3314-16

MULTIPLE EXPERTS FOR ROBUST FACE AUTHENTICATION

Stéphane Pigeon - Luc Vandendorpe*
UCL - Communications and Remote Sensing Laboratory
Place du Levant, 2 - 1348 Louvain-la-Neuve - Belgium

Most of the multimodal authentication schemes currently developed, combine speech and image-based features together and thus benefit from the high performance offered by the speech modality. Depending on the application, speech data is not always available or cannot be used. This paper takes these cases into account and investigates the best performance that can be achieved by a system based on facial images only, using information taken from both profile and frontal views.

Starting from two different profile-related modalities, one based on the profile shape, the other on the grey level distribution along this shape, we will issue a first profile-based expert whose performance is improved compared to each profile modality taken separately. A second expert will use the most invariant part of the frontal view, namely information from a rectangular grey level window centered around the eyes and nose features, in order to issue a frontal-based authentication. Different fusion schemes are studied and the best approach will be applied in order to efficiently combine our two experts. This will result in a robust image-based person authentication scheme that offers a success rate of 96.5% measured on the M2VTS multimodal face database.

Multiple Experts for Robust Face Authentication

Stéphane Pigeon and Luc Vandendorpe

Université catholique de Louvain
Communications and Remote Sensing Laboratory
Place du Levant 2, B-1348 Louvain-la-Neuve, Belgium

ABSTRACT

Most of the multimodal authentication schemes currently developed, combine speech and image-based features together and benefit from the high performance offered by the speech modality. Depending on the application, speech data is not always available or cannot be used. This paper takes these cases into account and investigates the best performance that can be achieved by a system based on facial images only, using information taken from both profile and frontal views. Starting from two different profile-related modalities, one based on the profile shape, the other on the grey level distribution along this shape, we will issue a first profile-based expert whose performance is improved compared to each profile modality taken separately. A second expert will use the most invariant part of the frontal view, namely information from a rectangular grey level window centered around the eyes and nose features, in order to issue a frontal-based authentication. Different fusion schemes are studied and the best approach will be applied in order to efficiently combine our two experts. This will result in a robust image-based person authentication scheme that offers a success rate of 96.5% measured on the M2VTS multimodal face database.

Keywords: multimodal person authentication, face, profile view, frontal view, fusion

1. INTRODUCTION

With the fast emergence of multimedia networks and their new services, the communication concept drastically evolved these last years. As the security issue has often been poorly addressed, the nature of the applications available over the network is mainly limited to non-commercial applications. Today, one has to reconsider the issue of secured access to local or centralized services taking the advantage of the latest breakthrough offered by the ever-growing multimedia environment. The main objective is to extend the scope of applications of network-based services by adding novel functionalities, enabled by multimodal verification strategies based on speech and face images. The aim of this paper is to develop an efficient face authentication algorithm, allowing to remotely verify the identity of a user using images taken from a distant camera. Before further describing the system developed here, let us first present some latest person recognition systems and their respective performance.

Basically, these systems can be divided into two main groups:

- systems that need a close or physical contact with the user like fingerprint analysis, hand shape recognition or iris recognition. These systems are often not well accepted by the users due to the close contact constraint but offer a high performance. They can hardly be used in a multimedia environment where the only sensors are a low-cost camera and/or a microphone.
- distant systems, dealing with an image captured from a distant camera or sound from a microphone. These systems are cheaper to implement (e.g. they can be software-based on a personal/multimedia computer), generally better accepted by the users for their convenience, but cannot compete with the performance offered by the first category. Therefore, multimodality (i.e. a combination of algorithms which make use of non redundant feature sets) is a key point in order to increase the efficiency and robustness of these methods.[1–4]

Further author information -
Stéphane Pigeon: Email: pigeon@tele.ucl.ac.be; WWW: http://www.tele.ucl.ac.be/PEOPLE/sp.html; Telephone: +32-10-478066; Fax: +32-10-472089
Luc Vandendorpe: Email: vdd@tele.ucl.ac.be; WWW: http://www.tele.ucl.ac.be/PEOPLE/lv.html; Telephone: +32-10-472312; Fax: +32-10-472089

Having in mind the development of a system that runs on multimedia platforms, we will restrict this introduction to systems that belong to the second category only. These systems are mainly based on frontal and profile view images as well as speech recordings. Before giving an overview of the performance that can be expected from these modalities, we would like to stress out how difficult it is to issue a comparison between the performance achieved by different algorithms, as referenced by their respective authors or in.[5] First, a wide range of databases is used. While some face databases well represent the data available in real applications, other are more questionable and often make use of a test set that has been acquired during the same recording session as the one used to build the reference models. This results in staggering - but meaningless - performance measures. On the other hand, databases like the FERET database (static images)[6] or the M2VTS multimodal database (video sequences and sound),[7] offer a good material in order to test the expected performance of authentication algorithms in real life scenarios.

