Kernel-based multimodal biometric verification using quality signals

J. Fierrez-Aguilar[a], J. Ortega-Garcia[a], J. Gonzalez-Rodriguez[a] and Josef Bigun[b]

[a]Univ. Politecnica de Madrid, DIAC EUITT Ctra. Valencia km. 7, 28031 Madrid, Spain;
[b]Halmstad University, Box 823, Kristian IVs vag, S-301 18 Halmstad, Sweden

ABSTRACT

A novel kernel-based fusion strategy is presented. It is based on SVM classifiers, trade-off coefficients introduced in the standard SVM training and testing procedures, and quality measures of the input biometric signals. Experimental results on a prototype application based on voice and fingerprint traits are reported. The benefits of using the two modalities as compared to only using one of them are revealed. This is achieved by using a novel experimental procedure in which multi-modal verification performance tests are compared with multi-probe tests of the individual subsystems. Appropriate selection of the parameters of the proposed quality-based scheme leads to a quality-based fusion scheme outperforming the raw fusion strategy without considering quality signals. In particular, a relative improvement of 18% is obtained for small SVM training set size by using only fingerprint quality labels.

Keywords: Biometrics, multimodal authentication, support vector machine, fingerprint, speaker, quality

1. INTRODUCTION

Automatic access of persons to services is becoming increasingly important in the information era. Although person authentication by machine has been a subject of study for more than thirty years,[1,2] it has not been until recently that the matter of combining a number of different traits for person verification has been considered.[3,4] There are a number of benefits of doing so, just to name a few: false acceptance and false rejection error rates decrease, the authentication system becomes more robust against individual sensor or subsystem failures and the number of cases where the system is not able to give an answer (e.g., bad quality fingerprints due to manual work in fingerprint verification or larynx disorders in speaker verification) vanishes. The technological environment is also appropriate because of the widespread deployment of multimedia-enabled mobile devices (PDAs, 3G mobile phones, tablet PCs, laptops on wireless LANs, etc.). As a result, much research work is currently being done in order to fulfil the requirements of applications for masses.

Two early theoretical frameworks for combining different machine experts in a multibiometric system have been described respectively by Bigun and Kittler.[4,5] From these studies, the former from a risk analysis perspective and the later from a statistical pattern recognition point of view,[6,7] it can be concluded (under some mild conditions which normally hold in practice) that the weighted average is a good way of conciliating the different authenticity scores from individual modalities of a multimodal verification system.

From a practical point of view, multimodal verification has also been studied as a two-class classification problem by using a number of machine learning paradigms, for example: neural networks, decision trees and support vector machines (SVM).[8–12] These studies have shown performance gains with trained classifiers, and favored support vector machines over neural networks and decision trees. As a conclusion, some design guidelines for a multibiometric system are known and well accepted.

Current trends in multimodal biometrics research include the exploitation of user-specific parameters,[13,14] and quality signals.[15,16] In this work, we propose and investigate a novel quality-based adaptive trained multimodal

Further author information: (Send correspondence to J.F.-A.)
J.F.-A.: E-mail: jfierrez@diac.upm.es
J.O.-G.: E-mail: jortega@diac.upm.es
J.G.-R.: E-mail: jgonzalez@diac.upm.es
J.B.: E-mail: Josef.Bigun@ide.hh.se

fusion scheme based on support vector machines. With adaptive fusion scheme, we mean that the fusion scheme readapts to each identity claim as a function of the estimated quality of the input biometric signal.

The paper is structured as follows. The fusion scheme based on support vector machines from which the proposed quality-based strategy is derived is first summarized in Sect. 2. In the following the proposed algorithm is described. The state-of-the-art components of a multimodal authentication application, namely minutiae-based fingerprint and GMM-UBM speaker verification subsystems,[17,18] are then briefly described in Sect. 3. Some experiments using the above-mentioned multimodal authentication prototype on real data are reported in Sect. 4, where some guidelines for the computation of the parameters involved are detailed and the benefits of the proposed adaptive fusion scheme are revealed. Conclusions will be finally given in Sect. 5.

