EXPLOITING GENERAL KNOWLEDGE IN USER-DEPENDENT FUSION STRATEGIES FOR MULTIMODAL BIOMETRIC VERIFICATION

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

Speech and Signal Processing Group (ATVS) - Universidad Politecnica de Madrid (UPM), Spain
email: {jfierrez, jortega, jgonzalez}@diac.upm.es

ABSTRACT

In this paper, a novel strategy for combining general and user-dependent knowledge in a multimodal biometric verification system is presented. It is based on SVM classifiers and trade-off coefficients introduced in the standard SVM training problem. Experiments are reported on a bimodal biometric system based on fingerprint and on-line signature traits. A comparison between three fusion strategies, namely user-independent, user-dependent and the proposed adapted user-dependent is carried out. As a result, the suggested approach outperforms the former ones. In particular, a highly remarkable relative improvement of 68% in the EER with respect to the user-independent approach is achieved. The severe and very common problem of training data scarcity in the user-dependent strategy is also relaxed by the proposed scheme, resulting in a relative improvement of 40% in the EER compared to the raw user-dependent strategy.

1. INTRODUCTION

Automatic extraction of identity cues from personal traits (e.g., fingerprints, speech, or face images) has given raise to a particular branch of pattern recognition (*biometrics*) where the goal is to infer identity of people from personal 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.

Biometrics provides a way to establish an identity based on '*who you are*', rather than by '*what you possess*' or '*what you know*' [2]. This concept not only ensures enhanced security but also avoids the need to remember and maintain multiple passwords.

Previous studies have shown 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 [3]. Even greater verification performance improvement can be expected through the use of multiple biometric characteristics [4]. The reader is referred to [4, 5, 6, 7] for some works on multimodal fusion.

2. MOTIVATION

The work presented in this paper is motivated by two general ideas whose benefits have been demonstrated in previous work:

This work has been supported by the Spanish Ministry for Science and Technology under projects TIC2003-09068 and TIC2003-08382. J. F.-A. also thanks Consejeria de Educacion de la Comunidad de Madrid and Fondo Social Europeo for supporting his doctoral research.

User-specific parameters: Recent advances in multimodal biometric verification systems have been accomplished by learning user-specific parameters [2, 7].

Adaptive learning: The adaptive learning framework offers a way to incorporate new or specific data into existing classification structures and combine them in an optimal manner [8, 9].

Regarding user-specific parameters, some related works are: i) threshold selection and degree of importance assigned to each of the biometric traits, as described in [2]; and, specifically related to this work, ii) the trained user-dependent fusion approach formulated in [7] whose experimental results showed a significant improvement in the system performance, and hence, provides a promising starting point for this work.

As far as adaptive learning, it is generally known that, unfortunately, in pattern recognition applications we rarely have the complete knowledge about the structure of the problem. In a typical case we merely have some general knowledge about the situation, together with a number of training data, which are particular representatives of the patterns we want to classify or, in this case, fuse by means of a classifier. In many cases, the amount of available training data is not sufficient and representative enough to guarantee good parameter estimation/learning and generalization capabilities.

To cope with this lack of robustness derived from partial knowledge of the structure of the problem, the use of robust adaptive decision/fusion strategies based on "all" the available information have been proposed [8]. As an example of the underlying philosophy, we may consider the fact that general information of the problem (such as user or task independent features) can constitute a rich source of information and a valuable starting point for user-specific recognition problems.

Based on the two general ideas above mentioned, the aim of this paper is to provide a specific framework for user-dependent multimodal biometric fusion incorporating the general knowledge provided by pooling user-independent data.

3. MULTIMODAL FUSION SCHEME

The proposed fusion scheme is derived from user-independent and user-dependent fusion strategies [7] based on SVM classifiers [10]. In first place, the notation is established and a brief description of the above mentioned trained fusion approaches is given. Then, the proposed fusion scheme is presented.

3.1. SVM-based multimodal fusion

Given a multimodal biometric verification system consisting of R different unimodal systems $r = 1, \ldots, R$, each one computes a similarity score $x_r \in \mathbb{R}$ between an input biometric pattern and the enrolled pattern of given claimant. Let the similarity scores, provided by the different unimodal systems, be combined into a multimodal score $\mathbf{x} = [x_1, \ldots, x_R]^T$. The design of a trained fusion scheme consists in the estimation of a function $f : \mathbb{R}^R \rightarrow \mathbb{R}$ based on empirical data so as to maximize the separability of client $\{f(\mathbf{x})|\text{client attempt}\}$ and impostor $\{f(\mathbf{x})|\text{impostor attempt}\}$ fused score distributions.

