Volume 89 Issue 11 November 2009 ISSN 0165-1684

SIGNAL PROCESSING

An International Journal

A publication of the European Association for Signal Processing (EURASIP)

Signal Processing 89 (2009) 2117–2127

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Combining different biometric traits with one-class classification

C. Bergamini [a], L.S. Oliveira [b,*], A.L. Koerich [a], R. Sabourin [c]

[a] Pontifical Catholic University of Parana (PUCPR), R. Imaculada Conceição, 1155, Curitiba, PR 80215-901, Brazil
[b] Federal University of Parana (UFPR), Department of Informatics, Rua Cel. Francisco Heráclito dos Santos, 100, Curitiba, PR, Brazil
[c] Ecole de Technologie Superieure, 1100 rue Notre Dame Ouest, Montreal, Quebec, Canada

ARTICLE INFO

Article history:
Received 18 September 2008
Received in revised form
13 April 2009
Accepted 20 April 2009
Available online 18 May 2009

Keywords:
Biometric fusion
One-class classification
SVM

ABSTRACT

It has been demonstrated in the literature that the combining of different biometric traits is a powerful tool to overcome the limitations imposed by a single biometric system. The fusion of different systems can be approached in different ways. In this work, we consider the pattern classification approach, where the scores of the various systems are used as features to feed the classifiers. More specifically, we are interested in one-class classifiers, and we show that one-class classification could be considered as an alternative to biometric fusion, especially when the data are highly unbalanced or when data from only a single class are available. The results reported for one-class classification on two different databases compares with the standard two-class SVM and surpasses all the conventional classifier combination rules tested.

1. Introduction

With to the impressive progress in information technology, the search is on for more robust and reliable authentication methods for a wide variety of applications. Biometrics has emerged as potentially providing a solution to deal with this kind of problem, because biometric data cannot be easily stolen or shared. Consequently, they have been used successfully in a wide variety of applications.

However, biometric systems are far from perfect [12,13]. For example, in the case of fingerprint verification, it is known that a small fraction of the population may be unsuited for the automatic identification because of genetic factors, or aging, or for occupational reasons. Face recognition imposes several restrictions on how the facial images are obtained, quite often requiring a fixed background and controlled illumination. Signature recognition requires contact with the writing instrument and effort on the part of the user. In summary, every biometric system has its limitations.

To overcome the constraints imposed by a single biometric system, several researchers have investigated the use of multiple sensors to capture different biometric traits. This field of research is known as multimodal biometrics [2,8,20]. The fusion of the various traits can be achieved at the feature extraction, matching score, or decision level. Fusion at matching score level has the advantage of using as much information as possible from each biometric modality, while at the same time enabling the integration of off-the-shelf biometric systems [22]. It is worth noting that a normalization step is generally necessary before combining scores from different matchers.

There are three types of multimodal biometric system. The first type is known as transformation-based. Here the matching scores are normalized (transformed) to place them on a comparable scale. There are several ways to implement the fusion of different matchers, such as sum, product, max, mean, weighted sum, etc. Besides choosing the best fusion strategy, the use of weights to indicate the importance of the matching scores provided by each

[*] Corresponding author at: Federal University of Parana (UFPR), Department of Informatics, Rua Cel. Francicsco H. dos Santos, 100, Curitiba, PR 81531-980, Brazil
E-mail address: lesoliveira@inf.ufpr.br (L.S. Oliveira).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.04.043

biometric trait should also be considered [9]. The second type of fusion is called density-based, and it relies on the estimation of the joint densities of the matching scores, and the fusion is carried out by statistical tests, such as the likelihood ratio test [15]. This type of fusion scheme achieves good performance if the densities can be well learned, given that a large number of representative training matching scores are available. The third strategy is classifier based. With this strategy the scores produced by each biometric system are considered as features to feed a classifier. In such a case, each input pattern should be labeled as either genuine or an impostor. A review of the literature reveals several works using classification strategies such as neural networks, *k*-NN, quadratic classifiers, and support vector machines (SVM) [29,19,6,11,5].

To the best of our knowledge, where the pattern classification approach to multimodal biometrics is used, the classification problem is considered as a two-class problem. The limitation of this strategy is that very often the amount of data available for training is not sufficient and not representative enough to guarantee good parameter estimation and generalization capabilities [4]. In some cases, only the patterns from one class (genuine or impostor) are available or the data are seriously unbalanced. In other cases, only samples of the target class can be used to build a model. The boundary between the two classes has to be estimated from the data of only one class (genuine or impostor). In other words, the task consists of defining a boundary around the target class, such that it accepts as many of the target samples as possible, while minimizing the chance of accepting outliers. This is known in the literature as one-class classification.

