# FUSION STRATEGIES IN MULTIMODAL BIOMETRIC VERIFICATION

*J. Fierrez-Aguilar, J. Ortega-Garcia and J. Gonzalez-Rodriguez*

{jfierrez, jortega, jgonzalez}.diac.upm.es
Biometrics Research Lab. – Universidad Politecnica de Madrid – Spain

**ABSTRACT**

The aim of this paper, regarding multimodal biometric verification, is twofold: on the one hand, to review some score fusion strategies reported in the literature and, on the other hand, to compare experimentally a selection of them using as monomodal baseline systems our template-based face, minutiae-based fingerprint and HMM-based on-line signature verification systems on the MCYT multimodal database. A new strategy is proposed and discussed in order to compute a multimodal combined score by means of Support Vector Machine (SVM) classifiers.

## 1. INTRODUCTION

Biometric signals and traits (fingerprints, speech, face images, etc.) contain identity information about the subject they belong to. Automatic extraction of these cues 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 (mainly, related to *security*, *forensics* and *remote managing*) 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 if we assume statistical independence between them [3]. Some works related to the *multimodal fusion* approach are [3]-[6].

Based on the above-mentioned research on multimodal fusion, the aim of this paper is twofold: on the one hand, to review the fusion strategies reported in the literature and, on the other hand, to compare experimentally some of them using our single trait systems and the MCYT multimodal 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. In a verification context, two situations of error are possible: an impostor is accepted (false alarm, FA) or the correct user is rejected (false reject, 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 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). Considering $R$ modalities, 2 classes ($\omega_1$ for clients and $\omega_2$ for impostors), and a given pattern $Z$ that generates the feature vector $\mathbf{x}_i$ for modality $i$, the classifiers (or experts) are considered to give the *a posteriori* probability for each class $k$: $P(\omega_k \mid \mathbf{x}_i)$. Several ways to implement the fusion of the modalities are then obtained (sum, product, max, …), based on the Bayes theorem and certain hypothesis, from which the *Sum Rule*:

$$\text{assign} \quad Z \rightarrow \omega_j \quad \text{if}$$

$$j = \arg\max_{k=1,2} \left\{ (1-R)P(\omega_k) + \sum_{i=1}^{R} P(\omega_k \mid \mathbf{x}_i) \right\}$$

outperformed the remainder in the experimental comparison, due to its robustness to errors in the estimation of $P(\omega_k \mid \mathbf{x}_i)$ made by the individual classifiers. From now on, this perspective will be referred to as *rule-based* (or *fixed*) *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 Neighbors 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.

### 2.2. Rule-based fusion vs. learning-based fusion

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.

### 2.3. Multimodal fusion via SVM

We have used the SVM not to provide a binary verification decision, as has been reported in related works [11][12], but to provide a fused score combining the outputs of the considered monomodal systems. 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 $X = (\mathbf{x}_i)_{i=1}^{l} \subset \mathbf{R}^R$ , where $l$ is the number of training vectors, $\mathbf{R}$ stands for the real line and $R$ is the number of modalities, is labeled with two-class targets $(y_i)_{i=1}^{l}$ , where $y_i \in \{-1,1\} = \{"\mathit{Impostors}","\mathit{Clients}"\}$ . $\Phi : \mathbf{R}^R \to F$ 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 $H(\mathbf{w},b) = \{\mathbf{f} \in F \,|< \mathbf{w},\mathbf{f} >_F +b = 0\}$ , (where $<\cdot,\cdot>_F$ denotes inner product in space $F$), is a good means of reducing the generalization risk. Vapnik also proved [13] that the optimal hyperplane can be obtained solving the convex quadratic programming (QP) problem:

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

where constant $C$ and slack variables $\xi_i$ are introduced because of the eventual non-separability of $\Phi(X)$ in space $F$. Applying the Karush-Kuhn-Tucker conditions to the problem in (1), the following sparse expression is obtained for the optimal hyperplane $H(\mathbf{w}^*,b^*)$:

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

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

$$D(\mathbf{x}_T) = sign\{< \mathbf{w}^*, \Phi(\mathbf{x}_T) >_F + b^*\} \tag{3}$$

defining $K(\mathbf{x}_i,\mathbf{x}_j) =< \Phi(\mathbf{x}_i),\Phi(\mathbf{x}_j) >_F$ as the kernel function and using (2) leads to:

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

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

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

In [12], the fusion strategy relied on the computation of the decision function $D$. 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 $s_T \in \mathbf{R}$ of the multimodal pattern $\mathbf{x}_T \in \mathbf{R}^R$ 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 verification 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 (FA and FR) of the verification task.

