Collaboration and the Knowledge Economy: Issues, Applications, Case Studies
Paul Cunningham and Miriam Cunningham (Eds)
IOS Press, 2008 Amsterdam
ISBN 978–1–58603–924-0

Multimodal Biometric Authentication in Secure Environments

Savvas ARGYROPOULOS, Yannis DAMOUSIS, Dimitrios TZOVARAS
Informatics and Telematics Institute, 1st km Thermi-Panorama Rd.,
P.O. Box 60361 Thermi-Thessaloniki, GR-57001, Greece
Tel: +30 2310 464160, Fax: + 30 2310 464164, Email: {Savvas.Argyropoulos, Damousis, Dimitrios.Tzovaras}@iti.gr

Abstract: This paper presents the multimodal biometric authentication framework of the EU Specific Target Research Project (STREP) called HUMABIO (Human Monitoring and Authentication using Biodynamic Indicators and Behavioural Analysis). The project aims to develop a modular, robust, multimodal biometrics security authentication system, which improves state-of-the-art methods in biometrics, such as face, speech, and gait and increases unobtrusiveness. For each modality, the data acquisition, feature extraction and matching procedures are briefly discussed. Subsequently, a comparative evaluation of multimodal score-level fusion methods ranging from simple classification schemes to more sophisticated machine learning algorithms, such as support vector machines and fuzzy expert systems is presented and their advantages over unimodal classification are highlighted. Experimental evaluation on data recorded in adverse environmental conditions shows that despite the increased unobtrusiveness the multimodal authentication system can achieve very satisfactory performance.

1. Introduction

Human identification has always been a field of primary concern in applications such as access control in secure infrastructures. In contrast to passwords or tokens which can be easily lost, stolen, forgotten, or shared, biometrics offer a reliable solution to the problem of identity management. Especially, the development of systems that integrate two or more biometric traits has received increased interest during the last years as the advantages of multimodal biometric systems become more evident. Most of the limitations imposed by unimodal biometric systems, e.g., low accuracy, high failure-to-enrol rate, sensitivity to noise etc., can be overcome in multimodal biometric systems [1].

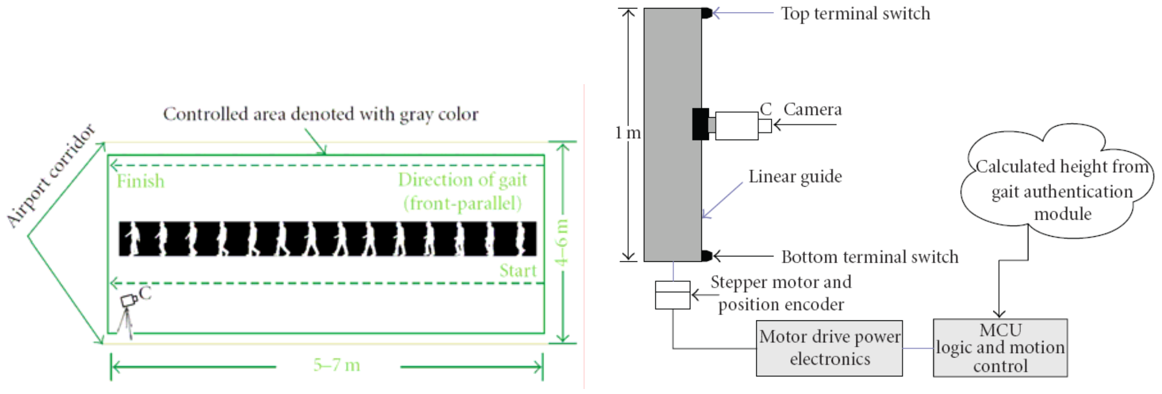
A major shortcoming of current biometric systems is the obtrusive process for obtaining the biometric feature. The subject has to stop, go through a specific measurement procedure, which depends on the biometric that can be very obtrusive, wait for a period of time, and get clearance after authentication is positive. Emerging biometrics such as gait and technologies such as automated person/face detection can potentially allow the non-stop (on-the-move) authentication or even identification which is unobtrusive and transparent to the subject and become part of an ambient intelligence environment.

