Multimodal Biometric Person Authentication*

Stéphane Pigeon – Luc Vandendorpe
UCL - Laboratoire de Télécommunications et Télédétection
Place du Levant, 2 – B-1348 Louvain-La-Neuve – Belgium
E-Mail: pigeon@tele.ucl.ac.be

January 12, 1999

Abstract

An authentication system checks the identity of a person who identified himself beforehand. This paper deals with an authentication by means of biometric experts running on facial images and voice recording. Its principal characteristic is to defer the decision to accept or reject a person at a higher integration level called the supervisor.

1 Introduction

This contribution describes a biometric person authentication system aimed at being easily implemented in a multimedia PC platform, and objectively studies its performance level in terms of False Rejections (i.e. the percentage of client accesses rejected by the system) and False Acceptances (i.e. the percentage of impostor accesses accepted by the system). In order to offer an increased robustness against the possible biometric changes which can affect a face day by day, multiple modalities have been taken simultaneously into account, namely two face-based modalities (the face as seen from the profile and the frontal views) and a speech-related modality. The information provided by these modalities - also referred as "experts", as each modality develops an expertise for a particular set of biometric features - is sent to the "supervisor". The supervisor implements the fusion of the different experts opinions and takes the final decision to accept or reject a given client identity claim. This work has been divided into two major sections. The first section describes the experts related to the profile, frontal and vocal modalities, and provides their individual performance. The second section is dedicated to the supervisor and compares various fusion techniques in different contexts. In this second section, we will show how a statistical approach is able to increase the robustness of the supervisor.

2 The biometric experts

2.1 Profile Shape Matching

The first modality consists of the authentication of the profile outline and is inspired from [1]. The algorithm is based on a chamfer matching that directly works on the profile contour encoded as x-y coordinates. The chamfer matching technique searches for the

*This work has been performed in the framework of the M2VTS Project (European ACTS programme).

best match between two binary images. Geometric transformations are used to distort one image (here referred to as the *candidate image*) to another (the *reference image*) in order to minimize a given distance measure between them. These binary images are often derived from the image edges. Here, we make use of the shape of the profile. The first step of the algorithm is to generate a *distance map* from the reference profile. This distance map associates with each pixel of the reference profile picture, its distance from the closest profile pixel (all profile pixels get thus the zero distance value). As the true Euclidian distance is costly to compute, we use a *sequential chamfer distance approximation* [2]. The use of a distance map drastically speeds up the distance computation between two given profile shapes as the global distance between these shapes is simply computed by summing all distances found along the first shape when superposed to the other's distance map. The global matching process is illustrated in Figure 1. First, the candidate profile is projected onto the reference distance map and a global distance is computed. By minimizing this distance – we used the simplex algorithm [3] – the optimum compensation parameters are found (translation, rotation and scale factor). Then, the residual distance between the best compensated and the reference profiles is sent to the supervisor to decide whether the two profiles belong to the same person or not.

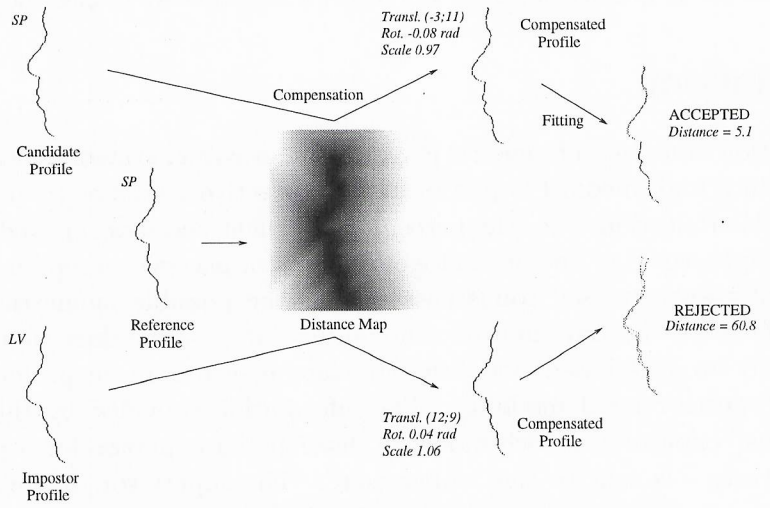

Figure 1: The Chamfer Matching Process

2.2 Grey Level Profile Matching

A second profile-related modality is based on the grey level information along the shape of the profile and includes features like mouth width and height, nostrils, nose depth, eyes and eyebrows as accessed from the profile view. Once the best compensation parameters have been found during the chamfer matching process, the same parameters are used to compensate the candidate profile grey level image in order to issue a pixel-by-pixel comparison with the grey levels of the reference image. The Mean Squared Error (MSE) is used to express the distance between the reference profile and the compensated candidate. Prior to the distance computation, one has first to normalize the grey level distribution between the two images to get rid of the illumination variability . A simple mean luminance normalization has been chosen. Due to the presence of hair, the whole profile view cannot be used to carry out the comparison and only a small area taken along the profile shape

has to be taken into account. Images are low-pass filtered prior to the matching. Examples are shown in Figure 2.

