Secured access to terminals and teleservices using biometrics verification

G. Richard[1]*, Y. Menguy, I. Guis, P. Lockwood*

Matra Nortel Communications

Rue JP Timbaud, 78392 Bois d'Arcy, France.

Abstract

This paper presents a typical application of secured access to a terminal and/or teleservices using biometrics (speech and frontal face) verification. This application is one of the prototype developed in the European ACTS project « M2VTS » which stands for Multi Modal Verification for Teleservices and Security Applications. The primary goal of this project was to address the issue of secured access to local and centralised services in a multimedia environment. The main objective was to extend the scope of application of network-based services by adding novel and intelligent functionalities, enabled by automatic verification systems combining multimodal strategies (secured access based on speech, image or other information). The objectives of the project were also to show that limitations of individual technologies (speaker verification, frontal face authentication, profile identification,...) can be overcome by relying on multi-modal decisions (combination or fusion of these technologies). Some of the Major results of M2VTS are presented in this paper.

1. Introduction

This paper presents a typical application of secured access to a terminal and/or teleservices using biometrics (speech and frontal face) verification. Traditional ways of access control to teleservices or terminals are based on identification (login name) with verification based on password or pin code. The main drawbacks of such an approach is that on the one hand it obliges the user to remember a password and that on the other hand the security vanishes as soon as this password is lost or stolen. Biometrics verification provides a novel and intelligent mean to enhance the security while at the same time enables the user to enter a secured terminal or teleservice without the need of remembering a dedicated password. Moreover, the use of multiple modalities allows to overcome the limitations of individual technologies (speaker verification, frontal face authentication, finger prints,...) and this by relying on multimodal decisions (combination or fusion of these technologies).

The work described in this paper is based on two independent modalities, namely speech (using speaker verification) and image (using frontal face verification).

These two modalities have the advantage of being non invasive. For example technologies such as iris verification are often negatively received by potential users due to the acquisition technique. Furthermore, users usually easily accept to be filmed (in the context of secured access) and to leave a voice signature. On the contrary with finger prints users are more reluctant to leave such information due to their « criminal » aspects but also due to the fact that all users will have to directly touch the fingerprints sensor.

The paper is organised as follows. The next section is dedicated to the description of the European ACTS project « M2VTS » (Multimodal Verification for Teleservices and Security Applications) in which most of the technology has been developed. The third section describes a prototype application developed by Matra Nortel Communications. The last section suggests some conclusions.

2. The European ACTS M2VTS project

2.1 Project overall description

M2VTS (« Multimodal Verification for Teleservices and Security Applications ») was a project supported by the European commission within the ACTS program (project AC-102). The primary goal of the M2VTS project was to address the issue of secured access to local and centralized services in a multimedia environment. The main objective was to extend the scope of application of network-based services by adding novel and intelligent functionalities, enabled by automatic verification systems combining multimodal strategies (secured access based on speech, image or other information). The major problem in user authentication is to achieve on the one hand toll performance: false acceptance rate as low as possible (minimise access to impostors), and false rejection rate as low as possible (a registered user should access to his system in any case), and on the other hand stand the wide range of conditions of use of such systems as well as provide ergonomically viable solutions.

The research was driven by the application needs and user requirements. Therefore, work had essentially been driven by four main goals:

1. Develop platforms for evaluation, implementation and fast prototyping of technology. Submit these platforms to

[1] Corresponding author. Gaël RICHARD, Matra Nortel Communications, Rue JP Timbaud, 78392 Bois d'Arcy, France. Email : gael.richard@matranortel.com

user tests in real situations; in order to measure the adequacy between user requirements and current maturity of the technology.

2. Develop algorithmic solutions for user authentication in a multi-modal context. Implement these solutions on the software platforms for fast prototyping. Refinements of the algorithms based on the results of the user tests in real situations
3. Develop prototypical applications for end users.
4. Perform final test at end users sites.

The project commenced in November 1995 and officially ended in December 1998.

