A Comparative Evaluation of Fusion Strategies for Multimodal Biometric Verification

J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero and J. Gonzalez-Rodriguez

Biometrics Research Lab., ATVS
Universidad Politecnica de Madrid, Spain
{jfierrez,jortega,jgonzalez}@diac.upm.es   www.atvs.diac.upm.es

Abstract. The aim of this paper, regarding multimodal biometric verification, is twofold: on the one hand, some score fusion strategies reported in the literature are reviewed and, on the other hand, we compare experimentally a selection of them using as monomodal baseline experts: *i*) our face verification system based on a global face appearance representation scheme, *ii*) our minutiae-based fingerprint verification system, and *iii*) our on-line signature verification system based on HMM modeling of temporal functions, on the MCYT multimodal database. A new strategy is also proposed and discussed in order to generate a multimodal combined score by means of Support Vector Machine (SVM) classifiers from which user-independent and user-dependent fusion schemes are derived and evaluated.

1 Introduction

Automatic extraction of identity cues from personal traits (e.g., fingerprints, speech, and face images) has given raise to a particular branch of pattern recognition (*biometrics*) where the goal is to infer identity of people from biometric data [1]. The increasing interest on biometrics is related to the important number of applications where a correct assessment of identity is a crucial point.

Our efforts at Biometrics Research Lab. (Universidad Politecnica de Madrid, Spain), have been focused on three basic biometric characteristics, namely, on-line signature –which is a behavioral trait-, face and fingerprint –which are physiological ones-, due to the following reasons: *i*) regarding fingerprint, due to its uniqueness and high discriminative capability; *ii*) regarding face, for its direct visualness with respect to *in situ* human interaction; and *iii*) regarding signature, for its personal, social and legal acceptability as an identification procedure.

Some studies [2] have showed that the performance of any single-trait verification system can be improved by *unimodal* (or *monomodal*) *fusion*, i.e., the combination of several verification strategies applied on the same input data. Even greater verification performance improvement can be expected through the use of multiple biometric characteristics assuming statistical independence between them [3]. Some works related to the *multimodal fusion* approach are [3]-[6].

In this contribution, and after reviewing some referenced approaches to score fusion in multimodal biometrics, especially those based on SVM classifiers as they

J. Kittler and M.S. Nixon (Eds.): AVBPA 2003, LNCS 2688, pp. 830-837, 2003.

have shown outstanding performance [12], we will derive fusion schemes, both user-independent and user-dependent, based on the computation of a multimodal combined score. Thereafter, fusion strategies will be compared using our signature, face and fingerprint systems on MCYT database [7].

In *verification* or *authentication* (the problem addressed here) a claim is made concerning the identity of a person and the biometric system has to take the binary decision of accepting or rejecting it based on the information extracted from the considered biometric trait regarding a predetermined threshold. In a verification context, two situations of error are possible: an impostor is accepted (false acceptance, FA) or the correct user is rejected (false rejection, FR). Performance measures of verification systems are related to the frequency with which these situations of error happen. One common performance measure is, for example, the so-called EER (*equal error rate*) which is the point attained when FA and FR rates coincide. Here, the performance of competing systems based on different fusion strategies will be compared by means of DET plots [8], which are graphical representations of FA vs. FR rates for a wide range of decision thresholds with a particular axis scaling.

2 Multimodal Fusion

2.1 Fusion Strategies

Biometric multimodality can be studied as a *classifier combination* problem [2], [9]. Kittler *et al.* considered in [9] the task of combining classifiers in a probabilistic bayesian framework and provided an example of multimodal biometric verification (fusing speech, frontal and profile images modalities). Several ways to merge the modalities are obtained (sum, product, max, min, …), based on the Bayes theorem and certain hypothesis, from which the *Sum Rule* (i.e., the combined score is obtained adding the monomodal scores which have been previously mapped to the [0,1] range) outperformed the remainder in the experimental comparison due to its robustness to errors made by the individual classifiers. From now on, this perspective will be referred to as *rule-based fusion*, because it does not takes into account the actual distribution of outputs from the experts.