Apart from the database issue, another source of disparity between the performance measure of different algorithms consists in the way this performance has been evaluated. As a major example, the performance of a recognition algorithm can be expressed in terms of *identification* or *authentication* efficiency.

An identification system compares the biometric features of the person to identify with all the entries of a client database. The identity of the client who has the closest feature set is assigned to the candidate. Such a system needs an important computing time (all entries have to be checked) but performs well in terms of correct recognition ratio. Indeed, even if the features of one client may vary over time, the client can still be correctly identified as long as the difference between the current feature set and the reference set is smaller than the inter-user distances. Unfortunately, such a system does not deal with the problem of possible imposters. Any imposter will be able to enter the system under the identity of the client whose feature set is the closest to the imposter. Even if an acceptance threshold is defined (a threshold on the feature distance above which the identification is rejected), the risk that an imposter can gain access to the system increases with the size of the client database. The performance of identification algorithms described in the literature are often close to 100%, but unfortunately information about their behavior against imposter access is rarely commented. Brunelli and Poggio[8] report an identification rate of 90% using geometrical features extracted from the frontal view image of 47 people. This rate rises to 100% for a template matching running on the same database. The Moghaddam and Pentland approach[9] is based on eigenfaces and offers an identification accuracy of 99% using frontal views of 155 individuals. Yu et al[10] report 100% correct identification over a database of 33 persons, by using fiducial marks extracted along the profile shape.

In an authentication scheme, the candidate has to claim his identity prior to accessing the system. This can be done by the introduction of a personal identification number or by a personal card that has to be read by the system. The features of the candidate are then compared with the entry associated with the claimed identity. Authentication succeeds if the distance between the candidate feature set and the claimed reference is below a given threshold which may depend of the claimed identity (individual thresholding). This system operates much faster compared to an identification scheme, since only one entry has to be checked. The performance of the system can be evaluated in terms of *False Rejection* (FR) rate, i.e. the percentage of client accesses rejected by the system, and *False Acceptance* (FA) rate, i.e. the percentage of imposter accesses accepted by the system. The *Success Rate* (SR) refers to $1 - FA - FR$. Goudail et al[11] report a SR of 93.5% using local autocorrelation coefficients computed on the facial images of 116 persons. Konen and Schulze-Krüger[12] developed a system based on an extension of the Elastic Graph Matching which runs on frontal images and achieves a SR of 96% on a database of 87 persons. Beumier and Acheroy's profile-based authentication scheme offers a SR of 90% over 41 persons when the profile shape extends about a 500-line resolution.[13]

In the framework of the M2VTS project, a European ACTS project dealing with multimodal person authentication, several algorithms were tested using a common multimodal database of 37 persons.[7] Using this database, an HMM based speech authentication method offered the highest success rate, as far as single modalities are concerned, with a SR of 97.5%.[14] By combining speech information with labial features found in the associated image sequence, the SR then increased up to 99.4%.[14] A Dynamic Grid Matching carried on frontal grey level images of the same database achieved a SR of 89%.[15] By combining these frontal and speech features together, the SR then rose to 99.5%.[15]

These excellent authentication rates of over the 99% are boosted in fact by the high performance offered by the speech modality. Depending on the application, speech data is not always available (person authentication using static mug shots) or cannot be used (person authentication in noisy environments). This paper takes these cases into account and investigates the performance of a system based on facial images only, using both profile and frontal information extracted from the profile and frontal views.

Starting from two different profile-related modalities, one based on the profile shape, the other on the grey level distribution along this shape, we will see how to issue a profile-based expert whose performance is improved compared to each profile modality taken separately. A second expert will use the invariant parts of frontal view (eyes and nose area) in order to issue a frontal-based authentication. Then different fusion schemes will be studied and the best approach will be applied in order to efficiently combine our two experts. This will result in a robust image based authentication scheme that offers a SR of 96.5% under the same conditions as the above-mentioned M2VTS results.

