Paper proposed for the Special Session on European projects IEEE Multimedia Systems'99

Multi Modal Verification for Teleservices and Security Applications (AC 102 - M2VTS)

G. Richard[1], Y. Menguy, I. Guis, N. Suaudeau, J. Boudy, P. Lockwood
Matra Nortel Communications, Rue JP Timbaud, 78392 Bois d'Arcy, France
C. Fernández, F. Fernández, D. Garcia-Plaza,
Ibermática, Avenida del Partenon, 16-18, Campo de las Naciones, 28042 Madrid, Spain
C. Kotropoulos, A. Tefas, I. Pitas,
Aristotle University of Thessaloniki, PO BOX 451, Thessaloniki 54006, Greece
R. Heimgartner, P. Ryser,
Cerberus, CH 8708 Mannedorf,Switzerland
C. Beumier, P. Verlinde,
Royal Military School, , Belgium
S. Pigeon
Université Catholique de Louvain, Belgium
G. Matas, J. Kittler,
University of Surrey
J. Bigün, Y. Abdeljaoued,
Ecole Polytechnique Fédérale de Lausanne, Switzerland
E. Meurville, L. Besacier, M. Ansorge,
Institute of Microtechnology, Rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland
G. Maitre, J. Luettin, S. Ben-Yacoub
IDIAP, Switzerland
B. Ruiz
CIII, Spain
J. Cortés
Banco Bilbao Vizcaya
K. Aldama
Banco Unidad Tecnica Auxilliar de la Policia

Summary

This paper presents the European ACTS project « M2VTS » which stands *for Multi Modal Verification for Teleservices and Security Applications.* The primary goal of this project is to address the issue of secured access to local and centralised services in a multimedia environment. The main objective is 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 are 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). The paper is organised as follows: in section 1, the main goals of the project are given. Then, in a second section, the list of participants is given and shortly discussed. Section 3 is dedicated to four major achievements of the project. Finally, some conclusions are drawn.

Key words: Multimodal interaction, multimedia databases, Multimedia software engineering tools

[1] Corresponding author. Email : gael.richard@matranortel.com

Multi Modal Verification for Teleservices and Security Applications (M2VTS)[2]

G. Richard[3], Y. Menguy, I. Guis, N. Suaudeau, J. Boudy, P. Lockwood, C. Fernandez, F. Fernández, C. Kotropoulos, H. Tefas, Pitas, R. Heimgartner, P. Ryser, C. Beumier, P. Verlinde, S. Pigeon, G. Matas, J. Kittler, J. Bigün, Y. Abdeljaoued, E. Meurville, L. Besacier, M. Ansorge, G. Maitre, J. Luettin, S. Ben-Yacoub, B. Ruiz, K. Aldama, J. Cortes

1.Objectives of the project

M2VTS (« Multimodal Verification for Teleservices and Security Applications ») is a project supported by the European commission within the ACTS program (project AC-102). The primary goal of the M2VTS project is to address the issue of secured access to local and centralised services in a multimedia environment. The main objective is 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 objectives of the project are also to show that limitations of individual technologies (speaker verification, front face authentication, profile identification,...) can be overcome by relying on multi-modal decisions (combination or fusion of these technologies) and can take benefit of the emerging multimedia environment: workstations, network computers, smart phones (that are more and more equipped with audio and video capabilities). The research is driven by the application needs and user requirements. Therefore, work has essentially been driven by four main goals:

1. Develop platforms for evaluation, implementation and fast prototyping of technology. Submit these platforms to 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 is now in its final year. At the time of writing this abstract, the project is now focusing on field tests of the final applications.

2.Participants to the project

In order to achieve the initial goals of the project, a consortium has been made around key players in Biometrics technology and Security application fields. The M2VTS project is led by **Matra Nortel Communications (Fr),** the French n°2 in telecommunications. Two other private companies are technically involved in the project, **Cerberus A.G.** (CH) on the one hand who commercialises a wide variety of security applications, and **Ibermatica S.A.** (SP) on the other hand who develops multimedia applications with secured access and in particular in the banking sector.