2. MULTIMODAL FUSION SCHEME

The proposed quality-based fusion scheme is derived from a raw user-independent fusion strategy based on SVM classifiers.[14,19] In first place, the notation is established and a brief description of the above mentioned approach is given. Then, the proposed quality-guided fusion scheme is presented.

2.1. 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 the claimant. Let the similarity scores, provided by the different unimodal systems, be combined into a multimodal score $\mathbf{x} = [x_1, \ldots, x_R]'$, where $'$ denotes transpose. 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.[19] In order to achieve a good level of generalization capability, the margin between the separator hyperplane

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

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:[19]

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

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_i is a positive constant that controls the relative influence of the two competing terms (the higher the C_i the higher the importance of associated training sample $(\mathbf{x}_i, y_i)$).

The optimization problem in (2), (3) and (4) is solved using the dual representation:[19]

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

subject to

$$\begin{aligned} &0 \leq \alpha_i \leq C_i, \quad i = 1, \ldots, N \\ &\sum_{i=1}^{N} \alpha_i y_i = 0 \end{aligned} \tag{6}$$

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 dot product kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i' \mathbf{x}_j$ leading to linear separation surfaces, and thus to weighted average trained fusion schemes as recommended in previous works,[4, 5, 12] has been used for the reported experiments.

The fused score s_T of a multimodal test pattern $\mathbf{x}_T$ is defined as follows[11]

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

which, applying the Karush-Kuhn-Tucker (KKT) conditions to the problem in (2), (3) and (4) 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 $SV = \{i | \alpha_i^* > 0\}$ indexes the set of support vectors. w_0^* is obtained from the solution to the problem in (5), (6) by using the KKT conditions.[20]

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

In this work we focus on user-independent SVM fusion. In this case, the training set $X = (\mathbf{x}_i, y_i)_{i=1}^N$ includes multimodal scores from a number of different clients and the obtained fusion rule $f(\mathbf{x})$ is applied at the operational stage regardless of the claimed identity.

2.2. Quality-Based Fusion Strategy

Let $\mathbf{q} = [q_1, \ldots, q_R]'$ denote the quality vector of the multimodal similarity score $\mathbf{x} = [x_1, \ldots, x_R]'$, where q_r is a quality value corresponding to similarity score x_r with $r = 1, \ldots, R$ and R is the number of modalities. In this work, the quality values q_r are computed as follows[15]

$$q_r = \sqrt{Q_r \cdot Q_{r,claim}} \tag{9}$$

where Q_r and $Q_{r,claim}$ are the quality label of the input signal for biometric trait r and the average signal quality of the biometric samples used by unimodal system r for modelling the claimed identity respectively. The two quality labels Q_r and $Q_{r,claim}$ are supposed to be in the range $[0, Q_{max}]$ with $Q_{max} > 1$ where 0 corresponds to the poorest quality, 1 corresponds to normal quality and Q_{max} corresponds to the highest quality. As a result $\mathbf{q} = [q_1, \ldots, q_R]'$ is computed from quality measures on the audio- or video-based input biometric signals (e.g., SNR or pitch deviations in case of voice utterances,[21] orientation certainty in case of fingerprint images,[22] etc.).

The proposed quality-guided fusion scheme (from now on also referred to as SVM_Q) is based on using the quality vector $\mathbf{q} = [q_1, \ldots, q_R]'$ as follows (the bimodal case $R = 2$ is described, generalization to the multimodal case is under investigation):

1. (SVM_Q Training) An initial fusion scheme (SVM) is trained as described in Sect. 2.1 by using

$$C_i = C \left(\frac{q_{i,1} q_{i,2}}{Q_{max}^2} \right)^{\alpha_1} \tag{10}$$

where $q_{i,1}$ and $q_{i,2}$ are the components of the quality vector $\mathbf{q}_i$ associated with training sample $(\mathbf{x}_i, y_i)$ and C is a positive constant. As a result, the higher the overall quality of a multimodal training score

the higher its contribution to the fusion scheme. Additionally, two SVMs of dimension one (SVM_1 and SVM_2) are trained by using training data from respectively first and second traits. Similarly to Eq. (10), $C_i = C(q_{i,j}/Q_{max})^{\alpha_1}$ for SVM_j with $j = 1, 2$.