Multimodal fusion can also be treated as a *pattern classification* problem [10]. Under this point of view, the scores given by individual expert modalities are considered as input patterns to be labeled as accepted/rejected (for the verification task). Verlinde *et al.* followed this approach and compared in [11] the following pattern classification techniques for multimodal fusion (sorted by relative decreasing performance): Logistic Regression, Maximum a Posteriori, *k*-Nearest Neighbours classifiers, Multilayer Perceptrons, Binary Decision Trees, Maximum Likelihood, Quadratic classifiers and Linear classifiers. In a recent contribution [12], the paradigm of *Support Vector Machines* (SVMs) has been compared with all the above-mentioned techniques carrying out the same experiments, outperforming all of them. From now on, this perspective will be referred to as *learning-based* (or *trained*) *fusion,* because it requires sample outputs from the experts to train the pattern classifiers.

Although it could be thought that learning-based fusion should have better performance than rule-based fusion, some examples have been reported in the literature where the Sum Rule have outperformed other learning-based approaches [3]. This rather surprising result motivates the experiments carried out from which we will show that an adequate design of the best reported learning-based fusion strategy (based on SVM) outperforms the Sum Rule approach.

2.2 Multimodal Fusion via SVM

We have used the SVM in order to provide not a binary verification decision, as it has been reported in related works [11][12], but rather a merged score combining the outputs of the considered monomodal experts. We will now introduce our approach providing references for further details.

The principle of SVM relies on a linear separation in a high dimension feature space where the data have been previously mapped, in order to take into account the eventual non-linearities of the problem [13]. In order to achieve a good level of generalization capability, the margin between the separator hyperplane and the data is maximized. Formally, the training set $X = (\mathbf{x}_i)_{i=1}^{l} \subset \mathbb{R}^R$, where l is the number of training vectors and R is the number of modalities, is labeled with two-class targets $(y_i)_{i=1}^{l}$, where $y_i \in \{-1,1\} = \{\text{"Impostors"},\text{"Clients"}\}$. $\Phi : \mathbb{R}^R \to F$ maps the data into a feature space F. Vapnik [13] has proved that maximizing the minimum distance in space F between $\Phi(X)$ and the separating hyperplane $H(\mathbf{w},b) = \{\mathbf{f} \in F \mid \langle \mathbf{w},\mathbf{f} \rangle_F + b = 0\}$, (where $\langle \cdot,\cdot \rangle_F$ denotes inner product in space F), is a good means of reducing the generalization risk. Vapnik also proved that the optimal hyperplane can be obtained solving the quadratic programming problem:

$$
\begin{array}{lll}
\text{Minimize} & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i & \\
\text{with} & y_i(\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_F + b) \geq 1 - \xi_i & i = 1,..,l \\
 & \xi_i \geq 0 & i = 1,..,l
\end{array}
\tag{1}
$$

where constant C and slack variables ξ_i are introduced to take into account the eventual non-separability of $\Phi(X)$ into F. Applying the Karush-Kuhn-Tucker conditions to the problem in (1), the following sparse expression is obtained for the optimum:

$$\mathbf{w}^* = \sum_{i \in SV} \alpha_i y_i \Phi(\mathbf{x}_i) \tag{2}$$

where $SV = \{i \mid \alpha_i > 0\}$ is the set of support vectors. Taking into account that the decision function $D(\cdot)$ that classifies a test pattern $\mathbf{x}_T$ is:

$$D(\mathbf{x}_T) = \text{sign}\left\{\langle \mathbf{w}^*, \Phi(\mathbf{x}_T) \rangle_F + b^*\right\} \tag{3}$$

defining $K(\mathbf{x}_i, \mathbf{x}_j) = \left\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \right\rangle_F$ as the kernel function and using (2) we obtain:

$$D(\mathbf{x}_T) = \operatorname{sign}\left\{ \sum_{i \in SV} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_T) + b^* \right\} \tag{4}$$

Problem (1) is solved for $(\alpha_i)_{i=1}^{l}$ and b^* in its dual form which, together with decision function (4), avoids manipulating directly the elements of F. The choice for the kernel has been in this case a Radial Basis Function (RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2 / 2\sigma^2) \tag{5}$$

In [12], the fusion strategy relied on the computation of the decision function $D(\cdot)$. A modification in order to obtain not a final classifier decision, but a combined multimodal score based on the proximity of the test pattern to the separating surface, is proposed here. The combined score $s_T \in \mathbb{R}$ of the multimodal pattern $\mathbf{x}_T \in \mathbb{R}^R$ will be calculated as:

$$s_T = \sum_{i \in SV} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_T) + b^* \tag{6}$$

Following this approach, the decision threshold parameter can be adjusted to reach different working points. This modification also permits to compare competing multimodal fusion strategies in terms of DET plots, trading-off the two error rates of the verification task.