M2VTS is clearly a technology oriented project and is thus built around several research institutes with extensive experience in the domain of biometrics technology**: Ecole Polytechnique Fédérale de Lausanne** (CH) for face recognition and fusion strategies, **Aristotle University of Thessaloniki** (GR) for face recognition, **Université Catholique de Louvain** (B) for profile recognition and fusion strategies, **University of Surrey** (GB) for face recognition, lip tracking and large multimodal database recording, **Royal Military Academy** (B) for 3D facial surface verification, **the Institute Dalle molle d'Intelligence Articificielle Perceptive** (CH) for speaker verification, lip tracking and fusion decision strategies, **Institute of Microtechnology/University of Neuchâtel** (CH) for speech algorithms optimisation and software integration onto hardware platforms, **University of Carlos III** (SP) for speaker verification.

[2] M2VTS is a European project of the ACTS program (project AC 102). Support for the Swiss partners is provided by OFES 95.0201. More information on the project can be found on the official web site : http://www.tele.ucl.ac.be/M2VTS.

[3] Corresponding author. Email : gael.richard@matranortel.com

Finally, to guarantee that the project leads towards applications that fulfill the need of potential end users, three other companies are involved in the project whose role is to participate to the application specifications and to test the final applications prototypes. Those companies are **Compagnie Européenne de Télésécurité** ***(European Telesecurity company)*****, Unidad Tecnica Auxilliar de la Policia** (*Basque police*) and the **Banco Bilbao Vizcaya**.

3.Major achievements of the project

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

3.1. Innovative Multimodal Verification Techniques

Speech and face recognition have exhibited a tremendous growth for more than two decades. A critical survey of the literature related to human and machine face recognition is found in [4], and in [2,5] 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 include:

- Frontal face recognition algorithms with very low error rate (4.9% to 7% Equal Error Rate (EER) 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: [6,7,8,9]);
- Profile recognition with very low error rate on «ideal conditions» images (7% EER) [10];
- Lip tracking techniques [3][11];
- Speech verification techniques leading to very low error rates (less than 3% 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) using Hidden Markov Model (HMM) classification technique, or text independent using arithmetic-harmonic sphericity measure [2,14,15,19];
- Facial surface analysis by 3D capture and analysis [13].

Furthermore, the primary interest of the project is to take benefit of these multiple modalities to build biometric systems for security application with several order of magnitude accuracy improvements. As such several innovative fusion techniques were developed such as clustering algorithms (split and merge algorithm), Bayesian fusion, Fisher linear discriminant and fusion classifiers. 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 [17]. On the first M2VTS database containing 37 persons [16], the best results obtained using three modalities are given below: the merit of each approach developed in the project is currently assessed on the extended M2VTS multimodal database of 295 persons (described below).

Modality	**TE**	**FAR/FRR**
Face	11.0%	3.6% / 7.4%
Profile	7 %	2% / 9%
Speech	6.7%	6.7% / 0.0%
Fusion (SVM)	**0.2%**	**0.2% / 0.0%**
Fusion (RBF-SVM	**0.1%**	**0.1% / 0.0%**
Fusion (Bayesian)		**0.5% / 0.0%**

Table 1: *Results obtained by individual modalities and by different fusion algorithms on a 37 persons database (see [17]). (FAR= False Acceptance Rate; FRR=False Rejection Rate; TE=Total Error ; SVM= Support Vector Machine ; RBF-SVM= radial basis function SVM) (see also [7][10][12]).*

3.2 Large Multimodal Databases

In this project a large database of talking and rotating heads was acquired for the purpose of training and testing multi-modal face and speech 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 below for an example on two subjects). This Extended M2VTS database (xM2VTSdb) extends the 37 persons database [16] acquired in the first year of the project and it is our belief that this database is the first of its kind and that it should encounter a clear success from the scientific and business communities. Concurrently, «real conditions» databases are also built during field tests. In particular, situations such as non-uniform lighting, smiling faces, or scale are represented.