2. (SVM_Q Authentication Phase) At this step, the three above-mentioned classifiers SVM, SVM_1 and SVM_2 are trained (i.e., the combining functions $f_{\text{SVM}}(\cdot)$, $f_{\text{SVM}_1}(\cdot)$ and $f_{\text{SVM}_2}(\cdot)$ introduced in (7) are available). An input multimodal biometric sample with quality vector $\mathbf{q}_T = [q_{T,1}, q_{T,2}]'$ (suppose $q_{T,1} > q_{T,2}$, otherwise interchange indexes) claims an identity and thus generates a multimodal similarity score $\mathbf{x}_T = [x_{T,1}, x_{T,2}]'$. The combined quality-based similarity score is computed as follows

$$f_{\text{SVM}_Q}(\mathbf{x}_T) = \beta f_{\text{SVM}_1}(x_{T,1}) + (1-\beta) f_{\text{SVM}}(\mathbf{x}_T) \tag{11}$$

where

$$\beta = \left(\frac{q_{T,1} - q_{T,2}}{Q_{max}} \right)^{\alpha_2} \tag{12}$$

As a result, the final fusion strategy is a quality-based trade-off between not using and using low quality traits.

3. UNIMODAL SUBSYSTEMS

3.1. Speaker Verification Subsystem

For the experiments reported in this paper, the GMM-based speaker verification system from Universidad Politecnica de Madrid used in the 2002 NIST Speaker Recognition Evaluation has been used.[18] Below we briefly describe the basics.[23]

Feature extraction. Short-time analysis of the speech signal is carried out by using 20 ms Hamming windows shifted 10 ms. For each analysis window $t \in \{1, 2, \ldots, T\}$, a feature vector $\mathbf{v}_t$ based on Mel-Frequency Cepstral Coefficients (MFCC) and including first and second order time derivative approximations is generated. Moreover, the feature vectors $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_T\}$ are supposed to be drawn from a user-dependent Gaussian Mixture Model λ which is estimated in the enrollment phase via MAP adaptation of a Universal Background Model λ_{UBM}. For our tests, the UBM is a text-independent 128 mixture GMM which was trained by using approximately 8 hours of Spanish mobile speech data (gender balanced).

Pattern comparison. Given a test utterance parameterized as V and a claimed identity modeled as λ, a matching score x'_{voice} is calculated by using the log-likelihood ratio

$$x'_{voice} = \log\left(p\left[V|\lambda\right]\right) - \log\left(p\left[V|\lambda_{UBM}\right]\right) \tag{13}$$

Score normalization. In order to generate a similarity score x_{voice} between 0 and 1, the matching score x'_{voice} is further normalized according to

$$x_{voice} = \frac{1}{1 + e^{-c_{voice} \cdot x'_{voice}}} \tag{14}$$

The parameter c_{voice} has been chosen heuristically on mobile speech data not used for the experiments reported here.

3.2. Fingerprint Verification Subsystem

For the experiments reported in this paper, a minutiae-based fingerprint verification system has been used.[17] Below we summarize the basics.[24]

Image enhancement. The fingerprint ridge structure is reconstructed according to: *i*) grayscale level normalization, *ii*) orientation field calculation,[25] *iii*) interest region extraction, *iv*) spatial-variant filtering according to the estimated orientation field, *v*) binarization, and *vi*) ridge profiling.

Feature extraction. The minutiae pattern is obtained from the binarized profiled image as follows: *i*) thinning, *ii*) removal of structure imperfections from the thinned image, and *iii*) minutiae extraction. For each detected minutia, the following parameters are stored: *a*) the x and y coordinates of the minutia, *b*) the orientation angle of the ridge containing the minutia, and *c*) the x and y coordinates of 10 samples of the ridge segment containing the minutia. An example fingerprint image from MCYT Database,[26] the resulting binary image after image enhancement, the detected minutiae superimposed on the thinned image and the resulting minutiae pattern are shown respectively in Fig. 1 from left to right.

Figure 1. Fingerprint feature extraction process.

Pattern comparison. Given a test and a reference minutiae pattern, a matching score x'_{finger} is computed. First, both patterns are aligned based on the minutia whose associated sampled ridge is most similar. The matching score is computed then by using a variant of the edit distance on polar coordinates and based on a size-adaptive tolerance box. When more than one reference minutiae pattern per client model are considered, the maximum matching score obtained by comparing the test and each reference pattern is used.

