Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 242401, 7 pages
doi:10.1155/2012/242401

Research Article

Four Machine Learning Algorithms for Biometrics Fusion: A Comparative Study

I. G. Damousis and S. Argyropoulos

Informatics and Telematics Institute, Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece

Correspondence should be addressed to I. G. Damousis, damousis@gmail.com

Received 29 October 2011; Accepted 12 January 2012

Academic Editor: Cheng-Jian Lin

We examine the efficiency of four machine learning algorithms for the fusion of several biometrics modalities to create a multimodal biometrics security system. The algorithms examined are Gaussian Mixture Models (GMMs), Artificial Neural Networks (ANNs), Fuzzy Expert Systems (FESs), and Support Vector Machines (SVMs). The fusion of biometrics leads to security systems that exhibit higher recognition rates and lower false alarms compared to unimodal biometric security systems. Supervised learning was carried out using a number of patterns from a well-known benchmark biometrics database, and the validation/testing took place with patterns from the same database which were not included in the training dataset. The comparison of the algorithms reveals that the biometrics fusion system is superior to the original unimodal systems and also other fusion schemes found in the literature.

1. Introduction

Identity verification has many real-life applications such as access control and economic or other transactions. Biometrics measure the unique physical or behavioural characteristics of an individual as a means to recognize or authenticate their identity. Common physical biometrics include fingerprints, hand or palm geometry, and retina, iris, or facial characteristics. On the other hand, behavioural characteristics include signature, voice (which also has a physical component), keystroke pattern, and gait. Although some technologies have gained more acceptance than others, it is beyond doubt that the field of access control and biometrics as a whole shows great potential for use in end user segments, such as airports, stadiums, defence installations, but also the industry and corporate workplaces where security and privacy are required.

Biometrics may be used for identity establishment. A new measurement that purports to belong to a particular entity is compared against the data stored in relation to that entity. If the measurements match, the assertion that the person is whom they say they are is regarded as being authenticated. Some building access schemes work in this way, with the system comparing the new measure against the company's employee database. Also authentication of the identity of a person is frequently used in order to grand access to premises or data.

Since authentication takes place instantaneously and usually only once, identity fraud is possible. An attacker can bypass the biometrics authentication system and continue undisturbed. A cracked or stolen biometric system presents a difficult problem. Unlike passwords or smart cards, which can be changed or reissued, absent serious medical intervention, a fingerprint or iris is forever. Once an attacker has successfully forged those characteristics, the end user must be excluded from the system entirely, raising the possibility of enormous security risks and/or reimplementation costs. Static physical characteristics can be digitally duplicated, for example, the face could be copied using a photograph, a voice-print using a voice recording, and the fingerprint using various forging methods. In addition static biometrics could be intolerant of changes in physiology such as daily voice changes or appearance changes.

Unimodal biometric systems have to contend with a variety of problems such as noisy data, intraclass variations, restricted degrees of freedom, non-universality, spoof

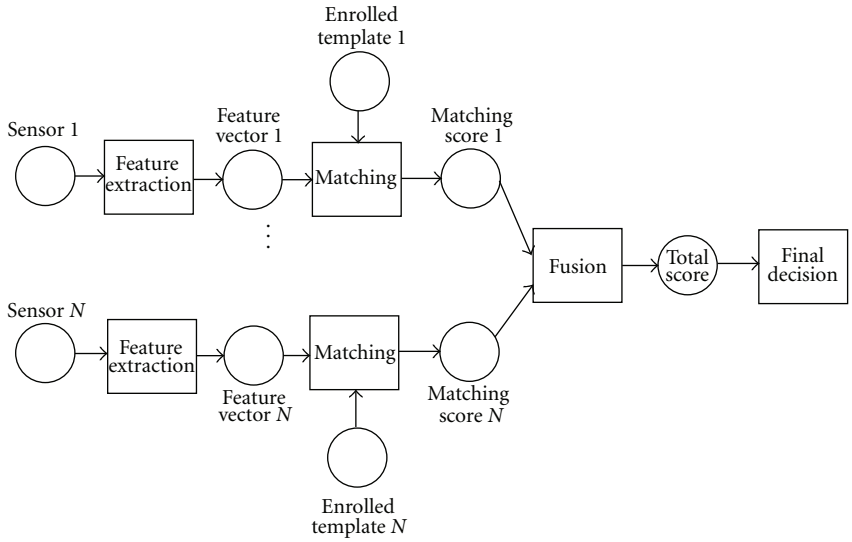

Figure 1: Fusion at the matching score level.

