are expected to be much smaller than the interset distance:

$$Q = \frac{R_1 + R_2}{H}$$

a low value of *Q* means the two classes are well separated.
Another separability measure is given by the Bhattacharrya distance [11]:

$$\beta = \int \sqrt{p\left(x/\omega_1\right)p\left(x/\omega_2\right)}dx$$

where *x* is the measurement vector, ω_1 and ω_2 are the two classes and β is bounded between 0 and 1. If either of the two conditional probabilities *p(x/y)* is equally zero the two classes are very well separated, while if the product is equal to one the two classes are superimposed. Consequently, the smaller the value of β the higher the separability between the two classes.
Assuming the probability density of the measurements vectors to be Gaussian, it is possible to compute:

$$B = -\ln\beta = \frac{1}{8}\left(m_1 - m_2\right)^t\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}\left(m_1 - m_2\right) + \frac{1}{2}\ln\left(\frac{\left|\frac{1}{2}\left(\Sigma_1 + \Sigma_2\right)\right|}{\sqrt{\left(\left|\Sigma_1\right|\left|\Sigma_2\right|\right)}}\right)$$

where m_1 and m_2 are the mean measurement vectors of the two classes (clients and impostors), Σ_1 and Σ_2 are the covariance matrices of the measurements of the two classes. There is a close relationship between the Bhattacharrya distance and the Bayesian error probability:

$$P_e \le \frac{1}{2}\beta$$

[3] As a consequence of the experimental procedure the training set and the test set are disjoint, except for the case where the image used to build the representation of one subject is also used for an impostor test.

[4] A matching vector is defined as the set of matching scores obtained from the matching engine of the system. The vector can be composed of a single element, if the matching involves a single facial feature, or many elements

	Correlation-based matching		Commercial system based on LDA
β	0.61	0.73	0.63
Pe	30.5%	36.5%	31.5%
1/Q	0.717	0.8	0.83
FAR	15%	21.5%	19.68%
FRR	17%	21.3%	20.34%

Table 1. Comparison between the performances of the face matching system described in section 2.1 and a commercial system based on LDA. The two columns at left are related to two different databases: the former with cooperative subjects, the latter with non-cooperative subjects.

A well known method for multivariate system analysis is the Fisher transform, which allows to project a vectorial function on a one-dimensional space. Through this technique it is possible to analyze the distributions of the measurement vectors of the two classes as two one-dimensional functions. The matching scores are used to determine the Fisher vector:

$$w = S_w^{-1}(m_1 - m_2)$$

$$S_w = S_1 + S_2 \quad ; \quad S_i = \sum_{j=1}^{N} \left(x_j^i - m_i\right)\left(x_j^i - m_i\right)^t$$

where N_1 and N_2 represent again the number of elements in each class and the other terms are defined as in the previous equations. The measurement vectors are projected on the Fisher's vector and the distribution of the two classes are computed. The resulting curves represent the probability densities of the missed clients and the allowed impostors, as a function of the matching score. The integrals of the two curves represent the FAR and FRR. From the resulting representation two results are inferred:

- The equal error rate of the system, which is the probability of equally accepting an impostor or rejecting a client. This measure is computed as the probability corresponding to the coordinate, on the horizontal axis, where the two probability density functions have the same area.
- The best discriminant threshold, which is the threshold to be applied to the computed matching scores to assure the best separation between the two classes. This is determined by the horizontal coordinate corresponding to the intersection point between the two curves.

Both these parameters define the goodness of the identity verification system, but the second one can also be applied as a threshold to perform recognition.

4. CONCLUSION

The validation and testing of face authentication systems is still an open problem. This is very important to assess the real performances of a biometric system. This paper presented some specifications with the aim of defining a test protocol to be applied to any image-based face authentication system.

The proposed protocol may be extended to more statistical tests (five are proposed here) maybe defining even better the separability of the client and impostor classes. Extensive experiments made on real life face images demonstrated the weakness of feature detection even for algorithms not based on feature matching. This is due to the fact that face registration is always necessary before computing the distance between faces. It has been shown how the test protocol can be also applied to recover the correct facial features and discard false matches.

5. REFERENCES

[1] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection". *IEEE Trans. on PAMI*, PAMI-19(7):711-20, 1997.

[2]S.J. McKenna and S. Gong. "Real-Time Pose Estimation". *Real-Time Imaging*, 4{5}:333-47, 1998.

[3] G.J. Edwards, T.F. Cootes and C.J. Taylor. "Face Recognition Using Active Appearance Models". In *Proc. of 5th European Conference on Computer Vision*, pp 581-95, Springer Verlag, 1998.

[4] R. Brunelli and T. Poggio. "Face recognition through geometrical features", In *Proc. of 2nd European Conference on Computer Vision*, pp 792-800, S. Margherita Ligure (Italy), 1992. Springer Verlag.

[5] Mogbaddam B., Jebara T. and Pentland A. "Bayesian face recognition". *Pattern Recognition*, 33(11):1771-82, Nov. 2000.

[6] Feng G.C., Yuen P.C., and Dai D.Q. "Human face recognition using pca on wavelet subband". *Journal of Electronic Imaging*, 9(2):226--33, 2000.

[7] Liu C. and Wechsler H. "Evolutionary pursuit and its application to face recognition". *IEEE Transaction on PAMI*, 22(6):570-582, 2000.

[8] M. Tistarelli. "Active/space-variant object recognition". Image and Vision Computing, 13(3):215-226, 1995.

[9] Phillips P.J., Wechsler H., Huang J. and Rauss P.J. "The feret database and evaluation procedure for face-recognition". *Image and Vision Computing*, 16(5):295-306, 1998.

[10] Phillips P.J., Hyeonjoon Moon, Rizvi S.A. and Rauss P.J. "The feret evaluation methodology for face-recognition algorithms". *IEEE Transaction on PAMI*, 22(10):1090-104, 2000.

[11] Grother P.J. "Cross validation comparison of NIST ocr databases". In *Proceedings of the SPIE*, volume 1906, pages 296-307, 1993.