This paper is organized as follows. In Section 2, we present the first profile-related modality which works on the *profile shape* and uses a chamfer matching technique to map the candidate profile to the reference profile. Section 3 introduces the second profile-related modality which is based on a grey level correlation between the candidate and the reference *profile images*, computed inside an area taken along the profile shape. These first modalities will be combined later on (Section 6) into a unique module called the *profile expert*. Section 4 deals with the last modality, which also issues a grey level correlation measure but makes use of a rectangular window located inside the *frontal image*, covering major invariant features like the eyes and nose. This modality will be referenced as the *frontal expert*. Section 5 presents the M2VTS database and the test protocol that has been used during our various experiments. The performance of each modality/expert is given in Section 6, when both global and individual thresholding schemes are used. In Section 7, different fusion strategies are tested and the best one is applied to the fusion of our profile- and frontal-based experts into a unique image-based person authentication algorithm. Finally, Section 8 concludes this work.

2. PROFILE SHAPE MATCHING

The first modality consists of the authentication of the profile outline and is inspired from.[16] The algorithm is based on a chamfer matching that directly works on the profile contour encoded as x-y coordinates. In this way, the performance of this modality mostly depends on the ability of the profile shape to dissociate different faces and not on the choice of a particular set of profile features and/or the quality of their extraction. Before introducing the profile matching method, we will first explain how to extract profile shapes from the color images taken from the M2VTS database.

2.1. Profile Segmentation

All profiles used in this work are taken from the M2VTS database profile views.[7] The database offers a nearly constant lighting over the different shots and a uniform grey background. As the background luminance is very close to the luminance of the skin, we have to use color information in order to extract the profile outline from the image. This extraction is automatically performed in two stages: first, the head is segmented from the background, then the profile is extracted from the head.

The head segmentation is performed by means of color clustering according to the method proposed by.[17]

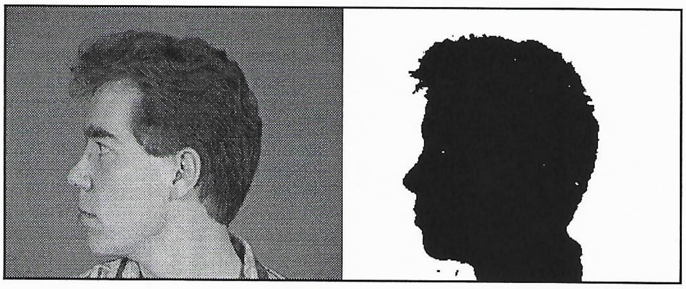

Figure 1. Result of the background segmentation by means of color clustering

Once the head is segmented, the profile must be extracted from the head. In order to achieve good recognition results, we have to restrict this extraction to the invariant parts of the profile only, and :

- exclude the forehead when it can be affected by the hairstyle

- exclude the area below the chin, as its contour highly depends on the tilt of the head which is likely to change from one shot to another
- exclude the lower part of bearded faces.

For each user, it is also important to select the same part of the profile from one shot to the other in order to avoid the introduction of a bias in the residual chamfer distance once the best compensation parameters have been found. As mentioned above, this fixed part extracted from the profile outline, has to take into account the characteristics of the face, excluding features like hair, moustache or beard from the profile. Therefore, at user-definition time, the choice is given between different extraction modes: a first mode selects the full profile and assumes short hair and the absence facial hair (Figure 2). A second mode only selects the lower part of the profile and has to be used when hair is suspected to cover the forehead. A third mode selects the upper area of the face for people wearing a moustache or a beard. The last mode is a combination of the two previous ones and selects the nose area only. All modes are normalized with respect to the nose height as illustrated in Figure 2 (first mode). More information about these different extraction modes can be found in.[16]

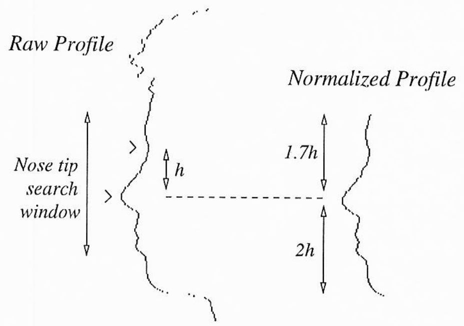

Figure 2. Profile extraction (1st mode)

2.2. Chamfer Profile Shape Matching

The chamfer matching technique searches for the 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* to generate the distance map.[18] By superposing the candidate image on this distance map and by summing up all distances found along the candidate profile, we get an estimate of the global distance that stands between them (Mean Squared criterion).