attacks, and unacceptable error rates. Some of these limitations can be addressed by deploying multimodal biometric systems [1, 2] that integrate the evidence presented by multiple sources of information. Indeed, the development of systems that integrate two or more biometrics is emerging as a trend. Such systems, known as multimodal biometric systems, are expected to be more reliable due to the presence of multiple, (fairly) independent pieces of evidence. These systems are able to meet the stringent performance requirements imposed by various applications. They address the problem of nonuniversality, since multiple traits ensure sufficient population coverage. They also deter spoofing since it would be difficult for an impostor to spoof multiple biometric traits of a genuine user simultaneously. Furthermore, they can facilitate a challenge response type of mechanism by requesting the user to present a random subset of biometric traits thereby ensuring that a "live" user is indeed present at the point of data acquisition.

In the present paper we present the development and evaluation of four machine learning algorithms for the fusion of the similarity scores of several biometric experts to form a multimodal biometrics system, aiming to raise significantly the recognition accuracy and reduce the false acceptance and false rejection rates. The supervised algorithms are trained using samples from a well known biometrics database and then validated using samples from the same database that are different from the training ones. The aim is to compare the developed algorithms with existing techniques and also find the most efficient one out of the four, in order to use it for the fusion of novel biometrics within project HUMABIO [3].

2. Supervised Fusion Algorithms

Fusion at matching score level is the most common approach of fusion in multimodal biometric systems [2]. This fact is mainly due to the easy accessibility and availability of the matching scores in many biometric modules. The input for a fusion algorithm at matching score level is the (dis-) similarity score provided by a biometric module (Figure 1). There are different approaches of merging scores at the matching score level. In this approach, output scores of the individual matching algorithms constitute the components of a multidimensional vector for example, a 3D vector is created if scores from a face, gait, and voice matching module are available. The resulting multi-dimensional vector is then classified using a classification algorithm such as Support Vector Machines (SVM), Fuzzy Expert systems (FES), neural networks, and so forth, to solve the two class classification problem of classifying the output vector into either "impostor" (unauthorized users) or "genuine" (authorized users, or *clients*) class. One advantage of the approach is the fact that scores may be inhomogeneous such as a mix of similarities and distances possibly located in different intervals. Thus, no pre-processing is required for classification fusion.

In supervised learning, the learning algorithm is provided a training set of patterns (or inputs) with associated labels (or outputs). Usually, the patterns are in the form of attribute vectors and once the attribute vectors are available, machine-learning methods can be applied, ranging from simple Boolean operators, to Bayesian classification and more sophisticated methods. The performance of a fusion algorithm relies on the tuning of the system. This tuning usually consists in a group of hyper-parameters that can be set manually (such as type of kernel in SVMs, number of chromosomes in Genetic Algorithms (GA), etc.) and another group that is set during the training phase. The training is used so that the algorithm can estimate (learn) the client and impostor spaces and is crucial for the performance of the fusion system.

In this study four state of the art fusion techniques were utilized, namely Support Vector Machines, Fuzzy Expert Systems, Gaussian Mixture Models and Artificial Neural

Networks. Each of these techniques follows a different philosophy for the fusion of the unimodal biometric inputs in order to produce an overall estimation of whether the person is a client or an impostor.

2.1. Support Vector Machine. A typical SVM implementation was developed [4]. A radial basis kernel function was used to map the input data to a higher dimensional space, in which they were linearly separable [5]. The radial basis kernel (RBF) was selected in order to handle probable nonlinearities between the input vectors and their corresponding class. It also has less hyperparameters than the polynomial kernel thus making training easier.

