Multimodality in BioSecure : Towards an evaluation protocol on virtual multi-modal databases

Bernadette Dorizzi, Sonia Garcia-Salicetti, Mohamed Anouar Mellakh
Département Electronique et Physique
Institut National des Télécommunications
9 Rue Charles Fourier, 91011 Evry France
{Mohamed.Anouar_mellakh@int-evry.fr, Sonia.Salicetti, Bernadette.Dorizzi}@int-evry.fr}

Abstract :
In this paper, we briefly present the BioSecure Network of Excellence and its objectives in terms of the development of biometric evaluation platforms. A particular focus is given in this project to the multimodal case, the evaluation of which requires a special attention due to the lack of large-size available databases. We show in this paper that the evaluation of score fusion methods (for two independent modalities) is possible even on standard size (roughly 100 persons) virtual databases at the price of a careful statistical protocol.

1. The BioSecure Network of Excellence: Biometrics for Secure Authentication[1]

Ensuring the safety of the citizens and society is a major concern nowadays. Moreover, the increasing use of information sensitive applications, such as e-commerce, e-banking and health monitoring, has triggered a real need for reliable, user-friendly, and widely acceptable control mechanisms for checking the identity of an individual. Biometrics, which bases the person authentication on the intrinsic aspects of a human being, appears as a viable alternative to more traditional approaches (such as PIN codes or passwords). European biometric research has already attained a high level of expertise. However, the present situation confirms that the penetration of biometrics has been less enthusiastic than predicted, especially in Europe. In order to overcome these current limitations, it would be necessary to balance the understanding of developments in the technology with the potential applications that will make use of them.

The NoE BioSecure has been started in June 2004, in the domain of biometrics, grouping the critical mass of expertise required to promote Europe as a leading force in the field.

The main objective of this network is to strengthen and to integrate multidisciplinary research efforts in order to investigate biometrics-based identity authentication methods, for the purpose of meeting the trust and security requirements in our progressing digital information society. This goal will be attained through various integrating efforts. A common evaluation framework (such as databases, reference systems and assessment protocols) will be developed, participating to standardisation efforts. Identifying and addressing the technical challenges linked to applications will lead to the definition of joint research activities, aiming at the facilitation of the employability and practical use of the technology.The BioSecure Network of Excellence will also promote mobility and international training. A large place will be given to dissemination through large scale events (i.e. conferences, common evaluation campaigns and residential workshops). These efforts will bring together the community and will facilitate the technology transfer to the industry.

A particular effort will be made in the multimodality domain which remains an open issue due in particular to the lack of large scale available databases. A virtual database obtained through the concatenation of several available databases such as BANCA[1,2], BIOMET[3], MYCT[4] will be used, in a first step, to test different strategies. This database will contain roughly 100 persons and particular care must be given to the associated testing protocols, taking into account the small size of the Database.

2. Interest of multimodality

Multimodality, specially fusioning scores coming from two independent modalities has already been the subject of an intensive research [5,6]. From a practical point of view, the use of several modalities can be considered in order to :

- Improve the efficiency of the global system

A single modality biometric system can be subject to a certain number of defects leading to an expected or unexpected high level of errors. Some errors can be due to *some noise* associated with the sensed data. It may be introduced in such data in many different ways: by sensors, by ambient conditions, or by the user. A high level of errors can also be generated by *Intra-class variability*: Biometric data may be variable from one acquisition to another (depending for instance on the emotional state of the person). Also, some modalities do not have a high enough discrimination capability across individuals: a biometric trait is in fact expected to be *differential across clients*, i.e. it has to vary significantly from one person to another. Some modalities do indeed permit the identification of a person (fingerprints, iris), while others are less enough discriminant for that task. .Finally, biometric systems may be attacked with forged data, or genuine data of a dead person may be presented to the sensor.

[1] http://www.biosecure.info

Using several **different modalities together** should help to deal with the points mentioned above, mostly when using complementary biometrics such as behavioral and physical, discriminative or not etc.. Indeed, multimodality has a clear impact on performance: research works have shown that multimodal systems enhance the authentication systems'performance significantly, relatively to unimodal systems. Such systems have by construction a higher discrimination capability and are more difficult to attack by impostors. Indeed combining fingerprint with hand shape, or face recognition may circumvent the usage of fake fingerprints, as faces and hands are more difficult to imitate than fingers. This is also the case for voice and lip movements which are naturally correlated.