Actually, we cannot directly compare the reference and the candidate profiles together. The candidate profile has first to be compensated for the possible geometric transformations that can affect it from one shot to the other, i.e. translation along the x and y axes (t_x, t_y), rotation in the x/y plane (θ_{xy}) and scale factor (z). Given a set of values for these transformation parameters, we build a *compensated profile* from the candidate profile. This compensated profile is superposed on the reference distance map and a global distance is computed. The best match between the candidate and the reference profiles is obtained by finding the set of transformation parameters minimizing this global distance. It thus reverts to minimize a cost function (distance) which, in our case, depends on t_x, t_y, θ_{xy} and z. This minimization is done through a classic multidimensional minimization method. We make use of the *Downhill simplex algorithm*, which requires only function evaluations and no derivatives.[19] The chamfer algorithm approximates the Euclidian distance with a maximum error of 6%. Hopefully, as the chamfer approximation is only used as a cost function during the matching process, this error does not influence the quality of the profile shape mapping and let us avoid the computation of the real Euclidian distance.

The global matching process is illustrated in Figure 3. First, the candidate profile is projected onto the reference distance map and a global distance is computed. By minimizing this distance, the optimum compensation parameters are found (t_x, t_y, θ_{xy} and z). Then, the residual distance between the best compensated and the reference profiles is used to decide whether the two profiles belong to the same person or not.