In this paper, we discuss combining different biometric matchers to improve the accuracy, efficiency, robustness, and fault tolerance of biometric systems. Our focus is to demonstrate that one-class classification can be considered as an alternative to combining different systems when only data from one class are available for training. We have carried out several experiments to better assess the results. First, we apply conventional combination rules such as sum, product, mean, etc. Then we use a pattern classification approach with the standard two-class SVM to combine the biometric systems.

The last part of the experiments is devoted to the pattern classification approach using a one-class SVM. We demonstrate that one-class classification can perform very well for biometric fusion when only data from one class (e.g. impostor) are available for training or the datasets are unbalanced. Two different databases, freely available for research purposes, were considered in our experiments. The first one is the biometric scores set, release 1 (NIST BSSR1) database [16] and the second is the MCYT scores database [17]. The results reported in this paper compare to the conventional two-class SVMs and surpass all the combination rules reported in the literature.

The remainder of this paper is organized as follows. Section 2 describes the databases used. Section 3 briefly reviews the one-class methods and describes the basics of the one-class SVM. Section 4 reports all the experiments that were carried out. Our conclusions are presented in the last section.

2. Databases

As mentioned previously, two databases were considered in this work. The first is the NIST biometric scores set, release 1 [16], which contains face and fingerprint matching scores. The database was built with the face and fingerprint data of 517 subjects. Two face matchers (C and G), and one fingerprint matcher (V) were used to produce the scores. Therefore, it contains four 517 × 517 similarity matrices: right index fingerprints scored by matcher V (R), left index fingerprints scored by matcher V (L), frontal face images scored by matcher C (C), and frontal face images scored by matcher G (G). Each similarity matrix contains 517 genuine scores and 266,772 (517 × 516) impostor scores. This database is depicted in Fig. 1.

The receiver operating characteristics (ROC) curves depicted in Fig. 2a show the baseline performance for the BSSR1 data. To better compare the results of the four matchers and make further combinations possible, a comparison criterion should be defined. One of the criteria most often used when considering the ROC is the area under the curve (AUC). The bigger the AUC, the better the system. In a perfect system, AUC = 1.0. In the context of the BSSR1 Set 1, the best performance is achieved by system C (face with matcher C), for which AUC = 0.989, as against 0.982, 0.981, and 0.962 from G, R, and L, respectively.

Another way to compare systems using the ROC is to define an FAR (e.g. 10^{-4}), and verify which system provides the best genuine acceptance rate (GAR). This is a common request of biometric systems, since they usually require very low FARs. In this case, it is easy to

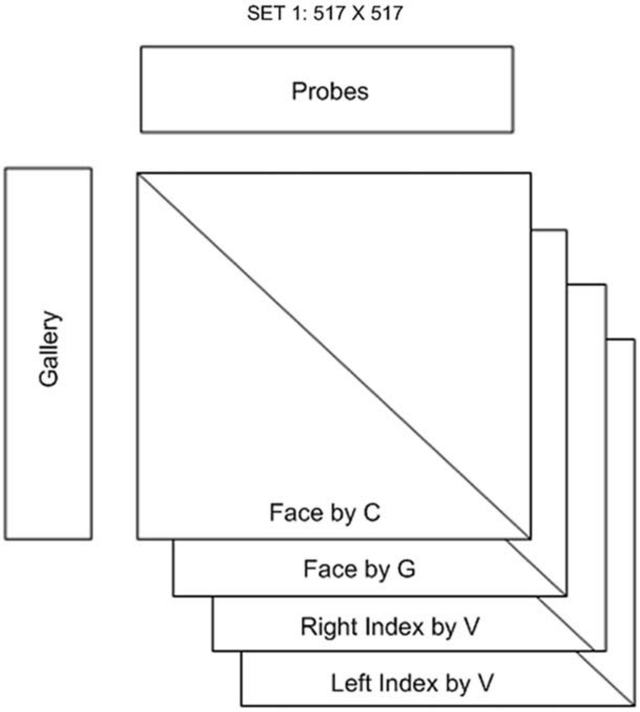

Fig. 1. Structure of the BSSR1 Set 1. Diagonal elements represent genuine (true mates) and the off-diagonal elements represent the impostor (true non-mates).