3 Experiments

3.1 Database Description

We have randomly selected 50 users from the MCYT Multimodal Database including fingerprint and on-line signature samples [7]. A subset of 50 different users from the XM2VTS face database [14] has been also randomly selected. From both subsets, and thanks to the independence of signature, fingerprint and face traits [3], we have created 50 chimeric individuals comprising fingerprint, signature and face traits.

The following training and testing procedure for monomodal systems had been established:

- *Training: i*) Fingerprint*:* Each client's index finger has been represented with 1 high-control minutiae pattern; *ii*) Signature*:* Each signature has been modelled with 6 samples, and *iii*) Face: Each face has been modelled with 4 samples, according to Configuration II of the Lausanne Protocol [15].
- *Testing: i*) Targets: 4 more samples of each trait (face, fingerprint and signature) have also been selected for tests (2 from evaluation and 2 from test data of the Lausanne Protocol in the case of face samples); *ii*) Impostors: 3 different impos-

tors (skilled forgeries in case of signature) for each client have been considered and, from each impostor, 5 samples have been selected.

Consequently, the subcorpus for the experiments consists of 50×4=200 client, and 50×3×5=750 impostor multimodal attempts.

3.2 Monomodal Baseline Systems

Standard performance individual verification systems (whose parameters have not been optimized) have been intentionally used because it makes the comparison of subsequent fusion strategies easier. In particular, we have considered: a face verification system based on a global face appearance representation scheme [16], a minutiae-based fingerprint verification system [17] and an on-line signature verification system based on HMM modelling of temporal functions [18].

3.3. Multimodal Experimental Procedure

For the rule-based fusion approach, all multimodal test scores (200 from clients and 750 from impostors) are used for testing the verification performance. For the learning-based fusion approach, user-independent and user-dependent strategies have been considered.

In the user-independent fusion scheme, a unique SVM is used for all users and the leave-one-out method [10], leaving out each one of the users, will be applied for testing (i.e., multimodal scores of one user will be combined with a SVM trained on the other users, generating thus 4 client and 15 impostor combined scores and this strategy will be carried out on the remaining 49 subjects, yielding 4×50=200 client and 15×50=750 impostor combined test scores). In the user-dependent fusion scheme, a different SVM is used for each user and the leave-one-out method will be again applied, in this case leaving out each one of the 4 client and 15 impostor multimodal scores of a particular user (i.e., for each user, the multimodal score leaved out will be combined with a SVM trained on the other scores of the selected user and the process will be repeated for each of the other scores yielding again 50×4=200 client and 50×15=750 impostor combined test scores).

3.4. Asymptotic Performance

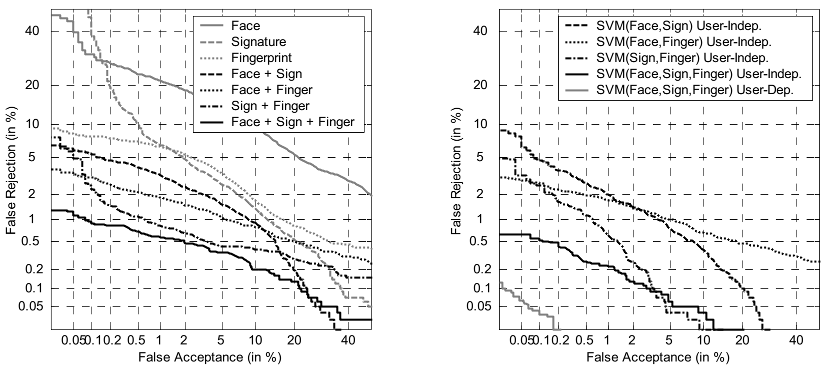
It has been demonstrated [11][12] that multimodal fusion schemes can have such a good performance that their comparison over a restricted size test data can be very difficult, if not impossible (leading even to error-free combined systems [12], due to the scarceness of data).