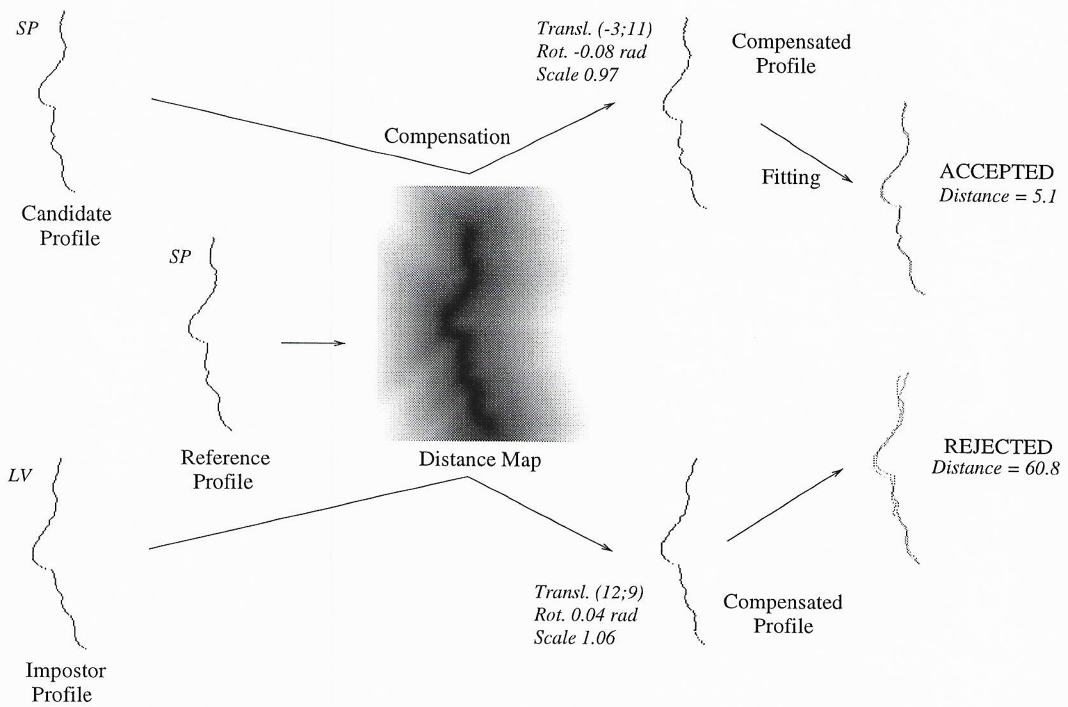

Figure 3. The Chamfer Matching Process

In order to avoid the simplex algorithm to converge towards a local minimum, attention must be paid to the initial parameters values used to initialize the algorithm. These values have to be as close as possible to the final solution for the algorithm to converge efficiently. t_x, t_y are estimated first by comparing the positions of the nose tip between the reference and the candidate profiles, z is found by comparing the two profile heights and θ_{xy} is arbitrary set to zero. These values are used to initialize a first chamfer/simplex algorithm that runs into a low-resolution mode (the candidate profile and the reference distance map are down-sampled by a factor of 4 in both x/y directions). As output, we get refined values for t_x, t_y and z and a pretty good estimation for θ_{xy}. All these values are used to start the final full-resolution search. At the end of the chamfer matching process, the residual distance between the best compensated candidate and its associated reference is obtained and used as a matching score. This score will be further processed in Section 6.

3. GREY LEVEL PROFILE MATCHING

The second profile-related modality is based on 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 comparison, one has to normalize the grey level distribution between the two images to get rid of the illumination variability. Two kinds of normalization have been investigated :

- Dynamic normalization: the grey level distribution of the image is extended to its maximum dynamic (0..255). Unfortunately, this method does not work properly and a problem occurs when bright areas (e.g. when teeth are visible) are present in one image and not in the other (e.g. mouth closed). Assuming the same lighting conditions for the two images, one will already be close to the maximum dynamic (teeth luminance values are close to 255) while the other has to be normalized. This results in an overall brightness change in the second image but not in the first and causes the grey level comparison to fail.

- Mean normalization: the mean of the two images is set to half the maximum dynamic (127) by adding an offset to every pixel. This method has been used in our experiments.

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. The best results are given for an area width of 25 pixels and when the grey level images are low-pass filtered prior to the matching. This low-pass filtering is meant to improve the quality of the matching in particularly noisy areas like the eyebrows. Some results of the grey level matching are shown in Figure 4.

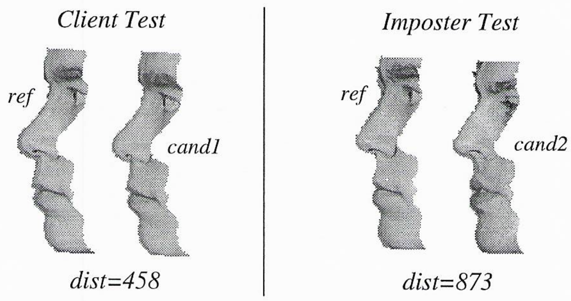

Figure 4. Grey level matching

4. 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, we make use of the same grey level normalization as described above. 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 - which are not low-pass filtered yet - using a technique that is similar to the technique proposed by[8]:

- from the frontal image, we compute the horizontal projection of horizontal gradient which offers a maximal peak at the eyes position and allows to precisely locate the vertical position of the eyes

- then, an horizontal 20 pixels-height strip is centered around the vertical eye coordinate and used to compute the vertical projection of the vertical gradient. This projection offers two distinct peaks at the horizontal locations of the two eyes.

- from these measures, we compute the coordinates of the point located in between the eyes. These coordinates are used to position a fixed-length matching window around the eyes and nose features (110x80 pixels).

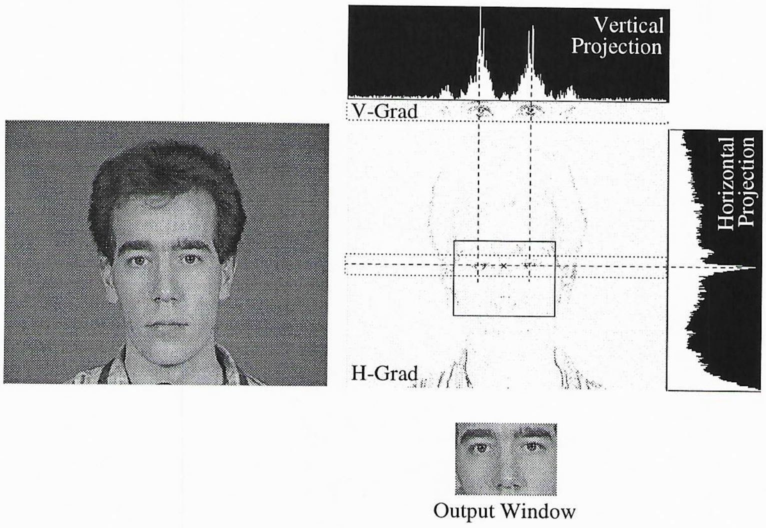

Figure 5. Frontal features localization

This procedure is illustrated in Figure 5.