- **Provide a more flexible system**

Considering two (or more) modalities does not mean using them at the same time. Indeed if we build a biometric system relying on both fingerprint and face and if a person cannot enroll its fingerprint, because of the bad quality of his finger, then it will be possible to use only his face image for verification. Non-availability of a biometric trait can also be temporary. Imagine a system functioning with iris and fingerprints. If one person during a short period has a problem with his eye, so that it is impossible to perform the iris scan, the fingerprint system can be used instead. The same thing occurs with people which would refuse to use a specific modality (for religious or health purposes for instance). So the multimodal aspect of the system allows a flexibility by providing an alternative to the identification process.

3. A Generic protocol for multimodal evaluation on virtual and real subjects

The evaluation of a multibiometric system is not an easy task : indeed, there are very few available multimodal databases (XM2VTS [7,8], BANCA [1,2], DAVID [9]]), most of which contain only two biometric modalities, usually face and voice. Also, multimodal databases available nowadays contain only about a hundred subjects, which makes difficult to extrapolate the success of a multimodal algorithm or method when being tested on a large population (thousands or millions of people). Moreover, multimodal databases more recently constructed as BIOMET[3], or under construction [10] have the tendency to contain more modalities (4 or 5) but not more subjects.

Many works in the multimodal fusion literature give results on about 100 real subjects, with no insight in the fact that such results may be in fact very biased. We address this problem in the present work and propose a new protocol for multibiometric systems evaluation on standard size databases of real subjects.

Moreover, it is also natural to wonder about the possibility of using databases of virtual subjects, that is an individual generated by combining different biometric traits (modalities) that belonging to different persons. If valid, this procedure would simplify multimodal data construction because it would be sufficient to merge two or more databases of approximately the same number of subjects, containing each different modalities, to generate a multimodal data corpus containing more modalities. Although this question is crucial for the progress of research in multimodal fusion, few works have exploited up to now the creation of virtual subjects for multimodal fusion [10,11]. As already mentioned, the first question that arises is: which is the validity of this procedure ? Then the next question is: if it is valid, which methodology should be used to evaluate multimodal systems on a given corpus of virtual subjects ? Our aim in this work is also to answer to such crucial questions.

In the following we resume, an experimental work that we performed previously and that is described in more details in [12]

Our methodology has been to create virtual subjects with data coming from a multimodal database of real subjects, that is the BIOMET database [3]. This permits us to do a comparative study of the behaviour of a bimodal fusion system (on-line signature and voice) on the real subjects and on several databases of virtual subjects generated from BIOMET. Indeed, the originality of this work is that we set the problem of using virtual subjects for systems evaluation relatively to the use of real subjects in multimodal databases. In fact, this procedure permits to have a better insight into what is in fact a real subjects database relatively to a virtual subjects one, and how evaluation should be performed in both cases.

As mentioned above, our work is limited to two modalities, voice and on-line signature, already combined in a previous work [13]. Of course, the choice of the modalities is a delicate question since it rises the problem of their mutual dependence/independence. We focus here in the combination of modalities that are a priori mutually independent, since it is only in this framework that we may consider building a virtual subject.

We combine such two modalities by a Support Vector Machine classifier, a statistical technique that has proven to give good results [13] and does not necessitate a priori scores normalisation at the price of a learning phase. We show in this framework that a bimodal (voice, signature) database of real subjects of standard size (around 100 persons) introduces a bias when evaluating the fusion system, because the size of the database does not permit to represent all the possible data variability in the bimodal sense. Moreover, we show that using databases of virtual subjects is equivalent in certain conditions (with a given protocol) to the use of a database of real subjects of standard size. We provide here an evaluation protocol on both types of databases.

3.1 Fusion of On-line Signature and Voice

This study considers two mono-modal biometric systems: a

signature verification system described in [14] and a text-independent Speaker Verification one described in [13]. The scores provided by each system are combined by means of a Support Vector Machine (SVM) [15].
We build what we call a bimodal database through the association of the scores of the two experts (Signature and Voice).
This bimodal database of 77 persons is then split in 2 subsets: one of 39 persons devoted to training the Support Vector Classifier, named *FLB* (Fusion Learning Base), and the other of 38 persons for testing purposes, named *FTB* (Fusion Test Base).
In order to reduce the bias related to the small number of persons in the database, we consider 50 different couples of training and test databases *(FLB,FTB)*, selected randomly, and compute average Errors Rates on the 50 generated *FTBs*.
For each person in FLB and FTB, we have at disposal 5 bimodal client accesses and in average 10 bimodal impostor accesses (this number varies across persons from 6 to 12 impostor accesses).
We create a virtual subject by pairing randomly signature data of a given subject to the speech data of another subject. In theory, we can create this way up to C_k^m data sets of virtual subjects, where k is the total number of clients in the database, and m is the number of modalities; in our case, m=2 and k=77, leading to 2962 data sets of virtual subjects. We chose to create 1000 data sets of virtual subjects as in [11], but in fact the question of the necessary number of data sets of virtual subjects is studied in more details in [12].
Every database of virtual subjects is split into a Fusion Learning Base (FLB) and a Fusion Test Base (FTB) as described above for the real database. We compute the mean False Acceptance Rate FA and the mean False Rejection Rate FR for the 1000 databases of virtual subjects, to obtain a "Virtual Mean DET Curve" (VMDC).