This article describes the multimodal biometric authentication system which was implemented during the course of the Human Monitoring and Authentication using Biodynamic Indicators and Behavioural Analysis (HUMABIO) FP6 EU project [2]. In particular, the application scenario for non-stop and unobtrusive authentication of employees in a controlled area is examined. Authentication is based on gait, voice, and face modalities in order to limit the cooperation of the user as much as possible, increase unobtrusiveness and user convenience and maximize user acceptance.

2. Objectives

HUMABIO is a Specific Targeted Research Project (STREP) that focuses its research on emerging and novel biometrics, aiming to enhance unobtrusiveness of biometrics-based access control systems. HUMABIO's application scenarios aim at increased unobtrusiveness for the subject, by taking into account varying factors and allow flexibility in the system operation. Such examples are the inclusion of noise models during the development of the voice recognition module for robust operation in noisy environments and the operation of the face module even with various facial expressions. However, increased unobtrusiveness has its toll on authentication accuracy. Even the more conventional HUMABIO biometrics present lower accuracy than the corresponding algorithms in the literature which refer to strictly controlled conditions. Multiple biometrics are combined within HUMABIO with the objective to increase the authentication accuracy of the multimodal system with respect to the biometrics it comprises. Based on criteria such as unobtrusiveness level, maturity of the technology, and biometric capacity, face, voice, and gait recognition biometrics were selected to be included in the airport application scenario of the HUMABIO system. It should be also noted that the authentication instead of identification scenario was targeted due to increased time constraints and requirements.

Unobtrusive authentication involves automatic authentication of authorized personnel that can move freely in restricted areas. The operational setup of the system, which is installed in a controlled area in Euroairport in Basel, Switzerland, is depicted in Figure 1 (a). The subject walks along a narrow corridor. When the subject enters the corridor the (claimed) identity is transmitted wirelessly to the system via radio frequency identification (RFID) tag. The aim of HUMABIO is to authenticate the claimed identity by the time the subject reaches the end of the corridor. As the subject walks through the corridor, the gait sequence is captured and the subject's height is estimated. Height information is used to calibrate the position of the camera and the microphone, as shown in Figure 1 (b). Face and voice recognition take place at the end of the corridor. By the time the subject reaches the camera and the microphone, their position is already calibrated allowing the unobtrusive face and voice recognition without the need of specific procedures for the collection of the biometric data as it is usually the case with current biometric solutions.



(a) (b)

Figure 1: (a) HUMABIO Airport Pilot and (b) Calibration of Camera and Microphone.

3. Description of Biometric Modalities

a. Face Classifier

Face feature extraction is carried out in three steps: face detection, face normalization and subspace projection. The first step for facial feature extraction is the accurate localization of face in an input image using a component based approach using detectors similar to those proposed in [3]. Subsequently, the detected face is normalized by applying a similarity transform (rotation, translation and scaling) to the image region containing the face [4]. In the final stage, the input is regarded as a *N*-dimensional pixel vector containing the concatenated rows of the normalized face image. The feature space for representing faces is then computed by Bayesian subspace analysis, presented in [5]. The dimensionalities for the subspaces were determined beforehand on a large training database containing faces of different individuals exhibiting various facial variations. Thus, the face classifier is robust to facial expressions, poses, illumination conditions and occlusions (e.g. glasses).

b. Speech Classifier

The speech signal is considered as a sequence of short-term frames (about 10 ms) that are processed by a Mel cepstral analysis method [6]; The cepstral transformation decorrelates the subband energies producing a low dimensional feature vector. Since the silence does not contain any speaker discriminant information, silence is withdrawn using an energy based voice activity detector. During the enrolment procedure, the subject is asked to pronounce a set of utterances covering at best the range of phonemes and speaking styles that could be used while being authenticated.