Unlike the grey level profile modality that makes use of the compensation parameters issued form the chamfer shape matching, we don't know 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 t_x, t_y, θ_{xy} and z parameters, and minimize the grey level distance between the two images. In order to speed-up the algorithm, approximative translation parameters (t_x, t_y) are provided to the first iteration of the simplex algorithm. These parameters are guessed by comparing the position of the point located in between the eyes of the candidate and the reference images, according to the 3-step algorithm described above. As first estimates, $\theta_{xy} = 0$ and $z = 1$. Moreover, different scale factors are applied for the horizontal (z_y) and vertical (z_x) directions, allowing to compensate for a small rotation of the head around the vertical axis. However, this double scale factor also allows to better distort one imposter image onto the claimed reference image and is likely to degrade the performance of the system in terms of false acceptances. Therefore, only small variations between these two scale factors should be allowed. The cost function to be minimized by the simplex algorithm is modified accordingly :

$$MSE(t_x, t_y, \theta, z_x, z_y) + \alpha |z_x - z_y| \tag{1}$$

where α has been manually set to allow a maximal difference of 10% between the two scale factors.

5. TEST SETUP

The M2VTS Multimodal Face Database[7] 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. The profile itself extends over a 100-150 pixel area. Profile views have been manually extracted from the *motion* sequences (see the M2VTS Database terminology) with a tolerance of about 90 degrees $\pm$ 15 degrees. The same tolerance applies for the manual selection of the frontal images. In our experiments, subjects do not wear glasses. Four different shots were used to perform our tests. These shots were taken at one week intervals. Different frontal views taken from the M2VTS database are shown in Figure 6. Figure 7 illustrates the different shots.

Our procedure for experimentation follows the M2VTS protocol.[7] 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 training database is used for providing the *reference* models for each client but also to calibrate the different acceptance thresholds required during the test session. The performance of the identification algorithms is evaluated

Figure 6. M2VTS Database : some frontal views

Figure 7. M2VTS Database : the different shots

by matching the 37 candidate persons (36 clients and 1 imposter) from the test database with the 36 reference clients. Such an experiment provides 36 *authentic* and 36 *imposture* tests. An authentic test consists of candidate claims which are true. An imposture test consists of candidate claims which are false.

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 imposter 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 profile (taken from test database) with each of its claimed references (the 3 shots taken from the training database) and by taking the best score (i.e. the lowest residual distance) as the final score.

6. PERFORMANCE OF THE SINGLE MODALITIES

A given match is accepted if its score (residual chamfer or grey level distance) is below a given threshold. Otherwise, it is rejected. The decision threshold can be applied to the whole database (single decision threshold) or may depend of the claimed identity (individual thresholding). For each threshold - or set of thresholds in the case of individual thresholding - some clients are rejected (i.e. the false rejection $FR(k)$) while a number of imposters are able to enter the system (i.e. the false acceptance $FA(k)$). By varying the decision threshold k, one can select a particular operating point $(FA(k), FR(k))$ depending on the final application. Individual thresholds are acquired during a training session and make use of the training dataset (36 clients/3 shots). During this session, the individual thresholds are computed as follows :

- for each client, take the 35 others to simulate imposter accesses
- take as a threshold for this client, the average score between the 5 best imposter accesses.
- the imposter scores are computed by taking the best score out of all possible matches between the 3 imposter references (taken from the 3 available shots) and the 3 client references.

Instead of taking the best imposter scores, one could have worked with the worst client scores in order to fix the individual thresholds. However, characterizing the behavior of the "best" imposter is more reliable compared to modeling the "worst" client, as we benefit from much more data to process. Therefore, imposter scores have been used here above. Moreover, computing an average over the best scores instead of working with the best score only, allows to better characterize the behavior of dangerous imposter accesses, and results in a better threshold selection.

The individual thresholding scheme described above provides 36 different thresholds and a single (FA, FR) operating point. In order to access other operating points, other individual threshold vectors may be generated by applying different scale factors to the original thresholding vector. Not surprisingly, an individual thresholding achieves a better performance than a global thresholding, as it allows to relax the constraint (i.e. increase the acceptance threshold) for users that are difficult to impost (i.e. users that are characterized by a distinctive feature set).