After the selection of the kernel, the process followed consists of identifying the best pair of C and γ, that is, the pair with the best cross-validation accuracy. Following the guidelines found in [6], the training set was divided into equal sized subsets and one subset was the validation dataset using the classifier that was trained on the remaining subsets. The process was repeated sequentially until all subsets acted a validation dataset. The selection of C and γ was done via grid search, that is, trying exponentially growing sequences of C and γ [7]. The penalty parameters for each of the two classes ("Genuine", "Impostor") was done via complete enumeration trials.

The final trained SVM model was then used with the optimal C, γ pair in order to check the classifier performance on the test dataset which is comprised of "unknown" patterns that were not used for the SVM training.

2.2. Fuzzy Expert System. A TSK FES [8] was developed as described in [9]. The FES's premise space consisted of three inputs ($NPI = 3$). Each premise input was segmented by three trapezoid membership functions described by the following equation:

$$\mu_i^j\left(x_{p,i}\right) = \max\left(\min\left(\frac{x_{p,i} - a_{i,j}}{b_{i,j} - a_{i,j}}, 1, \frac{d_{i,j} - x_{p,i}}{d_{i,j} - c_{i,j}}\right), 0\right), \quad (1)$$

where the parameters $a_{i,j}$ and $d_{i,j}$ locate the "feet" of the "jth" trapezoid of the "ith" premise input and the parameters $b_{i,j}$ and $c_{i,j}$ locate the "shoulders".

This segmentation leads to the creation of 27 three-dimensional fuzzy rules (Figure 2).

The firing strength of the rule $R(j)$, representing the degree to which $R(j)$ is excited by a particular premise input vector $\overline{X}_p$, is determined by

$$\mu_j\left(\overline{X}_p\right) = \prod_{i=1}^{\mathrm{NPI}} \mu_i^j\left(x_{p,i}\right). \quad (2)$$

The premise inputs were selected via extensive experimentation from the available biometric experts (shown in Table 1). Each fuzzy rule produces an output which is a linear function of the unimodal classifiers' scores $x_{c,i}$ shown in

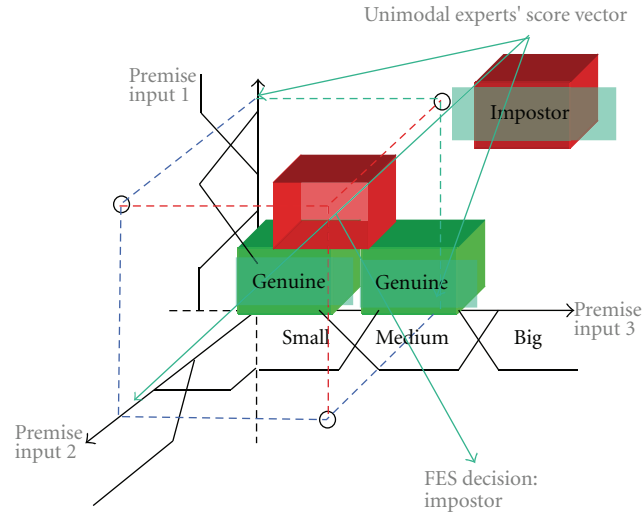

Figure 2: Three membership functions with linguistic expressions "low", "medium", "high" (score) are used for the partitioning of three premise inputs, leading to the formation of 27 fuzzy rules.

Table 1: EER of the unimodal experts in the XM2VTS database.

Expert	EER (%)
1 (Face)	1.83
2 (Face)	4.13
3 (Face)	1.78
4 (Face)	3.50
5 (Face)	6.50
6 (Voice)	1.09
7 (Voice)	6.50
8 (Voice)	4.50

Table 1 so as to include all of the available information from the unimodal biometric experts:

$$y_j = F\left(\overline{X}_c\right) = \lambda_0^j + \sum_{i=1}^{\mathrm{NI}} \lambda_i^j\, x_{c,i} + \sum_{k=1}^{\mathrm{NI}} \lambda_k^j \frac{1}{x_{c,i}}, \quad (3)$$

where NI is the number of unimodal classifiers.