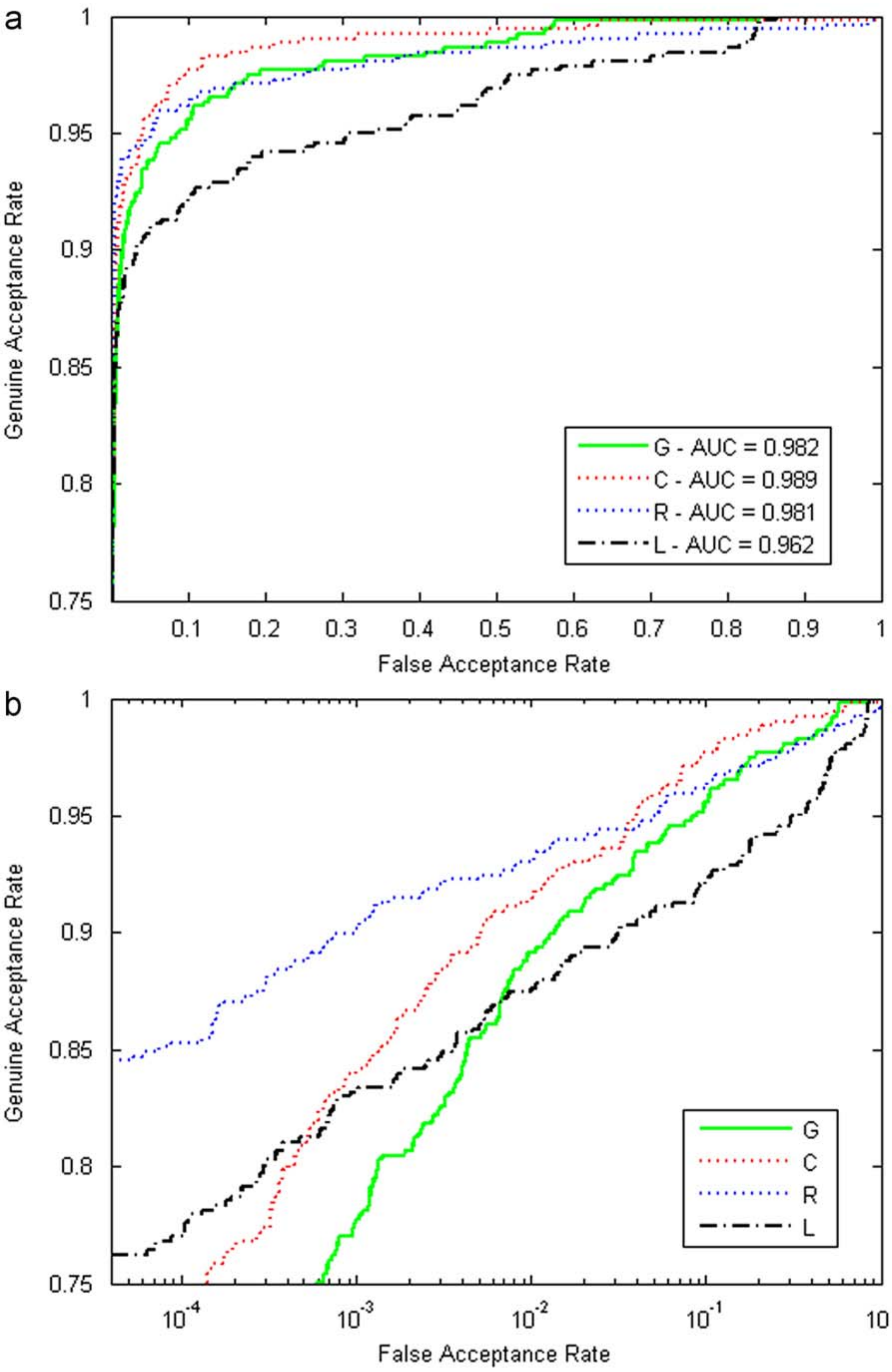

Fig. 2. Baseline performance for BSSR1 data on two different scales.

observe from Fig. 2a that system R (right fingerprint) outperforms all the others by producing a GAR of 0.85 for the FAR fixed at 10^{-4}. As we can see, even the system with the lower AUC (system L) performs better than C and G using this criterion.