The shape and grey level profile modalities were combined into a *profile expert*, by summing up their respective scores. In order to equally weight the two modalities, their scores were normalized with respect to their average client score over the training set. This profile expert may be considered as being independent from the frontal grey level modality - hereafter referred to as the *frontal expert* - as it gives access to the depth information of the face, i.e. information which cannot be accessed from a frontal view and thus cannot be part of the extracted frontal features. Thanks to this independence, the fusion of these experts will result in a drastic improvement of the performance of our authentication scheme, as shown in Section 7.

Table 1 summarizes the performance of the different modalities and experts by giving some particular operating points for both global and individual thresholding schemes. The *Equal Error Rate* (EER) stands for the point where $FA = FR$ and the *Success Rate* (SR) refers to the operating point where $1 - FA - FR$ reaches a maximum. The FR is also given for a FA of 1%.

	Thresholding	EER	SR	$FR_{FA=1\%}$
Profile Shape	Global	9%	83.5%	50%
	Individual	8%	87%	23%
Profile Grey level	Global	11%	78.5%	29%
	Individual	9%	82.5%	22.5%
Profile Expert	Global	8%	85.5%	18.5%
	Individual	7%	89%	11%
Frontal Expert	Global	11%	79.5%	25%
	Individual	8.5%	84.5%	17.5%

Table 1. Performance of the different modalities and experts

7. FUSION OF THE DIFFERENT EXPERTS

The fusion of different experts can be performed in two different ways: fusion at *decision* level or at *score* level. In the first case, different decisions are issued from each expert independently and then combined together according to simple rules like AND/OR operators or majority vote. On the other hand, one can also work by first combining the score *values* and then issue the final decision by thresholding the aggregated score.

A fusion scheme based on the scores (*soft fusion*) is likely to outperform a fusion that is based on individual decisions (*hard fusion*), since some information is lost after a thresholding is performed. We thus should work in the score value domain as long as possible during the fusion, then proceed to the final decision as the last step only. However, as long as simple soft fusion techniques are concerned, like a linear combination of the different scores as considered below, hard fusion may sometimes offer a better performance depending on the nature of the data to be merged, as further shown.

Both hard and soft fusion schemes have been applied to the fusion of our two experts according to the test protocol described in Section 5 and using individual thresholds. The fusion at decision level is based on AND and OR operators, while the fusion at score level performs a thresholding on the best linear combination of the two modalities scores.

Results are shown in Figure 8 where the different curves refer to the best performance achievable trough a given fusion scheme (minimum envelope of all possible combination of operating points). Making use of the M2VTS database and for the particular case of the fusion of our profile and frontal experts, the OR hard fusion offers the overall best results. Its performance is summarized in Table 2, and results in a drastic improvement compared to the performance of our individual experts.

	Thresholding	EER	SR	$FR_{FA=1\%}$
OR Fusion Scheme	Individual	2%	96.5%	3%

Table 2. Performance of the best fusion scheme (OR)

8. CONCLUSION

This paper presented a novel multimodal person authentication approach based on static images taken from the frontal and profile facial views. Two reliable independent authentication experts were developed. One is based on frontal features only and offers a Success Rate of 84.5%, the other makes use of information taken from the profile view and achieves slightly better results with a SR of 89%. These figures refer to a intensive testing session (more than 10,000 matches) on the M2VTS database (which does well represent data available from real applications) and make use of a fast individual threshold selection developed in this contribution. Different fusion schemes were studied : two hard fusion schemes using the AND and OR logical operators and a soft fusion scheme based on a linear combination of scores. In the particular case of the two image-based experts developed here, the OR logical operator achieved the best performance, with a SR of 96.5%. This result rates high considering the database that has been used and the fact that no speech information has been made available to our fusion manager.