The FES estimation of the client's authenticity is a synthesis (weighted average) of the 27 fuzzy rule outputs:

$$y = \frac{\sum_{j=1}^{\mathrm{NR}} \mu_j\left(\overline{X}_p\right) \cdot F_j\left(\overline{X}_c\right)}{\sum_{j=1}^{\mathrm{NR}} \mu_j\left(\overline{X}_p\right)}. \quad (4)$$

This estimation is compared to a threshold T and if it is higher than the FES classifies the pattern as genuine transaction while if it is lower it classifies it as impostor.

The parameters λ of the fuzzy rules' linear output functions (3), the parameters of the trapezoid membership functions (1) that define the position of the fuzzy rules in the premise space (or in other words the segmentation of the premise inputs via the shape and positioning of the membership functions), and also the decision threshold T are optimized by a real coded [9] genetic algorithm [10].

The GA fitness function was selected so as to minimize the false acceptance of impostors and maximize correct authentication rate through the evolution of the GA:

$$\text{Fitness function} = \frac{\text{correct\_out}}{\text{error\_out} + \text{false\_normal} + 1}, \quad (5)$$

where correct_out is the accumulated distance of the fuzzy output from the decision threshold in case of correct authentication (genuine or impostor), error_out is accumulated distance of the fuzzy output from the threshold in case of incorrect authentication for all training patterns, and false_normal is the number of falsely accepted impostor individuals.

In that way the solutions that have minimum false acceptance occurrences have higher fitness value. The same stands for solutions (chromosomes) that produce outputs that have larger distance from the threshold T in correct authentications (more robust solutions) and smaller distance from the threshold in case of erroneous authentications (thus driving wrong solutions towards the threshold and rectifying them).

2.3. Gaussian Mixture Model. Bayesian classification and decision making is based on probability theory and the principle of choosing the most probable or the lowest risk (expected cost) option. The Gaussian distribution is usually quite good approximation for a class model shape in a suitably selected feature space. In a Gaussian distribution lies an assumption that the class model is truly a model of one basic class. However, if the actual model is multimodal, this model cannot capture coherently the underlying distribution. Gaussian Mixture Model (GMM) is a mixture of several Gaussian distributions and can therefore represent different subclasses within a class [11]. The probability density function is defined as a weighted sum of Gaussians:

$$P(x;\theta) = \sum_{c=1}^{C} \alpha_c N(x;\mu_c,\Sigma_c), \quad (6)$$

where α_c is the weight of component c, $0 \le \alpha_c \le 1$ and

$$\sum_{c=1}^{C} \alpha_c = 1. \quad (7)$$

The parameter list $\theta = \{\alpha_1,\mu_1,\Sigma_1,\ldots,\alpha_c,\mu_c,\Sigma_c\}$ defines a particular Gaussian mixture probability density function. Estimation of the Gaussian mixture parameters for one class can be considered as unsupervised learning of the case where the samples are generated by individual components of the mixture distribution and without the knowledge of which sample was generated by which component.

A GMM was developed, which comprised of four mixture components. The weights of the components were estimated after extensive experimentation.

2.4. Artificial Neural Network. The fourth algorithm was a three-layer feed-forward neural network (NN) [12]. The layers consist of N input neurons, Y hidden neurons, and one output neuron where N is equal to the number of unimodal biometric experts (from Table 1, $N = 8$) and Y is set through experimentation equal to ten. The neurons are fully interconnected and a bias is applied on each neuron. The transfer function is selected to be sigmoid so as to address nonlinearities of the input data set. For the training of the weights, the typical back propagation method was used. The optimum number of training iterations and training parameters was set heuristically. Convergence was achieved after 500 iterations.