Formally, let the training set be $X = (\mathbf{x}_i, y_i)_{i=1}^N$ where N is the number of multimodal scores in the training set, and $y_i \in \{-1, 1\} = \{\text{Impostor}, \text{Client}\}$. The principle of SVM relies on a linear separation in a high dimension feature space $\mathbb{H}$ where the data have been previously mapped via $\Phi : \mathbb{R}^R \rightarrow \mathbb{H}; X \rightarrow \Phi(X)$, so as to take into account the eventual non-linearities of the problem [10]. In order to achieve a good level of generalization capability, the margin between the separator hyperplane

$$\left\{\mathbf{h} \in \mathbb{H} | \left\langle \mathbf{w}, \mathbf{h} \right\rangle_{\mathbb{H}} + w_0 = 0\right\} \tag{1}$$

and the mapped data $\Phi(X)$ is maximized (where $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ denotes inner product in space $\mathbb{H}$, and $(\mathbf{w} \in \mathbb{H}, w_0 \in \mathbb{R})$ are the parameters of the hyperplane). The optimal hyperplane can be obtained as the solution of the following quadratic programming problem [10]:

$$\min_{\mathbf{w}, w_0, \xi_1, \ldots, \xi_N} \left(\tfrac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \right) \tag{2}$$

subject to

$$y_i(\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_{\mathbb{H}} + w_0) \geq 1 - \xi_i, \quad i = 1, \ldots, N \tag{3}$$

$$\xi_i \geq 0, \qquad i = 1, \ldots, N \tag{4}$$

where slack variables ξ_i are introduced to take into account the eventual non-separability of $\Phi(X)$ into $\mathbb{H}$ and parameter C is a positive constant that controls the relative influence of the two competing terms.

The optimization problem in (2), (3) and (4) is solved using the Wolfe dual representation [11]:

$$\max_{\alpha_1, \ldots, \alpha_N} \left(\sum_{i=1}^{N} \alpha_i - \tfrac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \tag{5}$$

subject to

$$\begin{array}{l} 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, N \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \end{array} \tag{6}$$

where the introduction of the kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathbb{H}}$ avoids direct manipulation of the elements of $\mathbb{H}$. In particular, a Radial Basis Function (RBF) kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ has been used for the reported experiments.

In [7], the fused score s_T of a multimodal test pattern $\mathbf{x}_T$ is defined as follows

$$s_T = f(\mathbf{x}_T) = \langle \mathbf{w}^*, \Phi(\mathbf{x}_T) \rangle_{\mathbb{H}} + w_0^* \tag{7}$$

which, applying the Karush-Kuhn-Tucker (KKT) conditions to the problem in (2), can be shown to be equivalent to the following sparse expression

$$s_T = f(\mathbf{x}_T) = \sum_{i \in SV} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}_T) + w_0^* \tag{8}$$

where $(\mathbf{w}^*, w_0^*)$ is the optimal hyperplane, $(\alpha_1^*, \ldots, \alpha_N^*)$ is the solution to the problem in (5), (6) and $SV = \{i | \alpha_i^* > 0\}$ is the set of support vectors. w_0^* is obtained from the solution to the problem in (5), (6) by using the KKT conditions (see [11] for more details).

As a result, the training procedure in (5), (6) and the testing strategy in (8) are obtained for the problem of multimodal fusion.

3.2. User-independent and user-dependent fusion

In the user-independent case, the training set $X_{UI} = (\mathbf{x}_i, y_i)_{i=1}^{N_{UI}}$ includes multimodal scores from a number of different clients and the obtained fusion rule $f_{UI}(\mathbf{x})$ is applied at the operational stage regardless of the claimed identity.

Regarding the user-dependent case, a different fusion rule $f_j(\mathbf{x})$ is obtained for each client enrolled in the system $j = 1, \ldots, M$ by means of a training set comprising only multimodal scores of the specific client X_j. At the operational stage, the fusion rule $f_j(\mathbf{x})$ of the client j being claimed is applied.