Typically, the amount of speech recorded during the enrolment process could be from 30 seconds to several minutes. Those utterances are then used in order to create the voice profile of the subject. The voice profile is created using statistical modelling of the sequence of feature parameters which aims at estimating a probability that the sentence has been pronounced by a given speaker. Particularly, due to their simplicity, Gaussian Mixture Models (GMM) were used to approximate the true probability density functions. A GMM-based speaker profile is fully depicted by the mean and standard deviation vectors of each Gaussian (on the order of 128) and the Gaussian weights. The focus of the module development has been put on practical side issues such as the robustness to environment noise, the rejection of unreliable speech samples, the limited amount of enrolment data, and so forth. Several noise models were added to examine the robustness of the system in conditions that simulate real application environments.

c. Gait Classifier

The first step in the gait recognition module is the extraction of the walking subject's silhouette from the input image sequence. Subsequently, a set of transforms is applied to the silhouette to represent meaningful shape characteristics. In particular, the Radial Integration Transform (RIT) and Circular Integration (CIT) transform are applied due their capability to represent significant shape characteristics and their robustness to noise (e.g., illumination, different clothing, etc.). Additionally, the use of a new set of orthogonal moments is used based on the discrete classical weighted Krawtchouk polynomials due to their highly discriminative power. Thus, the gait recognition module output three matching scores, one for each descriptor: RIT, CIT, and Krawtchouk. Since the description the gait recognition module lies out of the scope of this paper, the interested reader are referred to [7] for further details.

4. Multimodal Biometric Fusion

Fusion at matching score level is the most common approach in multimodal biometric systems due to the easy accessibility and availability of the matching scores in most biometric modules [1]. In our approach, the output scores of the individual matching algorithms constitute the components of a multidimensional vector from face, gait, and voice classifiers, as illustrated in Figure 2. In the following, we briefly review two advanced machine learning techniques, support vector machines (SVM) and fuzzy expert systems (FES), which are widely used due to their reliability, performance, and effectiveness.

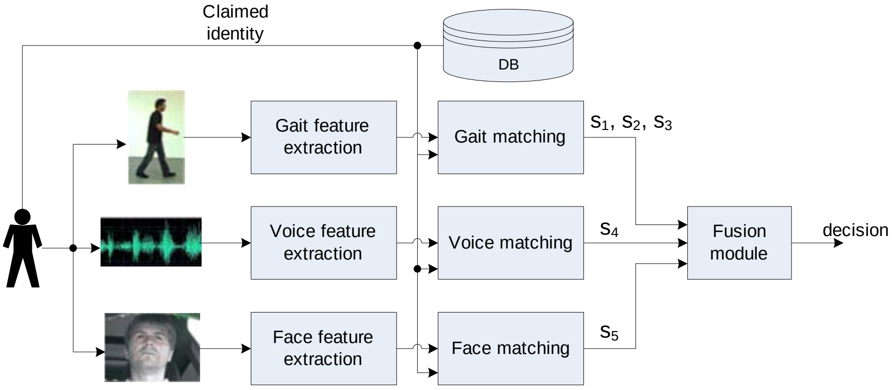


Figure 2: Multimodal Biometric Fusion at the Matching Score Level

a. Support Vector Machines

SVM map a given set of binary labeled training data to a high-dimensional feature space and separate the two classes of data with a maximum margin hyperplane [8], [9]. Thus, an initial fusion function $f\left(\overline{x}\right) = \left\langle w, \varphi\left(\overline{x}\right)\right\rangle + w_0$ is trained by solving the quadratic problem:

$$\min_{w,w_0,\xi_1,\ldots,\xi_N} \left(\frac{1}{2}\|w\|^2 + \sum_{i=1}^{N} C_i \cdot \xi_i \right)$$

$$\textit{s}\text{ubject to:}\quad y_i\left(\left\langle w, \varphi\left(\overline{x}\right)\right\rangle + w_0\right) \geq 1 - \xi_i,\ i = 1, \ldots, N$$

$$\xi_i \geq 0,\ i = 1, \ldots, N$$

where $\overline{x} = \left[s_1, s_2, s_3, s_4, s_5\right]$ and the function $\varphi: \Re^5 \rightarrow F$ maps the data into a feature space *F*. Thus, the training vectors are mapped into a higher dimensional space by the function *φ*. Then, the SVM algorithm finds a linear separating hyperplane with the maximal margin in this kernel dimensional space. Also, ξ_i denote the slack variables which are misclassified, C_i is the cost weight (or penalty parameter) associated with training data $\overline{x}_i$, and y_i represents the label of $\overline{x}_i$. It is easy to prove that the margin is maximized when [8]:

$$w = \sum_i a_i y_i \varphi(\overline{x}_i) \tag{1}$$

where α_i are positive real numbers that maximize:

$$\sum_i a_i - \sum_{ij} a_i a_j y_i y_j \left\langle \varphi(\overline{x}_i), \varphi(\overline{x}_j) \right\rangle \tag{2}$$

subject to:

$$\sum_i a_i y_i = 0,\quad a_i > 0 \tag{3}$$

The decision function can equivalently be expressed as:

$$f\left(\overline{x}\right) = \sum_i a_i y_i \left\langle \varphi(\overline{x}_i), \varphi(\overline{x}) \right\rangle - w_0 \tag{4}$$