## 3. EXPERIMENTAL COMPARISON

### 3.1. Database description

We have randomly selected 50 subjects from the MCYT Multimodal Database including fingerprint and on-line signature samples. For the experiments, a subset of 50 different subjects from the XM2VTS face database [14] have been also randomly selected. It has been supposed for convenience, and thanks to the independence of signature, fingerprint and face traits [3], that the individuals from MCYT and from XM2VTS coincide.

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-controlled minutiae pattern; *ii) Signature:* Each signature has been modeled with 6 samples, and *iii) Face:* Each face has been modeled with 4 samples, according to Configuration II of the Lausanne Protocol [14].
- *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 impostors (skilled forgeries in the 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 scores.

### 3.2. Monomodal baseline systems

Medium performance individual verification systems have been intentionally used because it makes the comparison of subsequent fusion strategies easier. In particular, and taking into account the above-mentioned training and testing database structure, we have considered: a 11.5%-EER template-based face verification system, a 2.6%-EER minutiae-based fingerprint verification system and a 4.8%-EER HMM-based on-line signature verification system. For a detailed description of these systems, see [15].

### 3.3. Multimodal experimental procedure

For rule-based fusion strategies, all multimodal test scores (200 from clients and 750 from impostors) are used for testing the verification performance. For learning-based fusion strategies, the leave-one-out method [10] is used to maximize the size of training and testing data for the learning machine, while maintaining their independence. Multimodal scores of one user are combined with a SVM trained on other users, generating thus 4 client and 15 impostor combined scores. This strategy is carried out on the remaining 49 subjects, yielding 4×50=200 client and 15×50=750 impostor combined test scores.

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*, which reduces side effects produced by data scarceness and avoids the uninformative zero EER result, has been used. The proposed statistically-motivated experimental procedure works as follows. Two Gaussian Mixture Models (GMM) with 4 components each are estimated respectively from client and impostor score histograms using the EM algorithm [10]. Then, 10,000 points from the resulting distributions are generated and used as input data for the performance testing DET plots. This procedure also simplifies the comparison of competing fusion strategies because it smoothes the performance plots.

### 3.4. Results

In Figure 1, the asymptotic performance of the monomodal baseline systems together with the performance of some rule-based combined systems are plotted. In this case, and trying to approximate the above-mentioned *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 (to undo de log-probability given by the HMM).

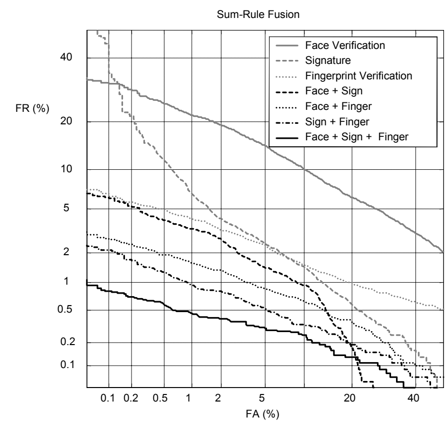


Figure 1. Performance of monomodal baseline systems and rule-based fusion strategies.

Results for the learning-based approach are plotted in the upper row of Figure 2. From left to right, DET performance plots of the RBF-SVM fusion strategy on normalized scores with three different kernel parameters are included. In all plots, the asymptotic detection performance of the Sum Rule on normalized scores is plotted in gray for better comparison.

In order to visualize the discrimination capability of the trained RBF-SVM fusion approach, client and impostor maps of signature and fingerprint scores before the fusion are plotted in the lower part of Figure 2. Decision boundaries and curves of equal combined multimodal score for one user of the leave-one-out procedure, whose client and impostor scores have been enlarged, have been also included.

## 4. CONCLUSIONS

A statistical-motivated experimental procedure has been introduced and applied to compare best published learning-based and rule-based fusion strategies by means of DET plots. Appropriate selection of parameters for the learning-based scheme has led to a fusion strategy that clearly outperforms the rule-based strategy.

Starting from a 11.5% EER face verification system, a 4.8% EER on-line signature verification system and a 2.6% EER fingerprint verification system, it has been shown that the Sum Rule reduced the EER to 1%. The RBF SVM fusion strategy performed even better reducing the EER to 0.3%.