3.3. Adapted user-dependent strategy

A user-dependent fusion scheme trading off the general knowledge provided by a user-independent training set $X_{UI} = (\mathbf{x}_i, y_i)_{i=1}^{N_{UI}}$ and the user specificities provided by a user-dependent training set $X_j = (\mathbf{x}_i, y_i)_{i=N_{UI}+1}^{N_{UI}+N_j}$ is proposed here. The suggested training procedure for client j is stated as follows:

$$\min_{\mathbf{w}, w_0, \xi_1, \ldots, \xi_{N_{UI}+N_j}} \left(\tfrac{1}{2} \|\mathbf{w}\|^2 + C_{UI} \sum_{i=1}^{N_{UI}} \xi_i + C_{UD} \sum_{i=N_{UI}+1}^{N_{UI}+N_j} \xi_i \right) \tag{9}$$

subject to

$$y_i(\langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle_{\mathbb{H}} + w_0) \geq 1 - \xi_i, \quad i = 1, \ldots, N_{UI} + N_j \tag{10}$$

$$\xi_{ij} \geq 0, \qquad i = 1, \ldots, N_{UI} + N_j \tag{11}$$

This can be seen as a user-dependent fusion scheme adapted from user-independent information. Sequential algorithms for the solution of the SVM optimization problem in (2), (3), (4) have been already proposed [9], and can be adjusted to deal with the problem in (9), (10), (11), first constructing the user-independent solution and then refining it by incorporating the user-dependent information. Nevertheless, in the present work, a simpler batch mode implementation has been used.

Following the derivation in [11], it can be shown that the dual representation of the problem in (9), (10), (11) is as follows:

$$\max_{\alpha_1, \ldots, \alpha_{N_{UI}+N_j}} \left(\sum_{i=1}^{N_{UI}+N_j} \alpha_i - \tfrac{1}{2} \sum_{i,k=1}^{N_{UI}+N_j} \alpha_i \alpha_k y_i y_k K(\mathbf{x}_i, \mathbf{x}_k) \right) \tag{12}$$

subject to

$$\begin{array}{l} 0 \leq \alpha_i \leq C_{UI}, \quad i = 1, \ldots, N_{UI} \\ 0 \leq \alpha_i \leq C_{UD}, \quad i = N_{UI} + 1, \ldots, N_{UI} + N_j \\ \sum_{i=1}^{N_{UI}+N_j} \alpha_i y_i = 0 \end{array} \tag{13}$$

For the experiments in the next section, the problems in (5), (6) and (12), (13) have been solved by using an interior point optimization solver.

4. EXPERIMENTS

Some experiments comparing the fusion schemes above described on a bimodal biometric verification system consisting of fingerprint and on-line signature subsystems have been carried out. Monomodal systems are first introduced providing references for further details. Then, the database and the experimental procedure are described. Finally, results are summarized.

4.1. Baseline monomodal systems

Individual verification systems with standard performance (parameters have not been optimized) have been intentionally used because it makes the comparison of subsequent fusion strategies easier. In particular, the experiments have been carried out on a bimodal biometric verification system including the minutiae-based fingerprint verification subsystem described in [12] and the on-line signature verification subsystem based on temporal functions and Hidden Markov Models reported in [13].

4.2. Database description

50 users have been randomly selected from the MCYT Bimodal Database including fingerprint and on-line signature samples [14]. The following training and testing procedures for monomodal systems have been established:

Training: i) Fingerprint: Each client's index finger has been represented with one high-control minutiae pattern; ii) Signature: Each signature has been modelled with 6 samples.

Testing: i) Clients: 4 samples of each trait (fingerprint and signature) have also been selected for tests; ii) Impostors: 3 different impostors (skilled forgeries in case of signature) for each client have been considered and, for each impostor, 5 samples have been selected.

Consequently, the subcorpus for the fusion experiments consists of $50 \times 4 = 200$ client, and $50 \times 3 \times 5 = 750$ impostor bimodal scores. Individual verification performance of the monomodal subsystems on the data set are depicted in Fig. 1 as DET curves.