2.2 Participants to the project

In order to achieve the initial goals of the project, a consortium was built around key players in Biometrics technology and Security application fields. Fourteen partners were involved in the M2VTS partners. Led by *Matra Nortel Communications (Fr*)**,** the French n°2 in telecommunications, the consortium included

- two other private companies (***Cerberus A.G.*** *(CH)* and ***Ibermatica*** **S.A.** (*SP))*
- 8 research institutes with extensive experience in the domain of biometrics technology *(Ecole Polytechnique Fédérale de Lausanne (CH), Aristotle University of Thessaloniki (GR), Université Catholique de Louvain (B), University of Surrey (GB), Royal Military Academy (B), the Institute Dalle molle d'Intelligence Articificielle Perceptive (CH), Institute of Microtechnology/University of Neuchâtel (CH) University of Carlos III (SP)).*
- 3 End users *(Compagnie Européenne de Télésécurité (European Telesecurity company), Unidad Tecnica Auxilliar de la Policia (Basque police), Banco Bilbao Vizcaya).*

2.3 Major achievements of the project

The major achievements of this project can be grouped in four categories : the development of innovative multimodal techniques, the recording of large multimodal databases, the development of hardware platforms where the multimodal technologies are integrated and the realisation of several applications for secured access. These achievements are described in [16] and summarized below :

2.3.1 Innovative Multimodal Verification Techniques

A critical survey of the literature related to human and machine face recognition is found in [3], and in [4] for speech recognition/speaker recognition. Within the M2VTS project, even if more emphasis was put on face recognition and speaker verification, other important modalities related to image were studied. In summary, the key techniques developed included:

- Frontal face recognition algorithms with very low error rate (about 7-8% on a database of 295 persons for the best algorithms). Most of these techniques run very efficiently (less than a few seconds on modern processors) (for further information one may consult: [5,6,7]);
- Profile recognition with very low error rate on «ideal conditions» images (7% EER) [8];
- Lip tracking techniques [2];
- Speech verification techniques leading to very low error rates (less than 2% EER). The speaker verification techniques are either Text dependent (i.e. the speaker is asked to pronounced a specific text that can be prompted on screen, or it can also be predefined password) using Hidden Markov Model (HMM) classification technique, or text independent using arithmetic-harmonic sphericity measure or HMM [1,11,12];
- Facial surface analysis by 3D capture and analysis [10].

Furthermore, the primary interest of the project was to take benefit of these multiple modalities by using innovative fusion techniques such as clustering algorithms (split and merge algorithm), Bayesian fusion, Fisher linear discriminant and fusion classifiers [9,14]. In the latter approach, fusion is viewed as a particular classification problem and techniques such as Support Vector Machine (SVM) and Logistic Regression proved to be particularly adapted to the fusion task [14].

The results obtained on the extended M2VTS database for speech and frontal face verification are summarized in Tables 1 and 2.

2.3.2 *Large Multimodal Databases*

In M2VTS, a large database of talking and rotating heads was acquired for the purpose of training and testing multi-modal face and speaker verification systems. In acquiring the database, two hundred and ninety five (295) persons from the University of Surrey visited a recording studio four times at approximately one-month intervals to insure sufficient variability. On each visit (session) two recordings (shots) were made. The first shot consisted of speech whilst in the second shot the subject was asked to rotate his head through a series of set positions (see figure 2, for an example on two subjects). In the third shots, 3D models were extracted using stereo active system developed by the Turing institute.

More information about the database can be obtained on the Web site of the University of Surrey, (*http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb*). This Extended M2VTS database (xM2VTSdb) extends the 37 persons database [13] acquired in the first year of the project. Concurrently, «real conditions» databases are also built during field tests (see figure 3). In particular, situations such as non-uniform lighting, smiling faces, or scale variation are represented. These « real conditions » databases are mandatory to test and improve the robustness of the authentication algorithms in difficult conditions such as non stable or asymmetric lighting, non-uniform background for image[15], and such as background noise for

speaker verification.

2.3.3 Platforms

In order to illustrate the usefulness, flexibility, and potentialities of the multi-modal verification techniques developed in M2VTS two hardware platforms were built.

- The first hardware platform consisted in a powerful and general multimedia platform based on a multiprocessor chip TMS320C80 from Texas Instruments (TI). This platform was actually composed of two very similar independent boards designed by Cerberus AG, namely the VisionPoint80 for the central site, and the CyberGuard80 for the remote site of a surveillance application with remote control.
- The second hardware platform, based on the MVC board designed by Matra Nortel Communications and dedicated to text-dependent speaker verification, was powered by more conventional DSPs (three processors TMS320C50 from TI). This board was designed to be plugged in a PC using the ISA bus. Thanks to a multi-channel telephony interface, it provides access to four PSTN lines of a PBX. Consequently, this second platform (MVC/PC) was usable for various access control applications such as local applications for physical access control and remote applications for access control to teleservices (telebanking, teleshopping or any teleservices application through Internet).