3.3 Platforms

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

The first hardware platform consists in a powerful and general multimedia platform based on a multiprocessor chip TMS320C80 from Texas Instruments (TI). This site through ISDN. From the application point of view, the platform supports the implementation of a broad range of application software, including sustained speech (text-independent speaker verification) and image processing, and videoconferencing standard H.320. The hardware being designed for real-time speech and image processing, a multi-modal verification can be processed in a few seconds. One of the application developed in the project is based on this hardware platform (Secured access with central monitoring/alarm verification (see section 3.5)) and is currently under field test.

The second hardware platform, based on the MVC board designed by Matra Nortel Communications and dedicated to text-dependent speaker verification, is powered by more conventional DSPs (three processors TMS320C50 from TI). This board is 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) is usable for various access control applications:

-Local applications for physical access control.

-Remote applications for access control to teleservices such as telebanking, teleshopping or any teleservices application through Internet.

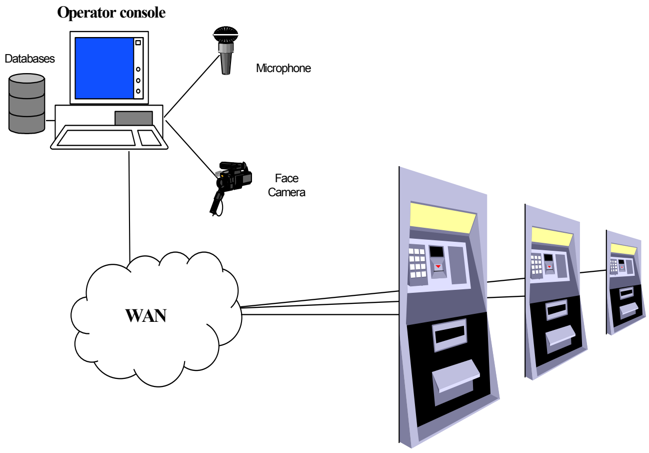
In an earlier stage of the project, this flexible software/hardware platform was used to develop several demonstrators that were placed at end user sites and evaluated using background technology. The results of these first field tests showed that in general the systems were not well accepted due to the lack of reliability and especially in difficult conditions (variable lighting, scale variation, background noise,...). This typically demonstrated that there was a clear need for new and more robust techniques for biometrics person authentication.

At the current stage of the project, the best multimodal techniques have been integrated in the flexible platform and tested in real conditions. These tests have already 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 and interesting results have already been produced for asymmetric lighting [18]. From this prototyping platform, several applications have been developed (see next section) and since the algorithms implemented in the applications already achieved far better performance than the background technology, there is no doubt that the acceptance by users will be significantly higher. Field tests at end user sites are now conducted and should confirm that expectation.

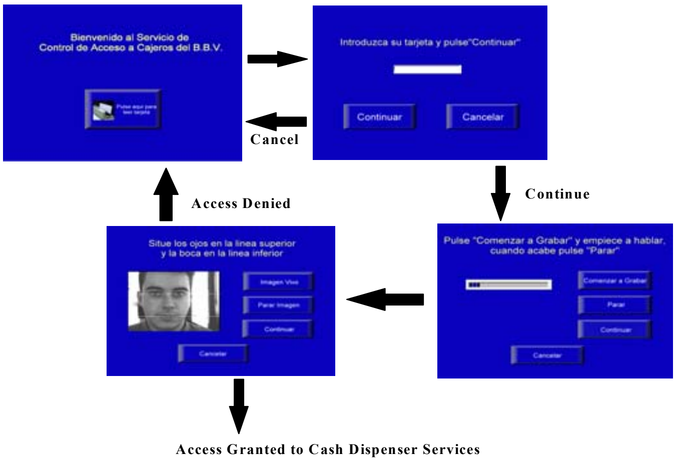
3.4. Applications

Seven prototypical applications integrating multimodal biometrics verification have been developed in this project by the three technically involved industrial partners. Most of these applications have been built using an Application Generation Tool developed in the project. Thanks to this tool, the project managed to develop several applications based on audio-visual verification. 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. Due to the limited available space for this abstract, only the cash dispenser teleservices application is described below :

A M2VTS application: cash dispenser Teleservices application*:* This application controls and verifies the identity of a person who accesses the different teleservices provided by a Bank (*Banco Bilbao Vizcaya)* through its cash dispenser net. The system is composed of a camera, that takes the frontal image of the face of the user and of a microphone or recording system of audio that permits the use of the voice, in addition to a magnetic cards reader that provides the system with the identification of the user. The user authentication and access to the bank services is performed in the cash dispensers. The dispenser will be based in a Pentium PC equipped with camera and microphone.