It is important to note that neither the learning algorithm nor the decision function need to represent explicitly the image of points in the feature space F, since both use only the dot products $\left\langle \varphi(\overline{x}_i), \varphi(\overline{x}_j) \right\rangle$. Hence, given the kernel function $K(\mathbf{X}, \mathbf{Y}) = \left\langle \varphi(\mathbf{X}), \varphi(\mathbf{Y}) \right\rangle$ one could learn and use the maximum margin hyperplane in the feature space without explicitly performing the mapping. In our case, the radial basis function (RBF) kernel was employed, which is given by $K(\mathbf{X},\mathbf{Y}) = e^{-\gamma\|\mathbf{X}-\mathbf{Y}\|^2}, \quad \gamma \geq 0$.

b. Fuzzy Expert Systems

Fuzzy expert systems (FES) use soft linguistic variables and a continuous range of truth-values in the interval [0, 1] [10]. In order to construct the fuzzy model structure, a number of premise inputs $\bar{X}_p = [x_{p,1}, \ldots, x_{p,NPI}]$ should be properly selected. These are the decision variables that constitute the premise space and allow the formulation of rules. Each premise variable is then partitioned by a certain number of fuzzy sets that cover adequately its universe of discourse. These fuzzy sets allow the linguistic description of a variable. The linguistic description (fuzzy sets) of the inputs is attained using membership functions of appropriate form.

In our model, each premise input $x_{p,i}$, $i=1,2,3$ is partitioned by three Trapezoid type membership functions. The linguistic description (partitioning) of the premise inputs results to the formation of several fuzzy regions $\mathbf{A}^{(j)}$, formed by the combinations of the memberships along each input. This leads to a number of *NR* = 3 x 3 x 3 = 27 rules, as depicted in *Figure 3*.

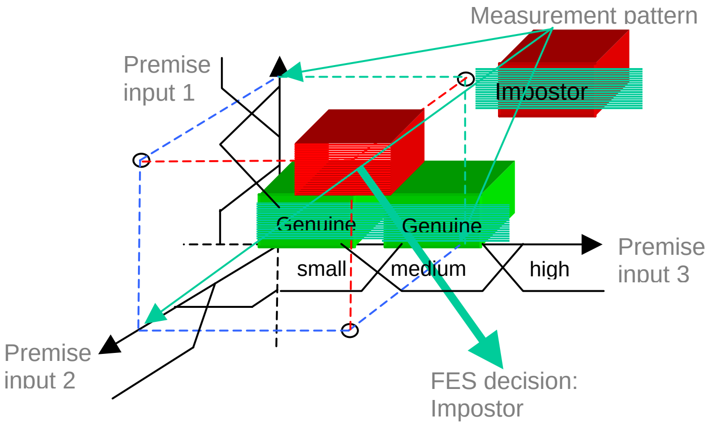


Figure 3: Three membership functions with linguistic expressions "small", "medium", "high" are used for the partitioning of three premise inputs, leading to the formation of 27 fuzzy rules.

Each rule $R^{(j)}$, *j*=1,..., *NR* corresponds to a particular category or case, having an IF-THEN description:

$$R^{(j)}: \quad IF \quad \bar{X}_p \quad is \quad A^{(j)} \quad THEN \quad y_j = F_j\left(\bar{X}_c\right) \tag{5}$$

where y_j = $\mathrm{F}(\bar{X}_c)$ represents the *j*-th rule output which is a crisp function of the input vector $\bar{X}_c = [s_1, s_2, s_3, s_4, s_5]$ consisting of the face, voice and gait matching scores:

$$y_j = F(\bar{X}_c) = \lambda_0^j + \sum_{i=1}^{NCI} \lambda_i^j x_{c,i} \tag{6}$$

where λ_i^j are weight coefficients and λ_0^j is a bias term. The lambda coefficients are determined from the training set using genetic algorithms.