3. Benchmark Database

The developed fusion schemes were tested on the publicly available XM2VTS face and speech database [13, 14]. The XM2VTS database contains facial and speech data from 295 subjects, recorded during four sessions taken at one-month intervals. It includes similarity scores from five face experts and three speech experts. The protocol consists of two sets: the development (training) set and the evaluation (validation) set. The development set, which is used for training, contains scores from three multimodal recordings for each of 200 client users and eight transactions from each of the 25 impostors. The evaluation set, which is used for testing the system, contains scores from two multimodal recordings of the (same) 200 client users and eight transactions for each of the 70 (new) impostors. The impostors in the evaluation set are different from the impostors in the development. Thus, the development set contains 600 (200×3) client transactions and 40000 ($25 \times 200 \times 8$) impostor transactions whereas the evaluation set contains 400 (200×2) client transactions and 112000 ($70 \times 200 \times 8$) impostor transactions.

Two metrics are computed: the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) defined as

$$\text{FAR} = \frac{\text{number of accepted impostors}}{\text{number of impostor transactions}},$$
$$\text{FRR} = \frac{\text{number of rejected clients}}{\text{number of client transactions}}. \quad (8)$$

The Half Total Error Rate (HTER) is also reported which is defined as

$$\text{HTER} = \frac{\text{FAR} + \text{FRR}}{2}. \quad (9)$$

The Equal Error Rate (EER) is computed as the point where FRR = FAR; in practice, FRR and FAR are not continuous functions and a crossover point might not exist. In this case, the interval [EERlo, EERhi] should be reported. Also, another useful tool for the evaluation of the performance of a biometric system is the Rate Operating Characteristic (ROC) curve, which is produced by plotting FAR versus FRR.

The EERs of the unimodal experts for the XM2VTS database are shown in Table 1.

Figure 3 shows indicative FAR-FRR diagrams for two face and voice unimodal biometrics experts of the XM2VTS database. It can be seen from the threshold T range that the expert scores that characterize someone as impostor or genuine differ significantly. However, this does not pose a problem for the fusion algorithms.