Score normalization. In order to generate a similarity score x_{finger} between 0 and 1, the matching score x'_{finger} is further normalized according to

$$x_{finger} = \tanh\left(c_{finger} \cdot x'_{finger}\right) \tag{15}$$

The parameter c_{finger} has been chosen heuristically on fingerprint data not used for the experiments reported here.

4. EXPERIMENTS

4.1. Database Description and Protocol

Cellular speech data consist of short utterances (the mobile number of each user). 75 users have been acquired, each one of them providing 10 utterance samples from 10 calls (within a month interval). The first 3 utterances

are used as voice modelling training data and the other 7 samples are used as client test data. The recordings were carried out by a dialogue-driven computer-based acquisition process, and data were not further supervised. Moreover, 10 real impostor attempts (i.e., each impostor knew the mobile number and the way it was pronounced by the user he/she was forgering) per user are used as testing data. Taking into account the automatic acquisition procedure and the highly skilled nature of the impostor data, near worst-case scenario has been prevailing in our experiments.

Fingerprint data from MCYT corpus has been used.[26] Below, some information related to the experiments we have conducted is briefly described.

MCYT fingerprint subcorpus comprises 330 individuals acquired at 4 different Spanish academic sites by using high resolution capacitive and optical capture devices. For each user, the 10 prints were acquired under different acquisition conditions and levels of control. As a result, each individual provided a total number of 240 fingerprint images to the database (10 prints $\times$ 12 samples/print $\times$ 2 sensors/sample).

Only the index fingers of the first 75 users in the database are used in the experiments. 10 print samples (optical scanner) per user are selected, 3 of them (each one from a different control level) are used as fingerprint modelling training data and the other 7 are used as testing data. We have also considered a near worst-case scenario using for each client the best 10 impostor fingerprint samples from a set of 750 different fingerprints.

All fingerprint images have been supervised and labelled according to the image quality by a human expert.[17] Basically, each different fingerprint image has been assigned a subjective quality measure from 0 (lowest quality) to 9 (highest quality) based on image factors like: incomplete fingerprint, smudge ridges or non uniform contrast, background noise, weak appearance of the ridge structure, significant breaks in the ridge structure, pores inside the ridges, etc. Fig. 4.1 shows four example images and their labelled quality.

Figure 2. Fingerprint images from MCYT corpus. Quality labelling from left to right: 0, 3, 6 and 9.

As a conclusion, data for evaluating the proposed fusion strategies consist of 75$\times$7 client and 75$\times$10 impostor bimodal attempts in a near worst-case scenario.

4.2. Multimodal Experimental Procedure

Multimodal authentication systems are usually compared with the baseline unimodal systems they consist of.[4, 5, 8, 9, 11, 12, 15] As recently pointed out in a panel discussion on multi-biometrics,[27] in order to reveal the benefits of using a number of different modalities, a more fair comparison between unimodal and multimodal systems should be based on using as many probe samples in the unimodal system as number of modalities in the multimodal scheme. We have adhered to this philosophy by combining the different probe samples (two in the experiments that follows) also with a trained SVM.

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.[20] For error estimation in multimodal authentication systems, variants of jackknife sampling using the leave-one-out principle are a common choice.[11, 28] In this work, a variant of bootstrap sampling has been used:[14, 15]

Multi-Modal 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. Errors on all bootstrap data sets all finally averaged.

Multi-Probe Fusion: For each user and considering one of the two biometric traits, 7 random matches between genuine scores and 10 random matches between impostor scores are computed (not permitting a match between a score and itself) so as to obtain 7 genuine and 10 impostor score pairs each one corresponding to two independent probe attempts. The two attempts are combined by using the same procedure described for multi-modal fusion.

For the experiments reported in the following, the problem in Eq. (5) subject to (6) has been solved by using an interior point optimization solver. $C = 100$ has been used in all tests.