The best multimodal techniques were integrated in the flexible platform and tested in real conditions. These tests have permitted to iterate a back and forth collaboration between industrial and academic experts in order to optimise and to enhance the robustness of the algorithms in real conditions.

2.3.4 Applications

Seven prototypical applications integrating multimodal biometrics verification were developed in M2VTS by the three technically involved industrial partners. These are secured access to local information systems, secured access to a building with or without central monitoring, teleservices application through Internet and on cash dispensers. One of these applications, a secured access to information services is described below.

3. An application of secured access to information services

The prototype developed by Matra Nortel Communications (MNC) integrates an application of secured access to information systems. The secured access is enhanced by using multimodal biometrics verification technology (face and speech verification). The face verification technology is based on the Elastic Graph Matching technology developed by EPFL in the M2VTS project but includes additional pre-processing tools developed by MNC to improve the robustness of the system (an automatic face grabber, a lighting correction tool, and an automatic face re-scaling module).

The speaker verification technology is based on Hidden Markov Models and was developed by Matra Nortel Communications. The fusion scheme used is, at the current stage, rather rudimentary but already permits to enhance the overall robustness of the system compared to each of the complementary modalities.

This application is built around a PC Pentium II 300 MHz with 64 MB RAM (equipped a Matrox Meteor II card for image acquisition and a sound blaster card for voice acquisition), a colour video camera (Sony EVI – D31), a microphone, which is connected to the amplifier of the camera and a Infra-red badge reader and multi-interface decoder (master BB+). A typical setting for this application is depicted below :

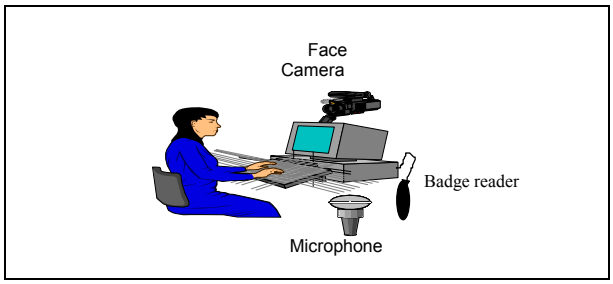


Figure 1: Typical setting for the secured access to information services application

It is important to emphasize that this application has very limited constraints for the user (no use of close talk microphone, the user do not need to be at a predefined distance from the camera). The secured access is achieved through the following steps :

- Identification of the user : the user presents its personal badge to a badge reader
- Speaker Verification : the user is asked to pronounce three numbers (randomly generated by the application) that are prompted on the screen.
- Frontal face verification : the user is asked to place his face inside the capture window. Thanks to automatic face tracking algorithm, the verification is automatically launched when a face is identified.
- Fusion and decision : based on the score obtained by the individual modalities, a decision is taken. As a result, the access to the information service is either granted or refused (if the user is an impostor).

The different steps of this application are illustrated in fig. 4.

4. Conclusion

This paper presented some of the results of the M2VTS project and focused in the last section on one of the prototypical application developed in the project. All applications developed in the project where demonstrated on fixed platforms, however, it is foreseen that in the framework of UMTS that future mobile terminals will integrate enhanced features such as video capabilities. These new capabilities will enable new services and in particular enhanced secured access to teleservices and information systems, where biometric verification could be performed

either directly on the terminal or remotely based on features extracted on the terminal.

5. References

[1] D. Genoud, F. Bimbot, G. Gravier, and G. Chollet. *«Combining methods to improve the phone based speaker verification decision»,* in ICSLP'96*, vol 3, pp 1756-1759, 1996*

[2] J. Luettin and N. A. Thacker*, "Speechreading Using Probabilistic Models",in Computer Vision and Image Understanding, Vol. 65, No. 2, pp. 163-178, 1997.*

[3] R. Chellapa, C.L. Wilson, and S. Sirohey, «Human and machine recognition of faces: A survey», *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995

[4] J. Campbell, «Speaker Recognition: A Tutorial*», proc. Of the IEEE*, Vol 85, No 9, Sept. 1997.