The second database used in this work is the MCYT scores database [17,4], which contains a dataset of bimodal matching scores (fingerprint and signature). The database was built with the fingerprint (N and Q) and signature (S) data of 75 subjects from the MCYT database, together with scalar fingerprint quality measures labeled by a human expert. Therefore, for each subject, it contains seven genuine scores and 10 impostor scores. Thus, each file contains 525 (75 × 7) genuine scores and 750 (75 × 10) impostor scores, for a total of 1275 scores. Fig. 3 depicts the performance of all the biometric systems in the MCYT database. The best performance is achieved by system S (signature), which has an AUC = 0.989, as against 0.979 and 0.628 from N and Q, respectively. Although this work deals with the normal MCYT database

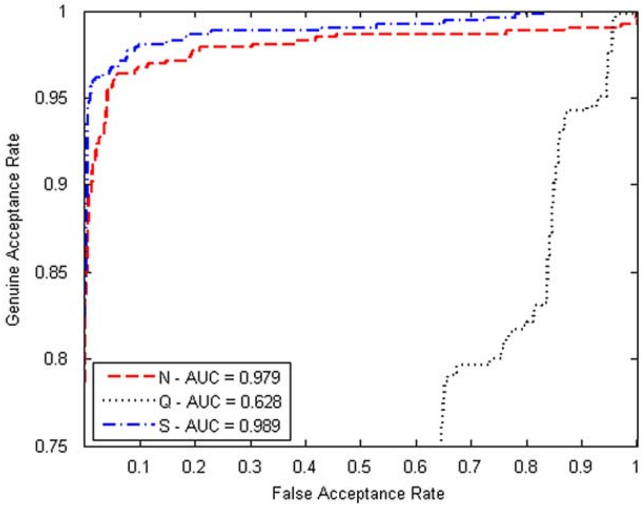

Fig. 3. Baseline performance for MCYT database.

(all high-quality images), this database also contains two other databases with 95% and 90% high-quality images from the normal database, respectively.

The MCYT database provides only normalized scores [4]. The fingerprint scores were normalized according to the following equation:

$$x_{finger} = \tanh(c_{finger} \times x'_{finger}) \tag{1}$$

where c_{finger} is a parameter defined empirically on a different dataset and x'_{finger} is the raw score produced by the fingerprint system. The signature scores were normalized according to the following equation:

$$x_{sig} = \exp(c_{sig} \times x'_{sig}) \tag{2}$$

where c_{sig} is a parameter defined empirically on a different dataset and x'_{sig} is the raw score produced by the signature verification system. The parameters c_{finger} and c_{sig} as well as the datasets used to define them are not reported in the references.

3. One-class classification

The problem with one-class classification lies in describing a target set of objects and detecting which new objects resemble objects in this training set. The problem with one-class classification is that, unlike conventional binary classification, information is only available from a single class. The objects in this class are called target objects, and so all the others are outlier objects.

Several different terms have been used to refer to one-class classification, such as outlier detection, novelty detection, and concept learning [7,25]. These various terms are usually used to represent the various problems with one-class classification. The most common of these is outlier detection, which consists of detecting those samples that do not resemble the bulk of the dataset in some way. One-class classification is also indicated for those problems where one of the classes is sampled very well, while the other class is seriously undersampled. This often happens when the measurements on the undersampled class might be very expensive or difficult to obtain.

Several models have been proposed in the literature for one-class classification. Most can be categorized into one of the two groups: density approaches and boundary approaches. As the name implies, a density approach uses a density method to directly estimate the density of the target objects [1,18,23]. In the testing procedure, a new sample is classified as an outlier if its surrounding region has a probability density below a specified threshold. The problem here is that sometimes it may be impossible to estimate the complete density of the data (e.g. small sample sizes). Boundary approaches have been developed focusing solely on the boundary of the data in order to overcome this kind of difficulty. Consequently, with these approaches, estimation of the complete probability density can be avoided and it becomes possible to learn from the data when the exact target density distribution is unknown. Moreover, it is sufficient that the user indicate only the boundary of the target class by using examples, and there is no need to model or sample the complete distribution [24].

The first attempts on this direction were made by Moya et al. [14] who trained a neural network with extra constraints to give closed boundaries. More convincing results were presented by Scholkopf et al. [21] who proposed to separate the target samples from the origin with maximal margin using a hyperplane. There is also the work of Tax and Duin [25].

3.1. SVM overview

In this section, we briefly describe the standard support vector machines proposed by Vapnik [27] and the modifications introduced by Scholkopf et al. [21] to build the one-class SVM. By providing this background, we aim to make our treatment of one-class classification in this paper complete.