Error rates are subsequently given either at the so-called Equal Error Rate (EER), i.e., the specific point attained when False Acceptance and False Rejection errors coincide,[29] or trading off False Acceptance and False Rejection errors by means of DET plots.[30]

4.3. Results

Verification performance results of unimodal and SVM-based multimodal systems without signal quality are plotted in Fig. 3 for $M = 20$ clients in the SVM training set.

In the following, the proposed quality-guided fusion scheme is studied. Regarding the quality measures, we have used the quality labels in MCYT database linearly normalized into the range $[0, 2]$ for fingerprint images. In case of voice utterances, uniform quality $q = 1$ is used in all cases.

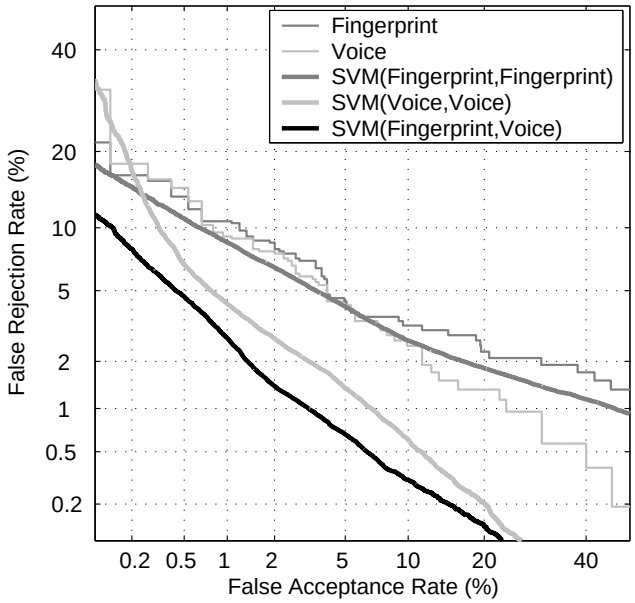
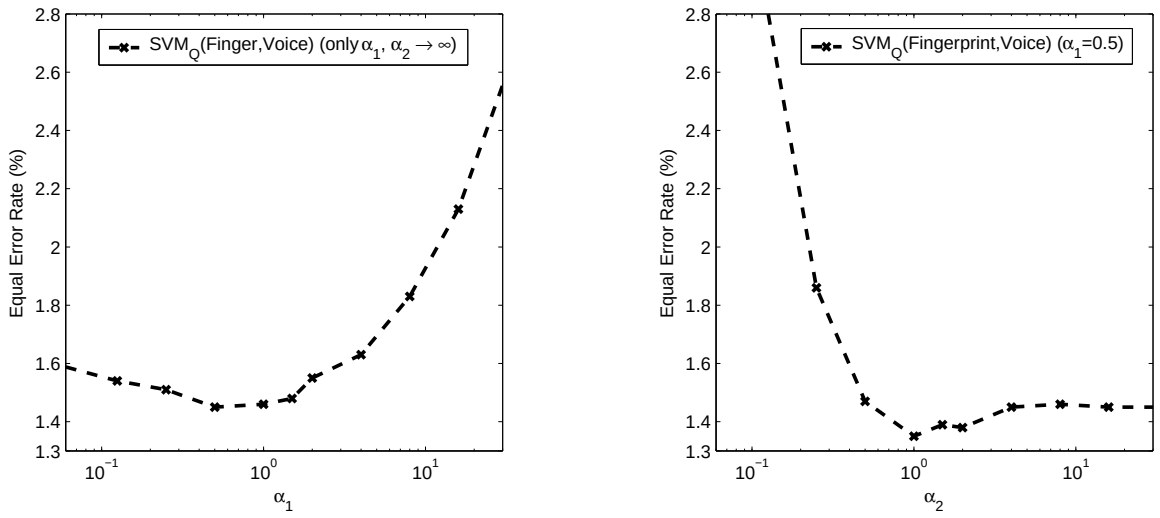


Figure 3. Verification performance of unimodal and SVM-based multimodal systems without signal quality.