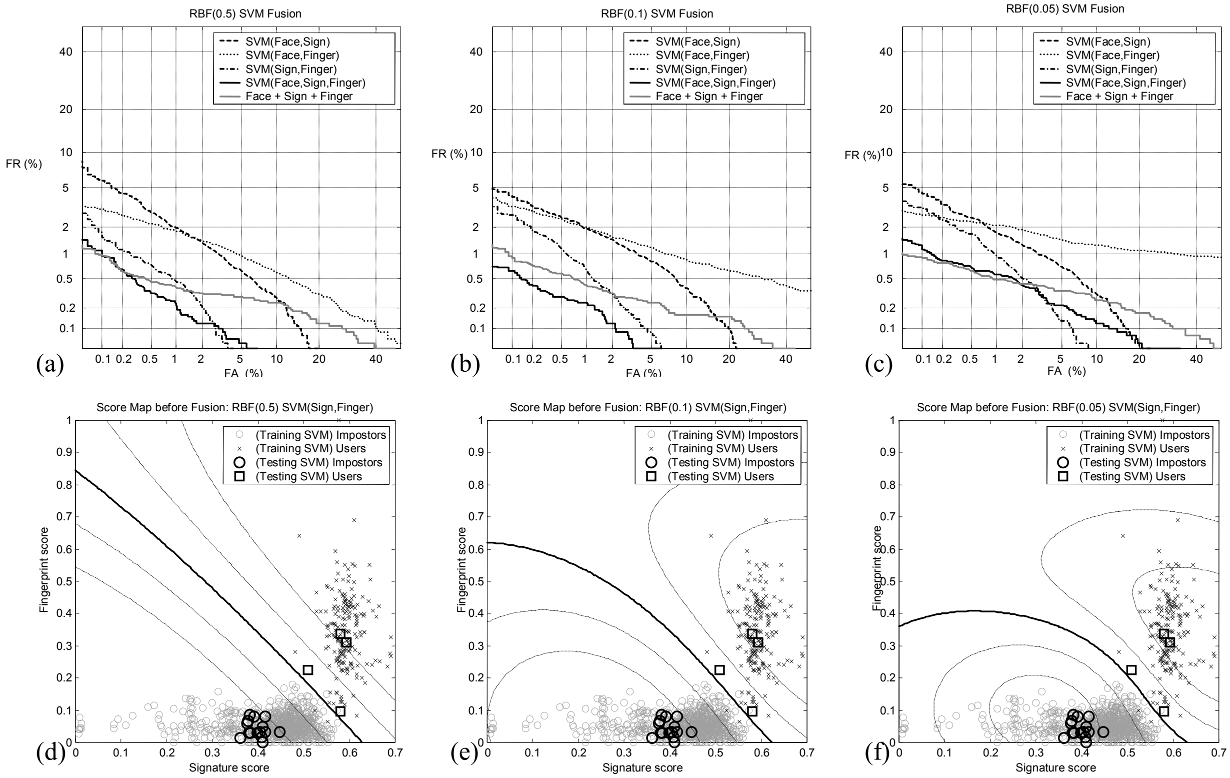


Figure 2. Learning-based fusion strategies. Upper row: performance of RBF-SVM fusion for different different kernel parameters (a),(b),(c): $\sigma^2 = \{0.5, 0.1, 0.05\}$ . Lower row: score map plots for different kernel parameters (d),(e),(f): $\sigma^2 = \{0.5, 0.1, 0.05\}$ .


## 5. 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.


## 6. REFERENCES


[1] A. Jain, R. Bolle and S. Pankanti (eds.), *Biometrics – Personal Identification in Networkwed Society*, Kluwer Academic Pub., 1999.

[2] L. Xu, A. Kryzak, C.Y. Suen, "Methods of Combining Multiple Classifiers and Their Application to Handwriting Recognition", *IEEE Trans. on Systems, Man and Cyber.*, vol. 22, no. 3, pp. 418-435, 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'01, pp. 354-359, Halmstad, Sweden, 2001.

[4] R. Brunelli, D. Falavigna, "Person Identification using Multiple Cues", *IEEE Trans. on Pattern Analysis and Machine Intell.*, vol. 17, no. 10, pp. 995-966, 1995.

[5] E. Bigun, J. Bigun, B. Duc, 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, Ed. Springer Verlag, 1997.

[6] L. Hong, A. K. Jain, "Integrating Faces and Fingerprints for Personal Identification", *IEEE Trans. on 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, K. Koutroumbas, *Pattern Recognition*, Academic Press, 1999.

[11] P. Verlinde, G. Chollet, M. Acheroy, "Multi-Modal Identity Verification using Expert Fusion", *Information Fusion*, no. 1, pp. 17-33, Ed. Elsevier, 2000.

[12] B. Gutschoven, 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, G. Maitre, "XM2VTSDB: The extended M2VTS database", Proc. of the 2nd Audio and Video-Based Person Authentication, pp. 72-77, Washington, D.C., March 1999.

[15] J. Ortega-Garcia, *et al*. "Fusion Strategies on Unimodal Fingerprint, Face, and On-Line Signature Verification Systems through MCYT Database", *IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics,* 2003. *(submitted for publication)*