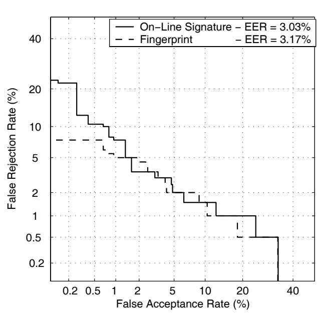


Fig. 1. Verification performance of baseline monomodal systems

4.3. Multimodal experimental procedure

Several methods have been described in the literature in order to maximize the use of the information embedded in the training samples during a test [11]. For error estimation in multimodal authentication systems, variants of jackknife sampling using the leave-one-out principle are a common choice [4, 7]. In this work, and depending on the experiment at hand, a variant of bootstrap sampling has been used:

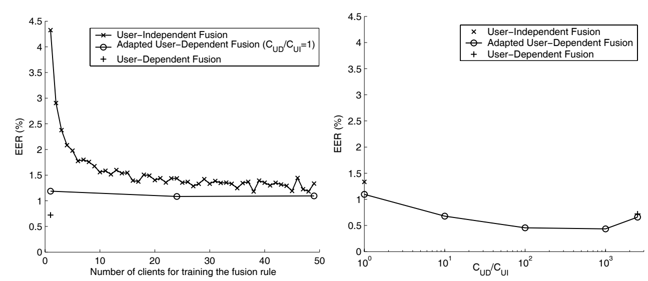
User-Independent Fusion: Bootstrap data sets have been created by randomly selecting M users from the training set with replacement. This selection process have been independently repeated 200 times to yield 200 bootstrap data sets. Each data set is used then to generate a user-independent fusion rule. Testing is finally performed on the remaining users not included in each bootstrap data set.

User-Dependent Fusion: For each user, 50 bootstrap data sets have been created randomly selecting with replacement (and forcing at least one sample in each class client/impostor) $N = 10$ samples. For each user and bootstrap data set, a different fusion rule is constructed and testing is performed on the remaining samples not included in the bootstrap data set.

Adapted User-Dependent Fusion: Bootstrap sampling of users is performed as in the user-independent case yielding 200 user-independent bootstrap data sets (UIBD). Multimodal scores of the remaining users not included in each UIBD are then sampled as in the user-dependent case. This yields 50 user-dependent bootstrap data sets per UIBD (which is used for training the user-dependent fusion rule) and per client not included in the UIBD. Testing is performed on the remaining samples not included in each user-dependent bootstrap data set.

4.4. Results

Results from a subset of experiments comparing user-independent, user-dependent and the proposed adapted user-dependent fusion strategies are summarized in Fig. 2.



(a) (b)

Fig. 2. Equal error rates of the studied fusion schemes

In Fig. 2 (a), verification performance of the bimodal authentication system is shown for an increasing number of clients in the fusion rule training set. In the case of user-independent fusion, error rate drops from 4.33% ($M = 1$ client for training) to 1.34% ($M = 49$). For the adapted user-dependent scheme ($C_{UI} = 100$ and $C_{UD} = 100$), the error rate decreases from 1.19% to 1.09%. On the other hand, the raw user-dependent fusion strategy goes down to 0.72% EER.

The main result of the present work is shown in Fig. 2 (b). In this case, $M = 49$ and $C_{UD} = 100$ are fixed and C_{UD}/C_{UI} is varied (hence trading off the influence of the user-independent and user-dependent information for training the fusion rule). As a result, a minimum of 0.43% EER is found for $C_{UD}/C_{UI} = 1000$.