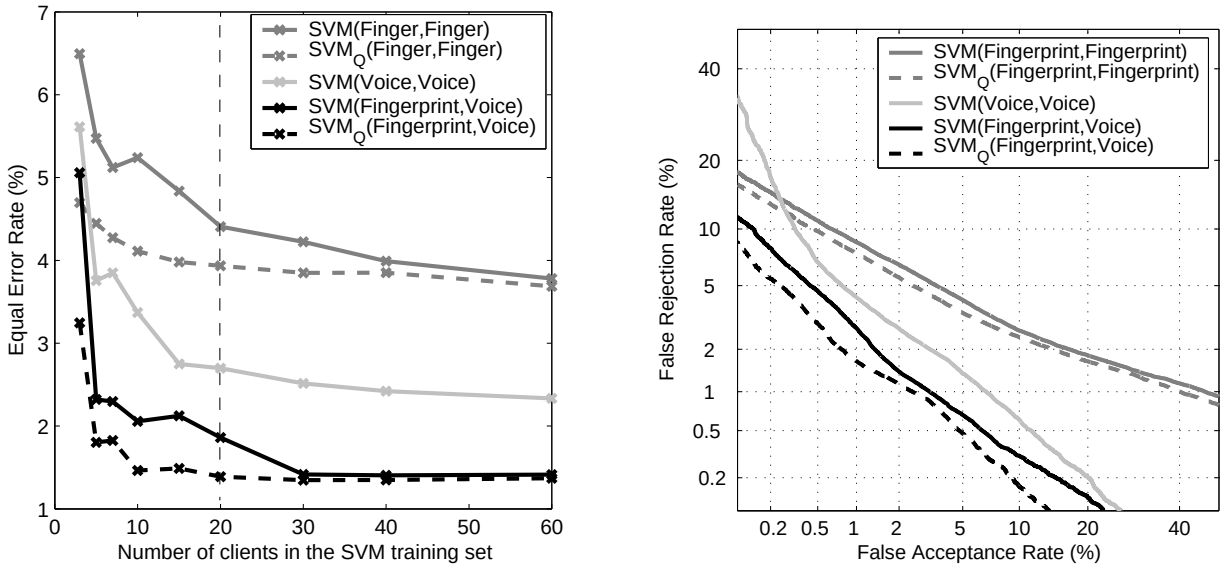
(a) (b)

Figure 4. Effects of varying parameters α_1 (a) and α_2 (b) on verification performance.

The effects on the verification performance when parameters α_1 and α_2 vary are first explored in Fig.4.

In Fig. 4 (a) verification performance of the bimodal authentication system is shown for increasing α_1 (i.e., increasing confidence on high quality multimodal training scores), while $\alpha_2 \to \infty$ so as to cancel the trade off in Eq. (11). Worth noting, a maximum of performance of 1.45% EER is obtained for $\alpha_1 = 0.5$.

In Fig. 4 (b) verification performance of the bimodal authentication system is shown for increasing α_2 (as α_2 decreases, the confidence on high quality test traits increases), while fixing $\alpha_1 = 0.5$. A maximum of performance of 1.35% EER is obtained for $\alpha_2 = 1$.



(a) (b)

Figure 5. Effects of the number of clients in SVM training set on verification performance and DET plots with (SVM_Q) and without (SVM) quality signals.