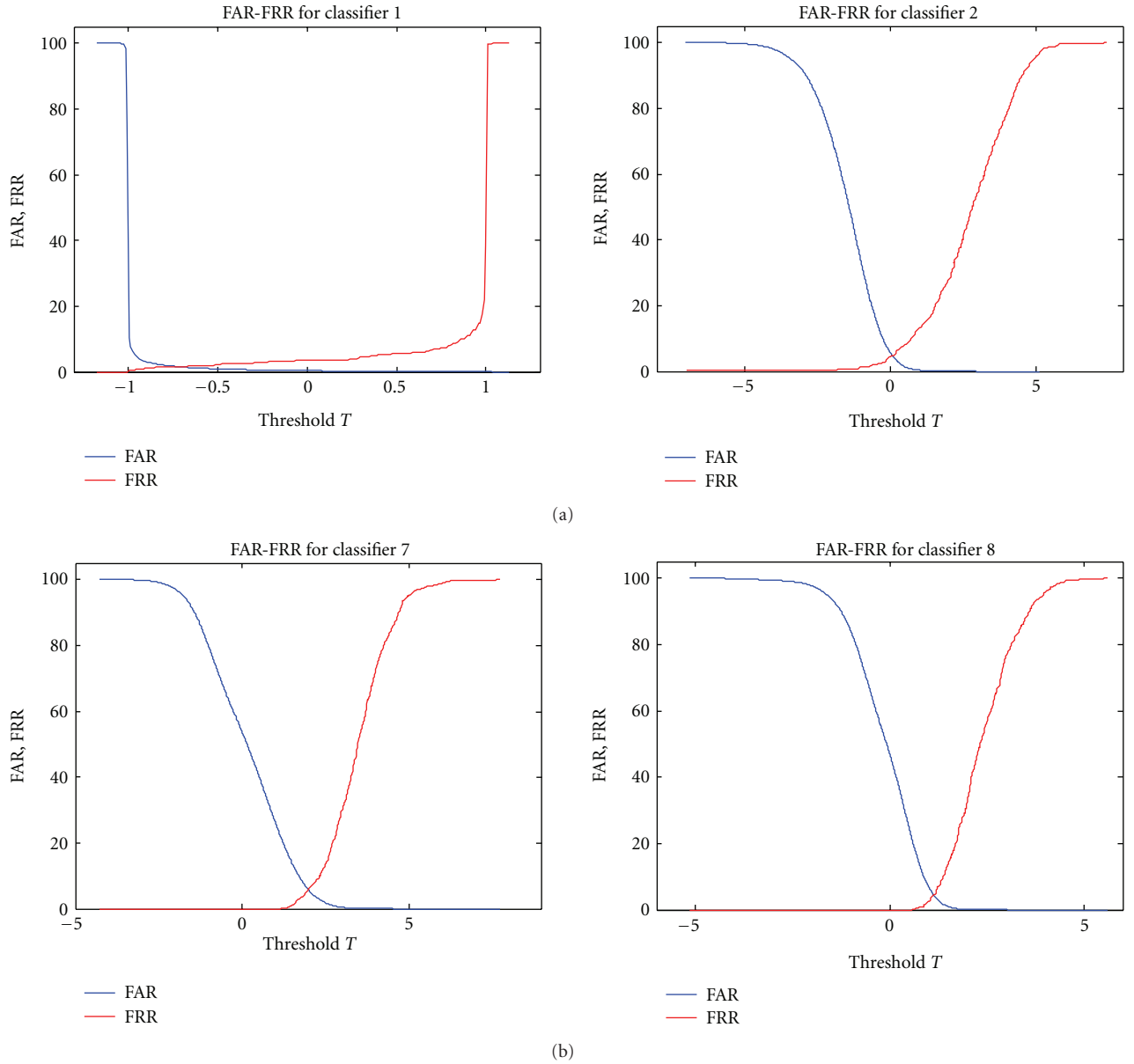

Figure 3: FAR-FRR diagrams and EER for (a) two face and (b) two voice biometrics experts of the XM2VTS database.

Table 2: Summary table of verification results of different fusion methods for the XM2VTS database.

Fusion method	Training set			Test set		
	FAR (%)	FRR (%)	HTER (%)	FAR (%)	FRR (%)	HTER (%)
SVM	0	0	0	0.0	0.5	**0.25**
FES	0	0.005	0.0025	0.15	0.75	**0.45**
NN	0.0025	0.33	0.17	0.0	1	**0.5**
GMM	0.0225	0.33	0.18	0.028	1	**0.51**
OR	12.63	0	6.31	19.46	0.0	**9.73**
AND	0.22	12.83	6.52	0.0	19.75	**9.87**

Table 3: Performance comparison of fusion schemes for the XM2VTS database.

Fusion scheme	HTER (%)
Product	3.50
Max	2.35
Min	1.13
Dempster-Shafer	0.76
Sum	0.75
Weighted sum	0.63
SVM	0.52
SumPro	0.31
Presented SVM	**0.25**

4. Test Results

4.1. Comparative Results. Table 2 summarizes the results of the investigated machine learning algorithms for multimodal fusion. More specifically, the classification of the XM2VTS patterns was performed using the SVM, GMM, FES, and NN fusion schemes. Furthermore, some simple combination rules (AND, OR) were also tested. The first conclusion we can reach from the results illustrated in the table is that all of the fusion schemes perform better than the best performing unimodal expert (i.e., expert 6 with EER 1.09%) and the classification absolute improvement using the SVM fusion method is 0.84% (SVM HTER = 0.25%, ~77% relative improvement over unimodal expert 6 classification error). This corroborates the statement that the effective combination of information coming from different experts can improve significantly the performance of a biometric system. Moreover, this table also confirms the superiority of the SVM fusion scheme over the other machine learning techniques. More specifically, the FAR and FRR using the SVM fusion classifier on the XM2VTS database are 0.0% and 0.5%, respectively. The superior performance of the SVM fusion classifier over the second best FES, which is 0.2%, is mainly attributed to the more efficient modelling of the feature space. Moreover, the SVM fusion expert performs very satisfactory when the number of the feature vectors (comprised of the matching scores in this case) is relatively large, as in the case of the XM2VTS database, where there are 8 unimodal experts.