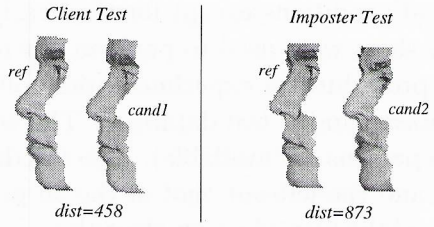

Figure 2: Grey level matching

2.3 Grey level Frontal Face Matching

Our frontal-based modality is similar to the grey level profile matching, in the sense that it also computes the grey level correlation between a candidate and a reference image, but using information from the frontal view instead of the profile. Prior to the matching itself, the same grey level normalization as described above is performed. Images are low-pass filtered in both x/y directions in order to improve the quality of the matching. Again, the MSE criterion is used to compute the distance between the two grey level images. In order to get rid of the face variability over the different shots, the grey level distance is computed inside a rectangular window that covers the most invariant features found inside the frontal view, namely the eyes/eyebrows and nose/nostrils features. This fixed window is automatically extracted from the input images using a technique that is similar to the technique proposed by [4]. Results are illustrated in Figure 3. Unlike the grey level profile modality that makes use of the compensation parameters issued form the chamfer shape matching, we don't know a priori which parameters to apply in order to match the candidate's eye/nose window onto the reference image. Again, we will use the simplex algorithm in order to find the optimal frontal compensation parameters, and minimize the grey level distance between the two images.

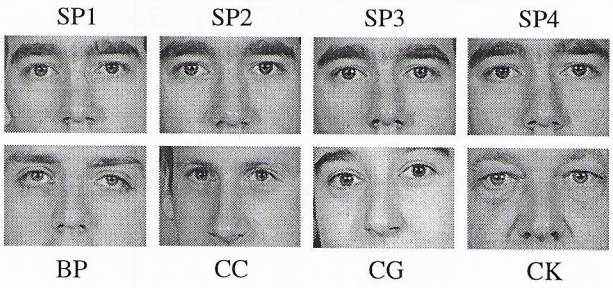

Figure 3: Frontal biometric expert : examples of matching windows

2.4 The Vocal Expert

The vocal expert used in this paper has been designed by the IDIAP [7] and is based on Hidden Markov Models. This method proved to be very powerful and is currently widely used in the field of speaker authentication. Further information can be found in [7]

2.5 Test Setup

The M2VTS Multimodal Face Database [5] has been used to test the methods proposed here. This database is made of 37 faces, offers an overall resolution of 286x350 pixel and has been acquired under real conditions except for the nearly constant lighting and fixed background. Four different shots were used to perform our tests. These shots were taken at one week intervals. Our procedure for experimentation follows the M2VTS protocol [5]. One experiment uses a *training* and a *test* database. The *training* database is built of 3 shots (4 are available) of 36 persons (37 available). The *test* database is built of the left-out shot of the left-out person and the left-out shot of the 36 persons present in the training database. The performance of the identification algorithms is evaluated by matching the 37 candidate persons (36 clients and 1 impostor) from the test database with the 36 reference clients. Such an experiment provides 36 *authentic* and 36 *imposture* tests. There are 4*37 possible experiments by leaving out one person and one shot, this means 4*37*36=5,328 client matches and 5,328 impostor matches. All these 10,656 matches were performed in order to evaluate the performance of our different algorithms and fusion schemes. At last, authentication results are computed by matching the candidate (taken from test set) with each of its claimed references (the 3 shots taken from the training set) and by taking the best score (i.e. the lowest residual distance) as the final score.

2.6 Performance of the Single Modalities

A given match is accepted if its score (e.g. residual chamfer or grey level distance) is below a given threshold. Otherwise, it is rejected. For each threshold, some clients are rejected (i.e. the false rejection $FR(k)$) while a number of impostors are able to enter the system under a client identity (i.e. the false acceptance $FA(k)$). For the purpose of the supervisor design, which better works fusing of relatively *independent* modalities, we will combine the two profile-related modalities into one unique *profile expert*, by summing up their scores. The frontal grey level matching is referred as the *frontal expert*. Table 1 summarizes the performance of the different experts by giving the *Equal Error Rate* (EER) (stands for the point where $FA = FR$) and the *Success Rate* (SR) (operating point where $1 - FA - FR$ reaches a maximum).

	EER	SR
Profile Expert	8%	85.5%
Frontal Expert	9%	83%
Vocal Expert	1.5%	97.5%

Table 1: Performance of the different experts

3 The Supervisor

Fusion performance can be expressed in both way : *a posteriori* results and *a priori* results. An "a posteriori" study characterizes the best performance that one is able to achieve given a particular fusion operator and two sets of client and impostor claims. Such study has been performed in [6] and involved hard fusion schemes (decision fusion : AND/OR fusion schemes) as well as soft fusion schemes (score fusion : linear combination of scores). A "a

priori" scenario tests the different supervisors under real operating conditions and involves new clients and impostors compared to the ones used at training time. This ensures that the data used for testing the supervisor is really independent from what has been used to build the supervisor's client and impostor "a priori" knowledge. The *supervisor training* and *supervisor test* sets have been generated by dividing the expert test subset in two groups, namely users 1-18 and users 19-37 (left out individuals relative to a given group are used as importers inside the same group).