In the present contribution, a statistical-motivated experimental procedure denominated as *Asymptotic Performance* [19], which reduces side effects produced by data scarceness and avoids the uninformative possible "0.0% EER" result, has been used. The proposed statistically-motivated experimental procedure works as follows. Two Gaussian Mixture Models with 4 components each are estimated respectively from client and impostor score histograms using the EM algorithm [10]. Client and impostor scores are then randomly generated according to the resulting distributions and used as input data for the performance testing DET plots. The number of samples is chosen heuristically in order to balance the computational complexity and the smoothness of the resulting plots according to the following: the more separated client and impostor distributions are expected, the greater the number of sampled points.

3.5. Results

In Fig. 1 (a), the asymptotic performance of the monomodal baseline systems together with the performance of some combined systems based on Sum Rule are plotted. In Fig. 1 (b), the asymptotic performance of the developed SVM-based fusion approach is also represented for some combinations of the considered modalities.

In this case, and trying to approximate the *a posteriori* probabilities described in [9], monomodal scores are normalized into the range [0–1] before the operation by means of a linear mapping in case of fingerprint and face, and with an exponential mapping in case of on-line signature (due to the log-probability score [18]).



(a) (b)

Fig. 1. Performance of: (a) Monomodal baseline systems and Sume Rule-based multimodal systems, (b) SVM-based multimodal systems (RBF with $2\sigma^2/R = 0.1$, where R is the number of modalities, see [19] for a evaluation of other kernel parameters)

In order to visualize the discrimination capability of SVM classifiers for user-independent and user-dependent fusion approaches, client and impostor maps of signature and fingerprint scores before the fusion are plotted in Fig. 2 for all users (a) and for a random user (b). In case of user-independent fusion (a), the decision boundary (i.e., multimodal combined score = 0) and curves of equal multimodal score for one user of the leave-one-out procedure, whose client and impostor scores have been enlarged, have been included. In case of user-dependent fusion (b), the same information related to the selected user is also provided.

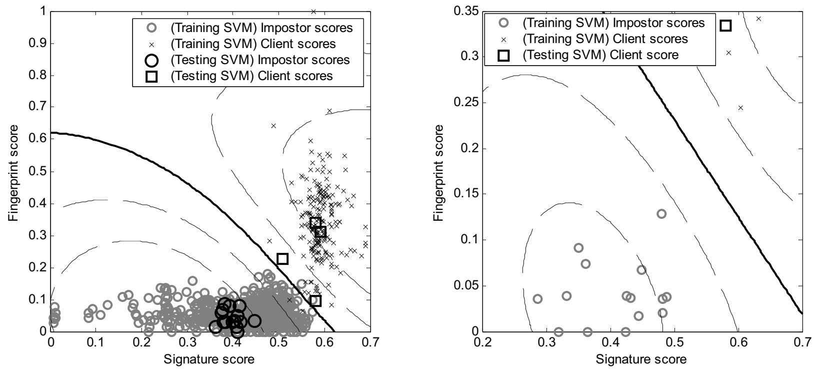


Fig. 2. Score map plot for user-independent (a) and user-dependent fusion (b). Curves of equal multimodal combined score are also plotted

4 Conclusions

A statistical-motivated experimental procedure has been introduced and applied in order to compare best referenced learning- and rule-based multimodal biometric fusion strategies by means of DET plots. A method for generating a multimodal combined score based on SVM classifiers has been proposed from which user-independent and user-dependent fusion strategies have been derived.

Appropriate selection of parameters for the learning-based approach has shown to provide better verification performance than the rule-based approach. In particular, starting from, approximately, a 10% EER face verification system, a 4% EER on-line signature verification system and a 3% EER fingerprint verification system, it has been shown that the Sum Rule reduced the EER to 0.5%. The RBF SVM fusion strategy performed even better reducing the EER to 0.3% and 0.05% respectively in case of user-independent and user-dependent fusion.

Encouraging initial results of the user-dependent learning-based approach motivate further research in order to exploit user specificities in the fusion stage of multimodal biometric verification systems.

Acknowledgements

This work has been supported by the Spanish Ministry for Science and Technology under project TIC2000-1669-C04-01. J. F.-A. also thanks Consejeria de Educacion de la Comunidad de Madrid and Fondo Social Europeo for supporting his doctoral research.