The multimodal biometrics system runs on the cash dispenser with the following user interface integrated in typical screens driven dialogue. The sequence of screens is as follows (see picture below): the first screen shows a welcome message and a button to begin the authentication process. The second screen asks the user to insert his card. When the card is read, the system evaluates its validity and allows to continue or to stop at this point. The third screen is the *audio verification screen*: the user presses the first button, then starts talking. The fourth and last screen is the *face authentication*: the user is asked to place his face between two horizontal lines and when ready to press the second button. The system processes all the information and allows the access to the cash dispenser services or denied it and the process re-starts from the beginning.



Bienvenido al Servicio de
Control de Acceso a Cajeros del B.B.V.

Introduzca su tarjeta y pulse"Continuar"

Pulse "Comenzar a Grabar" y empiece a hablar,
cuando acabe pulse "Parar"

Situe los ojos en la linea superior
y la boca en la linea inferior

4.Conclusion

The M2VTS project is nearly completed and has, in many respects, reached its original objectives. The multimodal verification techniques developed in the project are clearly on the leading front of the field and have led to numerous high quality publications. The recording of a large multimodal database (295 persons, 4 different sessions per person) also represents a clear impact of this project on the scientific community. Despite the strong commitment in technology development, the project led to several prototype applications, some of them being built around hardware elaborated in the project. These applications are currently tested on End users premises to validate the technology in real conditions. It is however probable that these tests will show that the applications perform well in «controlled conditions» but are not robust to environmental variations such as on the one hand, lighting changes, moving background, scale or translation variation (for image verification) and on the other hand, background noise and distant sound recording (for speaker verification). Future work in audiovisual verification should then be dedicated to robustness to environment variable conditions.

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] D. Reynolds, «Speaker Identification and Verification using Gaussian mixture speaker model», *Speech Com*. Vol 17, pp 91-108 (1995)

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

[4] 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

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

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

[7] C. Kotropoulos, and I. Pitas, «Face authentication based on morphological grid matching», *in Proc. of the IEEE Int. Conf. on Image Processing (ICIP 97)*, pp. I-105--I-108, Santa Barbara, California, U.S.A., 1997

[8] 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

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

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

[11] M. U. Ramos Sanchez and J. Matas and J. Kittler, «Statistical chromaticity-based lip tracking with B-splines*», proc. of IEEE-ICASSP97*, pp 2973-2976, Vol 4., 1997

[12] 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*

[13] 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

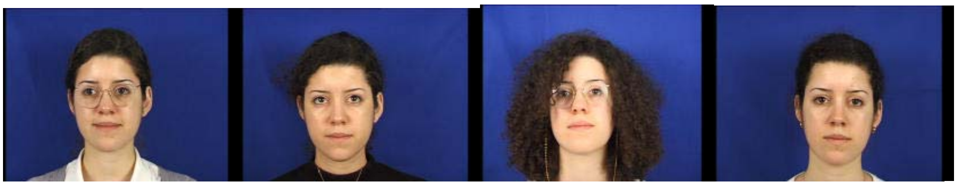
[14] P. Jourlin, J. Luettin, D. Genoud, and H. Wassner, «Acoustic-Labial Speaker Verification*» in Pattern Recognition Letters,* to appear

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

[16] S. Pigeon, and L. Vandendorpe, "The M2VTS multimodal face database," *in Lecture Notes in Computer Science: Audio- and Video- based Biometric Person Authentication (J. Bigun, G. Chollet and G.Borgefors, Eds.)*, vol. 1206, pp. 403-409, 1997

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

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



Session 1 Session 2 Session 3 Session 4

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