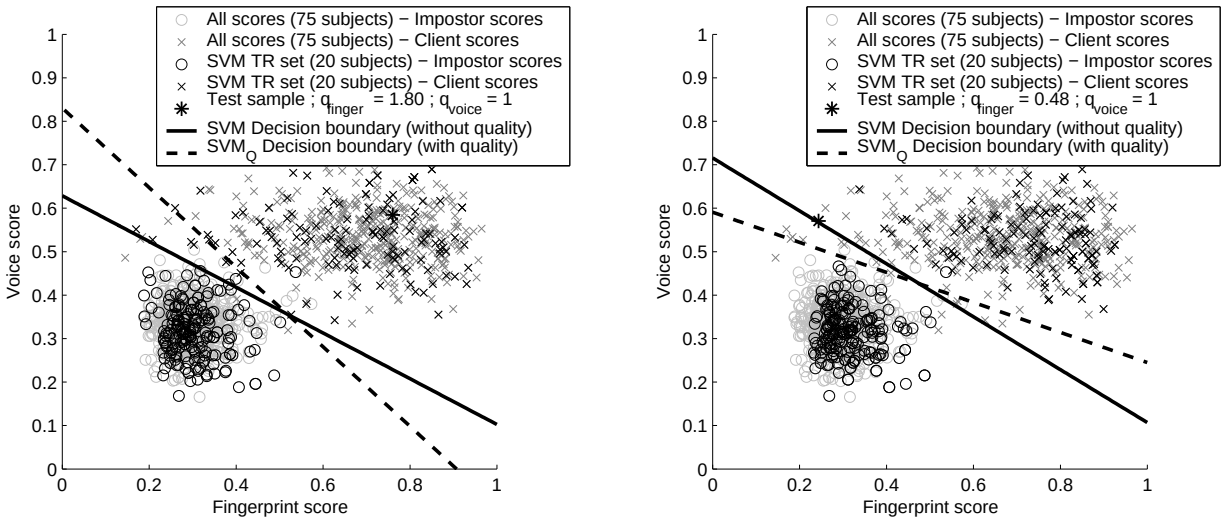


Figure 6. Training/testing scatter plot and decision boundaries for SVM fusion schemes with and without quality signals.

In the last experiment, we study the influence of increasing the number of clients M in the SVM training set over the verification performance. As it is shown in Fig. 5 (a), the error rate decreases monotonically with the number of clients in the SVM training set. In particular, a fast decay occurs for the first 10 clients (specially in the case the quality signals are used) and small improvements are obtained for more than 30 users. As can be observed, the proposed quality-based fusion scheme behaves particularly well in small training size conditions. Verification performance trade off plots with and without quality signals for $M = 20$ are given in Fig. 5 (b).

Finally, some examples that may provide an intuitive idea about how the fusion scheme is adapted depending on the image quality of the input fingerprints are shown. In particular, two different data sets of the bootstrap error estimation process are depicted in in Fig. 6 (a) and (b) respectively. User-independent decision boundaries (i.e., $f_{\mathrm{SVM}}(\mathbf{x}) = 0$ and $= f_{\mathrm{SVM}_Q}(\mathbf{x}) = 0$) have been included. In the case the score quality is considered, we observe that the SVM is adapted so as to increase or reduce the weight of the fingerprint score based on the fingerprint quality: the higher the image quality the higher the fingerprint weight and the lower the quality the lower the weight.

5. CONCLUSIONS

A kernel-based fusion scheme has been introduced. This scheme is based on SVM classifiers, trade-off coefficients introduced in the standard SVM training and testing procedures, and quality measures of the input biometric signals. The elements of a authentication application based on voice and fingerprint data have been described and some experiments using this prototype on real data have been reported.

In first place, the benefits of the combination of the two modalities are explored by using a novel experimental procedure comparing multi-modal verification performance tests with multi-probe tests of the individual subsystems. As a first result, verification performance of multi-probe individual systems (4.40% and 2.70% EER for fingerprint and voice subsystems using two probe samples) is improved by the bimodal system (1.65% EER), so the benefits are revealed. Appropriate selection of the parameters of the proposed scheme on the above prototype application leads to a quality-based fusion scheme outperforming the raw fusion strategy without considering signal quality. In particular, a relative improvement of 18% is obtained for small SVM training set size by using only fingerprint quality labels.

Future work includes the investigation of automatic quality measures for the different audio- and video-based biometric signals,[21,22] the generalization of the proposed scheme to the case of combining more than two modalities and the comparison of the reported scheme with other quality-based strategies.[15]

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry for Science and Technology under projects TIC2003-09068-C02-01 and TIC2003-08382-C05-01. J. F.-A. also thanks Consejeria de Educacion de la Comunidad de Madrid and Fondo Social Europeo for supporting his doctoral research.