5. Results

The multimodal biometric database for the training and evaluation of the authentication algorithms was formed by aggregating unimodal databases forming “virtual subjects” [11]. The face database consists of 29 subjects captures in two different sessions: one with neutral expression and one with facial expression (smiling or talking). The voice database contains 40 subjects from the YOHO database and consists of “combination lock” phrases (e.g. 36-24-36). Finally, the gait database consists of 75 people recorded at two different conditions: one with a slight difference in appearance (e.g., wearing a hat) and one with a different type of shoe. The *N*-th multimodal virtual user was created using the *N-th* user trait from each database. Thus, the multimodal database consists of 29 subjects and two recordings. The evaluation was performed using the first 15 subjects for training and the remaining 14 for testing. The sets slide for each run by one subject. Thus, the total number of genuine and impostor transactions in the training set is 15 x 2 x 29 = 870 and 15 x 14 x 2 x 29 = 12180, respectively. The test set contains 14 x 2 x 29 = 812 genuine and 14 x 13 x 2 x 29 = 10556 impostor transactions.

The performance of the biometric system is evaluated in terms of the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). Additionally, the performance of the unimodal classifiers is shown in Table 1 which illustrates their Equal Error Rate (EER). It must be stressed that the relatively high error rates compared to corresponding results in the literature are attributed to the flexible conditions during the measurements to increase unobtrusiveness. Furthermore, Table 2 summarizes the results of the investigated machine learning algorithms for multimodal fusion. More specifically, classification was performed using the SVM, FES, Gaussian Mixture Modelling (GMM) [12], and Neural Network (NN) fusion schemes. The first conclusion we can reach from the results illustrated in the table is that all the fusion schemes perform better that the best performing unimodal expert (voice classifier) and the gain is approximately 1.4%. This corroborates the statement that the effective combination of information from different experts can improve significantly the performance of a biometric system. Moreover, this table also confirms the superiority of the SVM fusion scheme. More specifically, the FAR and FRR error rates using the SVM fusion classifier are approximately 0.40% and 0.37%, respectively.

Table 1: EER of the Unimodal Biometric Classifiers

Biometric trait	Face	Voice	Gait	EER (%)
EER	12.8	1.76	18.88	12.8

Table 2: Evaluation Results for Initial Authentication in the Airport Pilot Using Different Fusion Methods

Fusion method	Training		Testing	
	FAR (%)	FRR (%)	FAR (%)	FRR (%)
SVM	0.16	0.11	0.40	0.37
FES	0.28	0.22	0.76	0.72
NN	0.65	0.71	0.92	0.98
GMM	0.95	1.03	1.05	1.11

6. Conclusions

In this paper, the multimodal biometric authentication framework of the HUMABIO project was presented. Specifically, the application scenario for non-stop and unobtrusive authentication of employees in a controlled area based on face, voice, and gait modalities was examined. The main challenge was to address the use of biometrics in specific pilot plans which allow flexible conditions and user unobtrusiveness while at the same time imposing stringent performance requirements. The critical point in such applications is the effective exploitation of the various unimodal experts and their integration in order to provide a global assessment about the person's identity authenticity and physiological state.

The HUMABIO taskforce identified a set of assessment types for the detailed evaluation of the system. These assessment types include the technical and performance assessment (i.e., whether the system can operate as designed), the impact assessment (i.e., to what extent the workplace and societal safety, the operational cost and efficiency, and the user comfort and Quality of Life (QoL) will be affected by the systems' introduction into the market), and the user acceptance assessment (i.e., whether the user groups involved may benefit from the system).

Moreover, a large number of participants were used to evaluate the HUMABIO system under various evaluation scenarios. The performance of both the users and the system were tracked through objective measurements (data from physiological, behavioural, and other biometrics, as well as from data from the systems' performance) and subjective measurement tools (i.e., questionnaires from both the users and the industrial clients). All the users participating in this project had to complete a series of evaluation scenarios, in order to evaluate the effectiveness and efficiency of HUMABIO operation modes. Subsequently, the users' objective and subjective performance was analysed to allow us to draw some conclusions regarding the success and efficacy of the project and its future potential.