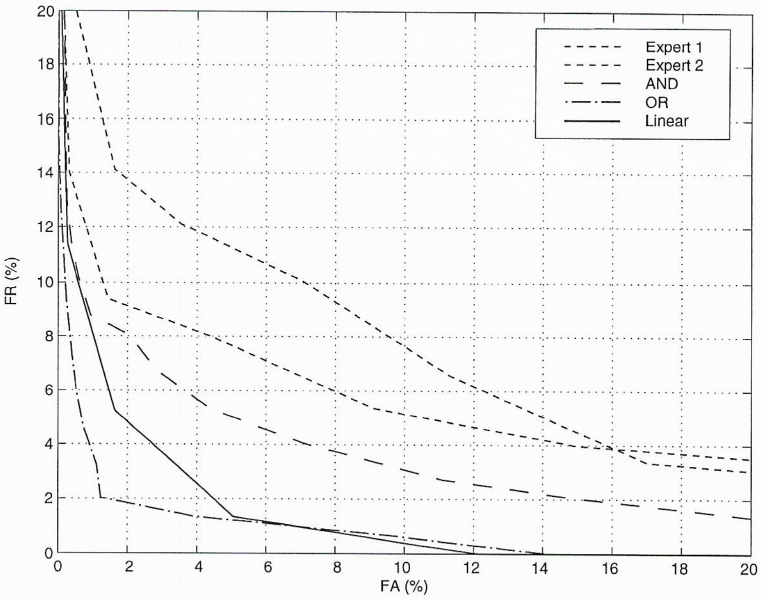

Figure 8. Performance of the hard and soft fusion schemes on the M2VTS Test Database

ACKNOWLEDGEMENTS

This work has been performed in the framework of the M2VTS Project (*Multi Modal Verification for Teleservices and Security applications*) granted by the European ACTS programme.

REFERENCES

1. M. Acheroy, C. Beumier, J. Bigün, G. Chollet, B. Duc, S. Fischer, D. Genoud, P. Lockwood, G. Maitre, S. Pigeon, I. Pitas, K. Sobottka and L. Vandendorpe, "Multi-Modal Person Verification Tools using Speech and Images", *Proc. European Conference on Multimedia Applications, Services and Techniques (ECMAST '96)*, Louvain-La-Neuve, Belgium, May 28-30, 1996, pp. 747-761.
2. R. Brunelli and D. Falavigna, "Person Identification Using Multiple Cues", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 10, Oct. 1995, pp. 955-966.
3. E. S. Bigün, J. Bigün, B. Duc, H. Bigün and S. Fisher, "Expert Conciliation for Multi Modal Person Authentication Systems by Bayesian Statistics", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 291-300.
4. J. Kittler, "Combining Evidence in Multimodal Person Identity Recognition Systems", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 327-334.
5. R. Chellappa, C. L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proc. IEEE*, Vol. 83, No. 5, May 1995, pp. 705-740.
6. P. J. Phillips, H. Moon, P. Rauss and S. A. Rizvi, "The FERET September 1996 Database and Evaluation Procedure", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 395-402.
7. S. Pigeon and L. Vandendorpe, "The M2VTS Multimodal Face Database (Release 1.00)", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14 1997, pp. 403-409. See also *http://www.tele.ucl.ac.be/M2VTS/*
8. 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.
9. B. Moghaddam and A. Pentland, "Face Recognition using View-Based and Modular Eigenspaces", *Automatic Systems for the Identification and Inspection of Humans, Proc. SPIE*, Vol. 2277, July 1994.

10. K. Yu, X. Jiang, H. Bunke, "Face Recognition by Facial Profile Analysis", *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, June 26-28, 1995, pp. 208-213.
11. F. Goudail, E. Lange, T. Iwamoto, K. Kyuma and N. Otsu, "Face Recognition System Using Local Autocorrelations and Multiscale Integration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 10, Oct. 1996, pp. 1024-1028.
12. W. Konen and E. Schulze-Krüger, "ZN-Face: A System for Access Control Using Automated Face Recognition", *Proc. International Workshop on Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, June 26-28, 1995, pp. 18-23.
13. C. Beumier and M. Acheroy, "Automatic Profile Identification", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 145-152.
14. P. Jourlin, J. Luettin, D. Genoud and H. Wassner, "Acoustic-Labial Speaker Verification", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 319-326.
15. B. Duc, G. Maître, S. Fischer and J. Bigün, "Person Authentication by Fusing Face and Speech Information", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 311-318.
16. S. Pigeon and L. Vandendorpe, "Profile Authentication Using a Chamfer Matching Algorithm", *Proc. First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '97)*, Crans-Montana, Switzerland, March 12-14, 1997, pp. 185-192.
17. J. Matas and J. Kittler, "Spatial and Feature Space Clustering: Applications in Image Analysis", *Proc. 6th Int. Conf. on Computer Analysis and Patterns*, Prague, Czechia, September 6-8, 1995.
18. 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.
19. W. H. Press, S. A. Teukolsky and W. T. Vertterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1988.