References

1. A. Jain, R. Bolle and S. Pankanti (eds.), *Biometrics – Personal Identification in Networkwed Society*, Kluwer Academic Publishers, 1999.
2. L. Xu, A. Kryzak and C.Y. Suen, "Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition", IEEE Trans. on Systems, Man and Cybernetics, vol. 22, no. 3, pp. 418-435, May-June 1992.
3. A. Ross, A. K. Jain, J. Z. Qian, "Information Fusion in Biometrics", Proc. of the 3rd Audio- and Video-Based Person Authentication, AVBPA, pp. 354-359, Halmstad, Sweden, 2001.
4. R. Brunelli and D. Falavigna, "Person Identification using Multiple Cues", IEEE Trans. Pattern Anal. and Machine Intell., vol. 17, no. 10, pp. 995-966, 1995.
5. E. Bigun, J. Bigun, B. Duc and S. Fisher, "Expert Conciliation for Multimodal Person Authentication Systems by Bayesian Statistics", Proc. of the 1st Audio and Video-Based Person Authentication, AVBPA'97, pp. 327-334, Springer Verlag, 1997.
6. L. Hong and A. K. Jain, "Integrating Faces and Fingerprints for Personal Identification", IEEE Trans. Pattern Anal. and Machine Intell., vol. 20, no. 12, pp. 1295-1307, 1998.
7. J. Ortega-Garcia, *et al.,* "MCYT: A Multimodal Biometric Database", Proc. of COST-275 Biometric Recognition Workshop, Rome (Italy), November 2002.
8. A. Martin, *et al.,* "The DET Curve in Assessment of Decision Task Performance", Proc. of ESCA 5th Eur. Conf. on Speech Comm. and Tech., EuroSpeech '97, pp. 1895-1898, Rhodes (Greece), 1997.
9. J. Kittler, M. Hatef, R.P.W. Duin and J. Matas, "On Combining Classifiers", IEEE Trans. Pattern Anal. and Machine Intell., vol. 20, no. 3, pp. 226-239, March 1998.
10. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 1999.
11. P. Verlinde, G. Chollet and M. Acheroy, "Multi-Modal Identity Verification using Expert Fusion", Information Fusion, no. 1, pp. 17-33, Elsevier, 2000.
12. B. Gutschoven and P. Verlinde, "Multi-Modal Identity Verification using Support Vector Machines (SVM)", Proc. of the 3rd Intl. Conf. on Information Fusion, 2000.
13. V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, 1999.
14. K. Messer, J. Matas, J. Kittler, J. Luettin and G. Maitre, "XM2VTSDB: The extended M2VTS database", Proc. of the 2nd Audio and Video-Based Person Authentication, AVBPA'99, pp. 72-77, Washington, D.C., March 1999.
15. J. Luettin, G. Maitre, "Evaluation protocol for the extended M2VTS database (Lausanne protocol)", Tech. Rep. 98-05 IDIAP, 1998. ftp.idiap.ch/pub/reports/1998/com98-05.ps.gz.
16. S. Cruz-Llanas, J. Fierrez-Aguilar, J. Ortega-Garcia and J. Gonzalez-Rodriguez, "A Comparative Evaluation of Global Representation-Based Schemes for Face Verification", Proc. Intl. Conf. on Image Processing, ICIP 2003 *(submitted for publication)*.
17. D. Simon, *et al.*, "An Improved Image Enhancement Scheme for Fingerprint Minutiae Extraction in Biometric Identification", In Proc. 3rd Intl. Conf. on Audio- and Video-Based Person Authentication, AVBPA 2001, Halmstad, Sweden, 6-8 June, 2001.
18. J. Ortega-Garcia, J. Fierrez-Aguilar, J. Martin-Rello and J. Gonzalez-Rodriguez, "Complete Signal Modeling and Score Normalization for Function-Based Dynamic Signature Verification", Proc. 4th Intl. Conf. on Audio- and Video-Based Person Authentication, AVBPA 2003 *(accepted for publication)*.
19. J. Fierrez-Aguilar, J. Ortega-Garcia and J. Gonzalez-Rodriguez, "Fusion strategies in Biometric Multimodal Verification", Proc. Intl. Conf. on Multimedia and Expo, ICME 2003 *(accepted for publication)*.