The most considerable impact of HUMABIO relates to the innovations that were developed in technology and biometrics. That is, in order for HUMABIO to be able to create a unique physiological signature for each individual, it had to explore the use of novel physiological indicators combined with state of the art behavioural measurements. Such a laborious task has never been attempted before, while, to-date, no systematic study using extensive physiological measurement databases has been presented. Additionally, the development of the HUMABIO system will lead towards a new way of authentication (one that resembles DNA authentication) that will "intimidate" intruders and will be unobtrusive to its users.

The next most important impact of the HUMABIO system can be identified in the realm of security and public safety. HUMABIO attempts to accomplish the combination of physiological and behavioural indicators that will allow de facto aliveness checks to the security/authentication system. Furthermore, this system will provide industry the ability to reduce (if not to eliminate) possible identity fraud and industrial espionage. The creation of such an innovative system will set new safety standards for a variety of application environments (e.g., laboratory, airport, etc.) and at the same time reduce the violation of the user's privacy (by using obstructive security checks). Consequently, the most probable candidates for the system's installation in this case would be environments with high security requirements (such as government agencies, R&D facilities, defence industry, etc.).

Additionally, HUMABIO's impact is not limited to industry, but can also extent to education and research. That is, this system can also offer significant advancements in fields related to human biology and psychology, health, technology, and biometrics. The HUMABIO system will offer researchers working in different modalities, the possibility to work on a modality fusion setting, which will allow an increased reliability in authentication algorithms. The creation of multimodal biometric databases and standardised evaluation methodologies will promote the quality of research conducted by scientists and will allow them to accelerate their research findings and quality.

Regarding exploitation and marketing, it should be noted that multimodal biometrics take only a 2.7% of the global market share in 2007 leaving a lot of room for expansion. In addition the unimodal biometrics developed within HUMABIO can be marketed as commercial products on their own. Several biometric modules are based on emerging biometrics such as EEG and ECG, while others such as anthropometric biometrics based on sensing seats are completely new but can be applied in a variety of systems such as vehicle security. For HUMABIO separate modules and as a whole a detailed bussiness plan has been produced and there is already interest from industries such as Volvo and Siemens for integration to existing security systems that they develop. The main actors identified as potential users of the HUMABIO system are final users (high security environments, supervisors/employees in government agencies, etc.), industrial users (application facilities, such as airports, laboratories, etc.), authentication systems and modules developers, authorities (e.g., police, security officers, etc.) and society (e.g., individuals, medical community, etc.).

References

[1] K. Jain and A. Ross, "Multibiometric systems," *Communication of the ACM*, vol. 47, no. 1, pp. 34-40, Jan. 2004.

[2] I. G. Damousis, D. Tzovaras, and E. Bekiaris, "Unobtrusive Multimodal Biometric Authentication: The HUMABIO Project Concept," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 265767, 11 pages, 2008.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[4] S. Z. Li, "Face detection," in Handbook of Face Recognition, J. A. K. Li, Stan Z., Ed. Berlin, Germany: Springer-Verlag, 2004.

[5] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, no. 11, pp. 1771–1782, 2000.

[6] F. Bimbot, et al., "A tutorial on text-independent speaker verification," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 4, pp. 430–451, 2004.

[7] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas, "Gait recognition using compact feature extraction transforms and depth information," *IEEE Trans. on Information Forensics and Security*, vol. 2, no. 3, pp. 623–630, Sep. 2007.

[8] N. Christianini, and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods," Cambridge University Press, 2000.

[9] S. Hearst et al., "Trends and controversies - support vector machines," *IEEE Intelligent Systems*, vol. 13, no. 4, pp. 18-28, Apr. 1998.

[10] H. J. Zimmermann, "Fuzzy Set Theory and its Applications," Kluwer, Boston, USA, 1996.

[11] N. Poh and S. Bengio, "Using Chimeric Users to Construct Fusion Classifiers in Biometric Authentication Tasks: An Investigation," ICASSP, pp. 1077-1080, Toulouse, France, 2006.

[12] Y. Stylianou *et al.*, "GMM-Based Multimodal Biometric Verification," eINTERFACE'05 Summer Workshop on Multimodal Interfaces, 2005.