[5] B. Duc, S. Fischer, and J. Bigun, «Face authentication with Gabor information on deformable graphs», *IEEE Trans. on Image Processing*, submitted 1997

[6] C. Kotropoulos, A. Tefas and I. Pitas, « Frontal Face Authentication using Variants of Dynamic Link Matching based on Mathematical Morphology », *IEEE Int. Conf. on Image Proc. (ICIP'98)*, Chicago, USA, I-122-I-126, 4-7 October 1998

[7] J. Matas, K. Jonsson and J. Kittler, «Fast face localisation and verification» *in Proc. of Brit. Machine Vision Conf.,* 1997

[8] Stephane Pigeon and Luc Vandendorpe, "Image-based multimodal face authentication", *in Signal Processing*, Vol 69, no 1, August 1998, pp 59-79.

[9] J. Kittler and M. Hatef and R.P.W Duin and J. Matas, «n combining classifiers» *accepted in IEEE Trans. Pattern Analysis and Machine Intelligence and to be published in 1998*

[10] C. Beumier, M.P. Acheroy, « Automatic Face Authentication from 3D Surface *», British Machine Vision Conf. BMVC 98*, Univ. of Southampton UK, 14-17 Sep, 1998

[11] P. Jourlin, J. Luettin, D. Genoud, and H. Wassner, «Acoustic-Labial Speaker Verification» *in Pattern Rec. Letters,* to appear

[12] B. Duc, G. Maître, S. Fischer, and J. Bigün, «Person Authentication by Fusing Face and Speech Information» *in Proceedings of AVBPA'97,* 1997

[13] S. Pigeon, and L. Vandendorpe, "The M2VTS multimodal face database," *in Proceedings of AVBPA'97,* 1997

[14] S. Ben-Yacoub, "Multi-Modal Data Fusion for Person Authentication using SVM", *IDIAP-RR-98-07.*

[15] A. Tefas, Y. Menguy, C. Kotropoulos, G. Richard, I. Pitas, P. Lockwood, « Compensating for variable recording conditions in frontal face authentication algorithms », *Proc of ICASSP'99.*

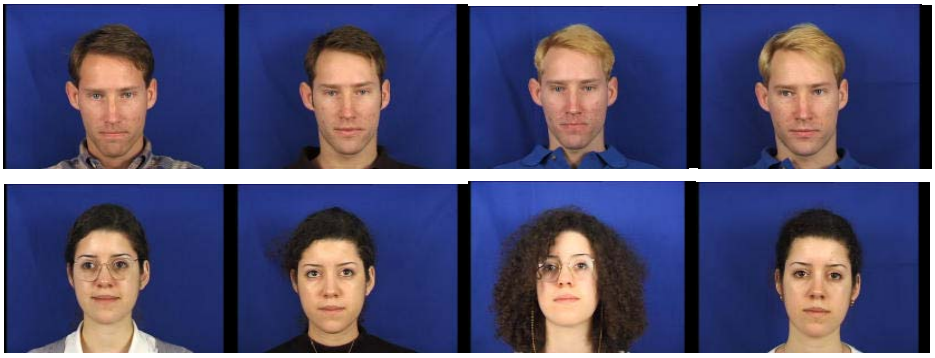
[16]G. Richard & al. « Multi Modal Verification for Teleservices and Security Applications (M2VTS) », *Proc. of IEEE ICMCS99, 7-11 June 1999, Firenze, Italy.*

Algorithm	Evaluation		Test	
	FA	FR	FA	FR
Elastic Graph Matching (EPFL)	8.4%	8.4%	8.1%	8.5%
Sphericity1 (IDIAP1)	1.17%	1.17%	1.6%	5.0%
Sphericity2 (IDIAP2)	0.17%	0.40%	1.41%	7.0%
HMM (IDIAP3)	0.015%	0.0%	1.48%	0.0%
Morphological Dynamic Link (AUT1)	8.11%	8%	8.23%	6.0%
Text independent based on GMM	4%	3%	10%	3%
Optimised Robust Correlation (Surrey)	7.5%	7.5%	7.76%	7.25%

Table 1: Results obtained on the extended M2VTS database (295 persons) for frontal face and speech verification