3.2 Comparative Fusion Experiences on real and virtual subjects

As a first step, we perform experiments in order to compare the performance on a real database (BIOMET) and on virtual databases built from the same persons. In Figure 1, we compare the DET curve [16] obtained on the BIOMET database to the 1000 DET curves corresponding to the 1000 databases of virtual subjects. Let's recall that the first curve represents average error rates over 50 different couples *(FLB,FTB)*. Figure 1 shows that the average DET curve on the BIOMET database is inside the band generated by the 1000 DET curves corresponding to virtual subjects sets. This first result permits to conclude that the system behaves on the database of real subjects (when averaging error rates on 50 partitions of the Fusion Learning and Test databases) as on any of the databases of virtual subjects. This also supports the mutual independence assumption between the two modalities that we consider, on-line signature and voice. Moreover, the use of virtual subjects data sets permits to have an estimation of performance variability, providing in fact a "confidence interval" for performance obtained on a real subjects data set of standard size (100 persons). In other words, the database of real subjects is a data set with an inherent bias. This bias is greatly increased if a single partition in a Fusion Learning and Testing Databases *(FLB,FTB)* is considered like widely done in the literature. Indeed, the statistics of bimodal data found in the test set (represented by the real subjects present in such set) may be very different from that present in the training set, leading this way to an unreliable and misleading evaluation of the fusion system. It is thus necessary to generate different couples *(FLB,FTB)* that correspond to different distributions of individuals in *FLB* and *FTB* respectively, and to average error rates over those trials.
The next experience confirms that 50 partitions or couples *(FLB,FTB)* are enough to reduce the inherent bias of the real subjects data set.

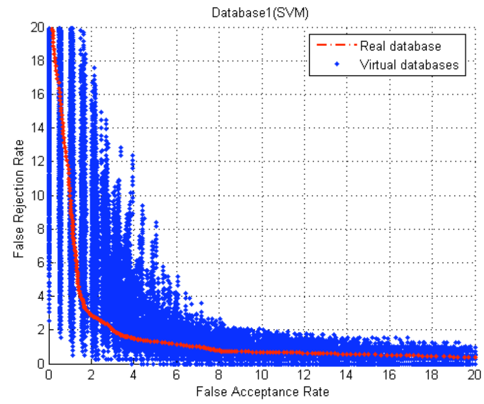

Fig. 1. DET curve for the real Database and 1000 associated virtual Databases

Indeed, we now compare, in a second step, the Virtual Mean DET Curve (VMDC) of the 1000 databases of virtual subjects with the mean DET curve on the BIOMET database. In Figure 2, we notice that the curves have exactly the same behaviour. This shows that it is in fact equivalent to evaluate the fusion system on 1000 virtual data sets to evaluating the fusion system on the database of real subjects by averaging results over 50 partitions *(FLB,FTB)* of such database.