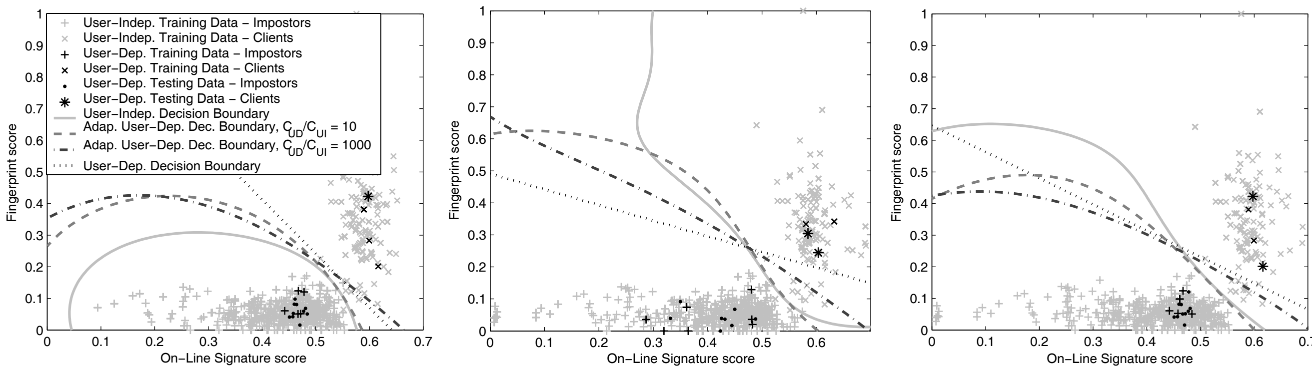


Fig. 3. Training/testing scatter plot and decision boundaries for the studied fusion schemes

In order to visualize the discriminative capability of SVM classifiers in the above described fusion approaches, client and impostor maps of signature and fingerprint scores before fusing are plotted in Fig. 3. In particular, three different data sets of the bootstrap error estimation process are depicted from left to right. User-independent, user-dependent and adapted user-dependent decision boundaries (i.e., multimodal combined score $s = f(\mathbf{x}) = 0$) have been included. The graphical effect of the latter approach being transformed from the user-independent to the user-dependent one is highly remarkable.

5. CONCLUSIONS

A user-dependent fusion scheme based on the general knowledge provided by pooling user-independent data has been introduced. This scheme is based on SVM classifiers and trade-off coefficients introduced in the standard SVM training problem. Appropriate selection of these parameters leads to an adapted user-dependent fusion scheme outperforming the raw user-dependent strategy due to the scarcity of training data for the later strategy.

In particular, and considering 3.03% EER signature and 3.17% EER fingerprint verification systems, it has been shown that user-independent and user-dependent multimodal fusion schemes reduced the EER down to 1.34% and 0.72%, respectively. The proposed adapted user-dependent strategy performed even better, reducing the EER down to 0.43%. As a result, the main problem of the user-dependent trained fusion approach, namely the training data scarcity, has been relaxed by the inclusion of user-independent information.

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

6. REFERENCES

[1] A. Jain, R. Bolle, and S. Pankanti (eds.), *Biometrics - Personal Identification in a Networked Society*, Kluwer, 1999.

[2] A. K. Jain and A. Ross, "Learning user-specific parameters in a multibiometric system," in *Proc. of ICIP*, 2002, vol. 1, pp. 57–60.

[3] L. Xu, A. Kryzak, and C.Y. Suen, "Methods of combining multiple classifiers and their application to handwritten recognition," *IEEE Trans. on SMC*, vol. 22, no. 3, pp. 418–435, 1992.

[4] J. Bigun, B. Duc, S. Fischer, A. Makarov, and F. Smeraldi, "Multi modal person authentication," in *Nato-Asi advanced study on face recogniton*, H. Wechsler et. al., Ed. 1997, vol. F-163, pp. 26–50, Springer.

[5] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. on PAMI*, vol. 20, no. 3, pp. 226–239, 1998.

[6] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. on NN*, vol. 10, no. 5, pp. 1065–1074, 1999.

[7] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero, and J. Gonzalez-Rodriguez, "A comparative evaluation of fusion strategies for multimodal biometric verification," in *Proc. of AVBPA*, 2003, pp. 830–837.

[8] C.-H. Lee and Q. Huo, "On adaptive decision rules and decision parameter adaptation for automatic speech recognition," *Proc. of the IEEE*, vol. 88, no. 8, pp. 1241–1269, 2000.

[9] A. Navia-Vazquez, F. Perez-Cruz, A. Artes-Rodriguez, and A.-R. Figueiras-Vidal, "Weighted least squares training of SV classifiers leading to compact and adaptive schemes," *IEEE Trans. on NN*, vol. 12, no. 5, pp. 1047–1059, 2001.

[10] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 2000.

[11] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2003.

[12] D. Simon-Zorita, J. Ortega-Garcia, J. Fierrez-Aguilar, and J. Gonzalez-Rodriguez, "Image quality and position variability assessment in minutiae-based fingerprint verification," *IEE Proc. VISP*, vol. 150, no. 6, 2003.

[13] 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," in *Proc. of AVBPA*, 2003, pp. 658–667.

[14] J. Ortega-Garcia et al., "MCYT baseline corpus: A bimodal biometric database," *IEE Proc. VISP*, vol. 150, no. 6, 2003.