Fusion algorithm	Experts	Evaluation		Test	
		FA	FR	FA	FR
Bayesian fusion	EPFL, IDIAP1, AUT1	1.1%	1.1%	1.9%	1.0%
SVM	EPFL, IDIAP1	0.59%	0.6%	0.92%	1.5%
SVM	EPFL, IDIAP2	0.8%	0.8%	1.02%	1.75%
SVM	Surrey, IDIAP3	0.0%	0.07%	1.07%	0.25%
SVM	Surrey, IDIAP2, IDIAP3	0.0%	0.0%	0.34%	0.5%
Fuzzy K-means	MDLA-AUT	2.21%	0.5%	2.44%	2.25%
Fuzzy K-means	MDLA	6.04%	6.5%	6.17%	3.5%
Fuzzy K-means	HMM-IDIAP	4.99%	5.0%	4.25%	4.53%
Non-linear SVM	Surrey ORC, IDIAP HMM and Sphericit			0.35%	0.50%
Weighted Linear	Surrey ORC IDIAP HMM	0.6%	0.6%	0.86%	0.25%

Table 2: *Results obtained by different fusion algorithms on the 295 persons database. (FAR= False Acceptance Rate; FRR=False Rejection Rate ;*



Session 1 Session 2 Session 3 Session 4

Figure 2: The Extended M2VTS database (xm2vtsdb: Images of the same two subjects grabbed from the video taken at each of the separate sessions.

Figure 3: A « real condition » M2VTS database (from left to right 1. Normal, 2.Lighting changes, 3. Smiling faces, 4. Scaling)

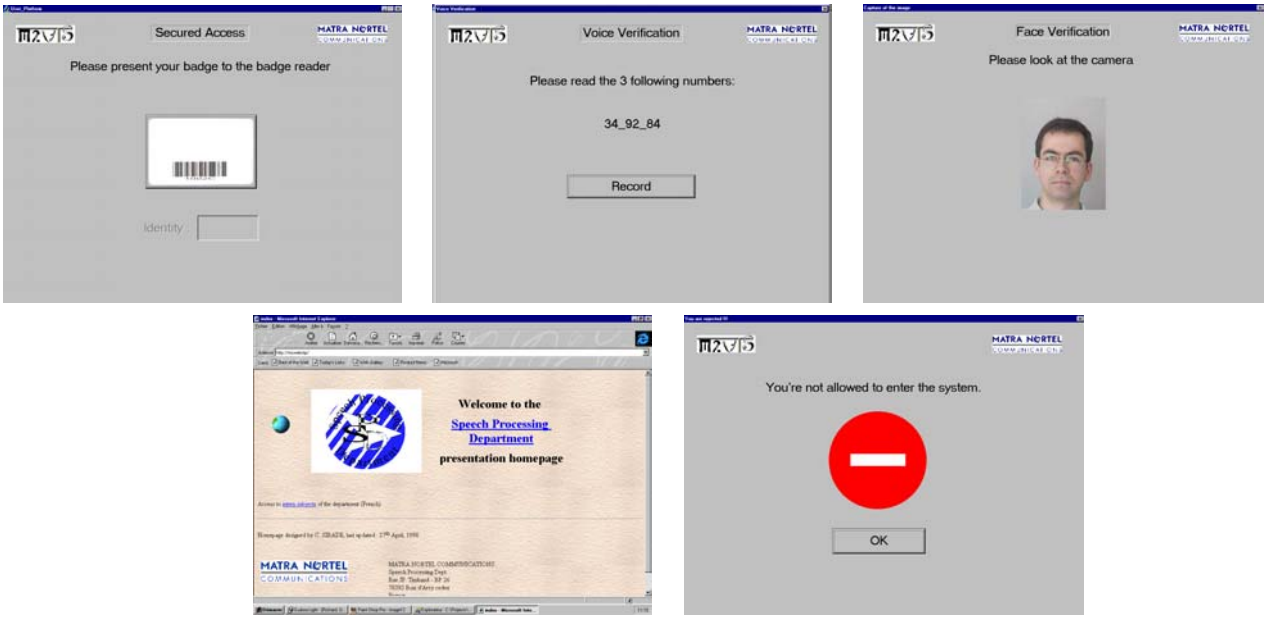


Figure 2: The Matra Nortel Communications prototypical platform: typical operation succession : 1 ; User Identification (upper left) ; 2) speaker verification (upper middle) ; 3) Image verification (upper right) ; 4) Fusion and decision (the user is accepted and accesses to the information service (lower left), or refused (lower right).