4.2. Comparison with State-of-the-Art Methods. An experimental study was also conducted to compare the developed schemes with state-of-the-art fusion methods on the XM2VTS database. The same unimodal experts and the same protocol were employed so that any performance gain or decrease can be attributed only to the fusion algorithm. The following table illustrates the HTER values of various fusion schemes, as reported in [15].

The first conclusion from Table 3 is that the reported results are inferior compared to the results presented in the previous section. Specifically, the best result, in terms of HTER, is 0.31%, whereas the best result of the algorithms presented in this paper is 0.25%. Moreover, it can be seen that the accuracy achieved in [15] for SVM classification is lower than the authentication accuracy produced by our implementation of the SVM fusion scheme. This could be attributed partly to the different implementation of the algorithm. The comparison results validate the superiority of the developed SVM scheme and indicate its appropriateness for the application scenarios examined within HUMABIO [16, 17].

5. Conclusions

Four machine learning algorithms were developed for the fusion of several biometric modalities in order to detect the most efficient one for use within the project HUMABIO. The algorithms were Gaussian Mixture Models, an Artificial Neural Network, a Fuzzy Expert System, and Support Vector Machines. The algorithms were trained and tested using a well-known biometric database which contains samples of face and speech and similarity scores of five face and three speech biometric experts. The fusion results were compared against existing fusion techniques and also against each other, showing that the fusion schemes presented in this paper produce better biometric accuracy from conventional methods. From the four algorithms, the most efficient one proved to be the support vector machines-based one offering significant performance enhancement over unimodal biometrics, over more traditionally combined multimodal biometrics, but also over the SoA.

Acknowledgments

This work was supported in part by the EC under Contract FP6-026990 HUMABIO [3, 16]. I. G. Damousis, when the research reported in this paper took place, was with the Informatics and Telematics Institute of the Centre for Research and Technology Hellas in Thessaloniki, Greece (e-mail: damousis@gmail.gr). S. Argyropoulos, when the research reported in this paper took place, was with the Informatics and Telematics Institute of the Centre for Research and Technology Hellas in Thessaloniki, Greece (e-mail: savvas@ieee.org).

References

[1] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[2] A. K. Jain and A. Ross, "Multibiometric systems," *Communications of the ACM*, vol. 47, no. 1, pp. 34–40, 2004.

[3] I. G. Damousis, D. Tzovaras, and E. Bekiaris, "Unobtrusive multimodal biometric authentication: the HUMABIO project concept," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, Article ID 265767, 11 pages, 2008.

[4] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[5] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2000.

[6] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," *Test*, vol. 1, no. 1, pp. 1–16, 2010.

[7] R. E. Fan, P. H. Chen, and C. J. Lin, "Working set selection using second order information for training support vector machines," *Journal of Machine Learning Research*, vol. 6, pp. 1889–1918, 2005.

[8] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[9] I. G. Damousis and D. Tzovaras, "Fuzzy fusion of eyelid activity indicators for hypovigilance-related accident prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, pp. 491–500, 2008.

[10] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York, NY, USA, 1989.

[11] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.

[12] T. Masters, *Signal and Image Processing with Neural Networks*, John Wiley & Sons, 1994.

[13] N. Poh and S. Bengio, "A score-level fusion benchmark database for biometric authentication," in *Proceedings of the 5th International Conference on Audio, and Video-Based Biometric Person Authentication (AVBPA '05)*, vol. 3546 of *Lecture Notes in Computer Science*, pp. 1059–1070, July 2005.

[14] XM2VTS database, http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/.

[15] L. Shoushan and Z. Chengqing, "Classifier combining rules under independence assumptions," in *Proceedings of the 7th international Conference on Multiple Classifier Systems (MCS '07)*, vol. 4472 of *lecture Notes in Computer Science*, pp. 322–332, 2007.

[16] HUMABIO project, http://www.humabio-eu.org/.

[17] A. Vatakis et al., "Deliverable 7.1 HUMABIO Pilot plans," 2008.