3.1 Exhaustive supervisor design

The first supervisor family studied in this work implements a simple and rather intuitive technique. On the basis of the training data, the internal parameters of the supervisor are optimized in order to maximize its performance. These parameters are thresholds in case of AND/OR decision fusion, or weights in case of a linear score fusion, for example. The search for the best supervisor parameters is made in an exhaustive way, i.e. by considering all possible combinations and by retaining the one which optimizes a given performance criterion. Once these parameters are fixed, the supervisor is evaluated on the test set. We will call such supervisor, an *exhaustive* supervisor. Although the results achieved on the supervisor training set are remarkable, the performance obtained on the test set is disappointing (i.e. not much better than the best expert). This gives reasons to think that the exhaustive training technique proposed here is far from being robust and that slightly different client/impostor score distributions are able to perturb the exhaustive supervisor (this phenomenon is called *overtraining*: a supervisor can take so well into account the particular characteristics of the training set that it becomes unable to deal with another set which offers slightly different characteristics).

3.2 Statistical supervisor design

The statistical supervisor that has been used here, is based on Fischer's work and makes use of a linear decision border to separate two given populations (the clients and the impostors in our case). Let us denote π_c and π_i the client and impostor populations respectively. An individual is randomly taken among one of these two classes and is authenticated using p experts. Each expert provides a score z_n, $n = 1..p$ and $z = [z_1...z_p]'$ represents the vector of the observed scores. From z, we must decide whether the candidate is a client or an impostor. Let us denote $f_c()$ and $f_i()$ the probability density functions of vector z within the π_c and π_i populations respectively. A simple decision rule classifies the candidate among the clients if

$$f_c(z)/f_i(z) \geq k \tag{1}$$

where k represents an acceptance threshold whose value depends on the FA/FR compromise that one wants to achieve. In our problem, $f_c()$ and $f_i()$ are unknown and must be estimated from the training data. A common assumption consists in approaching the real distributions by normal distributions. The mean and variance parameters are unknown, but may be estimated from the training data. Equation (1) can then be rewritten as follows : $D_L(z) \geq ln(k) = k^*$, where

$$D_L(k) = (z - \frac{1}{2}(\widehat{\mu}_c + \widehat{\mu}_i))' \ \widehat{\Sigma}^{-1} \ (\widehat{\mu}_c - \widehat{\mu}_i) \tag{2}$$

Σ represents the covariance matrix computed across both imposter and client training scores. Fisher has been the first to use this function for classification goals. As $D_L(z)$ is linear in z, it has been commonly called *Linear Discriminant Function* (LDF). The procedure to be followed in order to check the identity of a candidate consists thus to calculate $\widehat{\mu}_c$, $\widehat{\mu}_i$ and $\widehat{\sigma}$ by using the training data (what is carried out once and for all), then $D_L(z)$ and to compare $D_L(z)$ with the threshold k^*. If $D_L(z) \geq k^*$, the candidate is accepted.

4 Conclusions

Table 2 summarizes the performance achieved by the statistical supervisor and highlights the robustness of the suggested method: the performance predicted from the training procedure is quite faithful to what is really obtained. Also, the performance level obtained during the test phase is always better than what could be achieved using an exhaustive supervisor, and particularly when fusing all available experts. In such a case, we obtain a success rate of 99.2%. The error rate which we would have observed by making use of the best expert, i.e. the vocal expert, is thus reduced by a factor three.

Experts	Criteria	Learn	Test
		FA FR	FA FR
Pro+Front	Min EER	EER 4.7	4.0 4.7
	Max SR	SR 93.4	SR 90.1
Pro+Front+Voice	Min EER	EER 0	0.2 0.7
	Max SR	SR 100	SR 99.2

Table 2: Performance of the statistical supervisor (values expressed in %).

References

[1] S. Pigeon and L. Vandendorpe, "Profile Authentication Using a Chamfer Matching Algorithm", *Proceedings AVBPA '97*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 185-192.

[2] Gunilla Borgefors, "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 6, Nov. 1988, pp. 849-865.

[3] W. H. Press, S. A. Teukolsky and W. T. Vertterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1988.

[4] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, Oct. 1993, pp. 104-1052.

[5] S. Pigeon and L. Vandendorpe, "The M2VTS Multimodal Face Database (Release 1.00)", *Proceedings AVBPA '97*, Crans-Montana, Switzerland, March 12-14 1997, pp. 403-409. See also *http://www.tele.ucl.ac.be/M2VTS/*

[6] S. Pigeon and L. Vandendorpe, "Image-based Multimodal Face Authentication", *Signal Processing*, Vol. 69, no. 1, pp. 59-79, October 1998.

[7] P. Jourlin, J. Luettin, D. Genoud and H. Wassner, "Acoustic-labial Speaker Verification". *Pattern Recognition Letters*, Vol. 18, no. 9, pp. 853-858, September 1997.