REFERENCES

1. T. Kanade, "Picture processing system by computer complex and recognition of human faces," in *doctoral dissertation, Kyoto University*, November 1973.
2. B. S. Atal, "Automatic recognition of speakers from their voices," *Proceedings of the IEEE* **64**, pp. 460–475, 1976.
3. R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Trans. on Pattern Anal. and Machine Intell.* **17**(10), pp. 955–966, 1995.
4. E. S. Bigun, J. Bigun, B. Duc, and S. Fischer, "Expert conciliation for multi modal person authentication systems by bayesian statistics," in *Proc. of IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, J. Bigun, G. Chollet, and G. Borgefors, eds., pp. 291–300, Springer, 1997.
5. J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. on PAMI* **20**(3), pp. 226–239, 1998.
6. E. S. Bigun, "Risk analysis of catastrophes using experts' judgments: An empirical study on risk analysis of major civil aircraft accidents in europe," *European J. Operational Research* **87**, pp. 599–612, 1995.
7. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, Wiley, 2001.
8. S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. on Neural Networks* **10**(5), pp. 1065–1074, 1999.
9. P. Verlinde, G. Chollet, and M. Acheroy, "Multi-modal identity verification using expert fusion," *Information Fusion* **1**(1), pp. 17–33, 2000.
10. B. Gutschoven and P. Verlinde, "Multi-modal identity verification using support vector machines (SVM)," in *Proc. of the Intl. Conf. on Information Fusion, FUSION*, pp. 3–8, IEEE Press, 2000.
11. 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 IAPR Intl. Conf. on Audio- and Video-based Person Authentication, AVBPA*, pp. 830–837, Springer, 2003.
12. A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters* **24**(13), pp. 2115–2125, 2003.
13. A. K. Jain and A. Ross, "Learning user-specific parameters in a multibiometric system," in *Proc. of the IEEE Intl. Conf. on Image Processing, ICIP*, **1**, pp. 57–60, 2002.
14. J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Exploiting general knowledge in user-dependent fusion strategies for multimodal biometric verification," in *Proc. of the IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, 2004. (accepted).
15. J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Multimodal biometric authentication using quality signals in mobile communications," in *Proc. of IAPR Intl. Conf. on Image Analysis and Processing, ICIAP*, pp. 2–13, IEEE CS Press, 2003.
16. K.-A. Toh, W.-Y. Yau, E. Lim, and L. C. a C.-H. Ng, "Fusion of auxiliary information for multi-modal biometrics authentication," in *Proc. of Intl. Conf. on Biometric Authentication, ICBA*, 2004. (accepted).
17. 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 Proceedings Vision, Image and Signal Processing* **150**(6), pp. 402–408, 2003.
18. D. Garcia-Romero *et al.*, "ATVS-UPM results and presentation at NIST'2002 speaker recognition evaluation," 2002.
19. V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 2000.
20. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2003.

21. D. Garcia-Romero, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Ortega-Garcia, "On the use of quality measures for text-independent speaker recognition," in *ESCA Workshop on Speaker and Language Recognition, Odyssey*, 2004. (accepted).
22. E. Lim, X. Jiang, and W. Yau, "Fingerprint quality and validity analysis," in *Proc. of the Intl. Conf. on Image Processing, ICIP*, **1**, pp. 469–472, 2002.
23. D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital Signal Processing* **10**, pp. 19–41, 2000.
24. A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity authentication system using fingerprints," *Proceedings of the IEEE* **85**(9), pp. 1365–1388, 1997.
25. J. Bigun and G. H. Granlund, "Optimal orientation detection of linear symmetry," in *First International Conference on Computer Vision, ICCV (London)*, pp. 433–438, IEEE Computer Society Press, (Washington, DC.), June 8–11 1987.
26. J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro, "MCYT baseline corpus: A bimodal biometric database," *IEE Proceedings Vision, Image and Signal Processing* **150**(6), pp. 395–401, 2003.
27. K.-W. Bowyer, "When is multi-modal better than uni-modal in biometrics?," in *Workshop on Multimodal User Authentication*, December 2003.
28. J. Bigun, B. Duc, S. Fischer, A. Makarov, and F. Smeraldi, "Multi modal person authentication," in *Nato-Asi advanced study on face recogniton*, H. W. et. al., ed., **F-163**, pp. 26–50, Springer, 1997.
29. D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, , and A. K. Jain, "FVC2000: fingerprint verification competition," *IEEE Trans. on Pattern Anal. and Machine Intell.* **24**(3), pp. 402–412, 2002.
30. A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of decision task performance," in *Proc. of ESCA Eur. Conf. on Speech Comm. and Tech., EuroSpeech*, pp. 1895–1898, 1997.