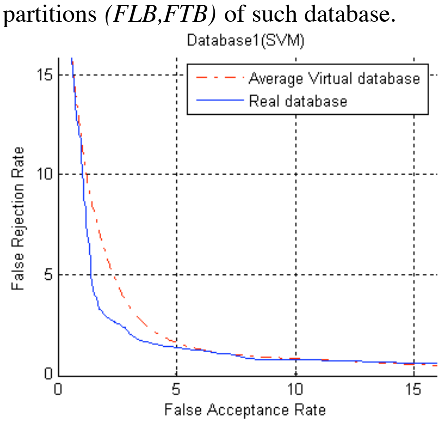

Fig. 2. VDMC vs. average Error Rates on the real database

4. Conclusions

We have studied in this work the problem of evaluation of score fusion algorithms on relatively small size real-person databases of bimodal score values as well as the question of using virtual persons (built through different pairing of the mono-modal scores) instead of real ones. The data at our disposal comes from 77 subjects of the BIOMET database and we considered two a priori independent modalities : on-line signature and speech..Several databases of virtual subjects were constructed from BIOMET bimodal data. Our first conclusion is that a standard size database (about 100 subjects) of real subjects behaves exactly as a virtual subjects set of the same size when evaluating the multibiometric system. This of course supports the mutual independence assumption of the two biometric traits that we consider. In other words, this confirms a natural intuition that a database of real subjects has an inherent bias, since each subject represents a specific combination of the modalities considered, and about 100 instances are not enough to cover all the possible variance of such combination, not even for two modalities. To cope with this fact, we propose a protocol for multibiometric systems evaluation on standard size databases (about 100 subjects) of real subjects, consisting in creating several partitions (we have shown that 50 partitions is a good compromise) of the data set in a Fusion Learning Base and a Fusion Test Base *(FLB,FTB)* and in averaging error rates over such 50 trials for each value of the threshold. Indeed, our statement is that evaluating a fusion system on only one partition *(FLB,FTB)* like usually done in the literature, gives biased and thus unreliable results, even if the subjects that are in the database are real! Moreover, we have shown that it is equivalent to evaluate a fusion system on the database of real subjects by averaging error rates over 50 partitions *(FLB,FTB)*, and on 1000 virtual subjects data sets if a mean False Acceptance Rate and a mean False Rejection Rate are computed on the 1000 data sets for each value of the decision threshold. As a conclusion, we have also proposed a protocol for evaluating a multibiometric system on virtual subjects data sets. Finally, we can conclude that, in the case of mutual independence of the modalities that are considered, the use of virtual subjects with the protocol above given is a powerful tool to estimate the performance variability, providing a “confidence interval” for performance obtained on a real subjects data set of standard size (100 persons). It is thus recommended for a complete and reliable evaluation of multibiometric systems, such as the ones that will occur in the framework of the BioSecure Network of Excellence.

Acknowledgments :
The authors would like to thank G. Chollet (ENST) and his team for having provided the scores of the speech verification system.

References

[1] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, Jo. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruíz, and J.-P. Thiran. The BANCA Database and Evaluation Protocol. Audio- and Video-Based Biometric Person Authentication (AVBPA), Guilford, 2003.
[2] http://www.ee.surrey.ac.uk/Research/VSSP/banca/
[3] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. Leroux-Les Jardins, J. Lunter, Y. Ni, D. Petrovska-Delacretaz, "BIOMET: a Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities", Proc. of 4th International Conference on Audio and Video-Based Biometric Person Authentication, pp. 845-853, Guildford, UK, July 2003
[4] J. Ortega-Garcia et al. MCYT : A multimodal biometric database. In Proc. Of the COST-275 Workshop : The Advent of Biometrics on the Internet, p.123-126, 2002
[5] A. K. Jain & A. Ross, Multibiometric Systems, Communications of the ACM, January 2004/Vol. 47, N°1
[6] J. Kittler, M. Hatef, R. Duin and J. Matas, “On combining classifiers, IEE Trans. On Pattern Analysis and Machine Intelligence 20, 3 (Mar. 1998), IEE, NY, 226-239
[7] K Messer J Matas J Kittler, J Luettin and G Maître, “XM2VTSDB: The Extended M2VTS Database”, Second International Conference on Audio and Video-based Biometric Person Authentication, 1999.
[8] http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/results/
[9] J.S.D Mason, F. Deravi, C. Chibelushi & S. Gandon BT DAVID - Final Report, Speech and Image PDETessing Research Group, Dept. Of Electrical and Electronic Engineering, University of Wales Swansea, UK.
[10] A. Ross, A. Jain, "Information Fusion in Biometrics", Pattern Recognition Letters 24, pp. 2115-2125, 2003.
[11] M. Indovina, U. Uludag, R. Snelick, A. Mink, A. Jain, « Multimodal BiometricAuthentication Methods : A COTS Approach », Workshop on Multimodal User Authentification (MMUA), pp. 99-106, Santa Barbara, California, USA, Dec. 2003.
[12] S.Garcia-Salicetti, M.A.Mellakh, B. Dorizzi, “A generic protocol for multibiometric systems evaluation on virtual and real subjects”, submitted to AVBPA 2005
[13] B. Ly Van, R. Blouet, S. Renouard, S. Garcia-Salicetti, B. Dorizzi, G. Chollet : «Signature with text-dependent and text-independent speech for robust identityverification », Workshop on Multimodal User Authentification (MMUA), pp. 13-18, Santa Barbara, California, USA, Dec. 2003.
[14] B. Ly Van, S. Garcia-Salicetti, B. Dorizzi, “Fusion of HMM’s Likelihood and Viterbi Path for On-line Signature Verification”, Biometric Authentication Workshop (BioAW), Lecture Notes in Computer Science (LNCS) 3087, pp. 318-331, Prague, Czech Republic, May 2004.
[15] V. Vapnik, “The Nature of Statistical Learning Theory”, *Statistics for Engineering and Information Science,* Second Edition, Springer, 1999.
[16] J.P. Egan, "Signal Detection Theory and DET analysis", Academic Press, New York, 1975