The idea behind the SVM is to map the input vectors into a high-dimensional feature space using the "kernel trick" and then to construct a linear decision function in this space so that the dataset becomes separated with a maximum margin. Let dataset $(x_1, y_1), \ldots, (x_l, y_l), x \in \Re^n, y \in \{1, -1\}$ be a training set. The standard SVM should solve the following primal problem:

$$\begin{cases} \min\limits_{\mathbf{w},b,\xi} & \dfrac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum\limits_{i=1}^{l}\xi_i \\ \text{s.t.} & y_i(\mathbf{w}^T\Phi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \;\; i = 1, \ldots, l \end{cases} \tag{3}$$

where Φ is the kernel function. The solution $\mathbf{w}$ and b of this equation forms the linear decision function. ξ is known as a slack variable. The parameter C indicates how severely errors must be punished. The choice of C may have a strong effect on the behavior of the classifier for difficult classification problems, e.g. if the errors are punished too much, the SVM can overfit the training data. For computational reasons, instead of solving the problem of Eq. (3) directly, the SVM solves its dual problem as follows:

$$\begin{cases} \min\limits_{\alpha} & \dfrac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ \text{s.t.} & 0 \le \alpha_i \le C, \ \ i = 1,\ldots,l \\ & y^T\alpha = 0 \end{cases} \tag{4}$$

where $Q_{ij} \equiv y_i y_j \Phi(x_i)^T \Phi(x_j)$. Then, the solution of Eq. (4) is used to compute w and b in Eq. (3). To avoid computing the dot product in the high-dimensional feature space, the SVM uses a kernel function. One of the most commonly used kernels is the RBF (radial basis function) kernel, $K(x_1, x_2) = e^{-\gamma(x_1-x_2)(x_1-x_2)}$.

Like the traditional SVM, the one-class algorithm maps the input data into a high-dimensional feature space (via a kernel) and iteratively finds the maximal margin hyperplane that best separates the training data from the origin. It can be viewed as a classical two-class SVM where all the training data lies in the first class, while the origin is taken as the only member of the second class. It uses a parameter $v \in \{0, 1\}$ to control the tradeoff between training error and model complexity [26].

Given a training set without any class information, $x_i \in \Re^n, i = 1,\ldots,l$, the primal form of the one-class SVM is as follows:

$$\begin{cases} \min\limits_{\mathbf{w},\xi,\rho} & \dfrac{1}{2}\mathbf{w}^T\mathbf{w} - \rho + \dfrac{1}{vl}\sum\limits_{i=1}^{l}\xi_i \\ \text{s.t.} & \mathbf{w}^T\Phi(x_i) \ge \rho - \xi_i \\ & \xi_i \ge 0, \ \ i = 1,\ldots,l \end{cases} \tag{5}$$

The solution $\mathbf{w}$ and ρ of Eq. (5) form the linear decision function. The dual problem of the one-class SVM is as follows:

$$\begin{cases} \min\limits_{\alpha} & \dfrac{1}{2}\alpha^T Q\alpha \\ \text{s.t.} & 0 \le \alpha_i \le \dfrac{1}{vl}, \ \ i = 1,\ldots,l \\ & e^T\alpha = 1 \end{cases} \tag{6}$$

where $Q_{ij} = K(x_i, x_j) \equiv \Phi(x_i)^T\Phi(x_j)$. Then, the solution α of Eq. (6) is used to compute w and ρ in Eq. (5).

Fig. 4 presents a simple example of the one-class SVM to illustrate how the data can be separated from the outliers. Considering the context of the biometric fusion, Fig. 4a shows the distribution of the impostors (the support is indicated by the circle that encloses the data) and a small group of genuine data (in this case, the outliers) in the input space. An outlier is any data instance that lies outside the support of the training data. After using a suitable kernel to project the data onto the feature space, the data distribution is shown in Fig. 4b. The hyperplane $\mathbf{w}$ separates the training data from the origin by a maximal margin $\rho/\|\mathbf{w}\|$. Data mapped to the same side of the origin will be given a negative one-class SVM value ($f_{oc} < 0$), whereas those mapped to the side of the training data will have positive values.

In this work, we have used the RBF kernel. However, unlike the traditional two-class formulation, there are no explicit penalties for false positives. Consequently, larger values of γ in the RBF kernel are required to achieve tight approximations for the performance region. What we can observe is that the SVM tends to degenerate into Parzen window estimators as larger values for γ are used.

4. Experiments

As stated elsewhere, the main objective in combining several matchers is to improve the reliability of the system. In the ROC context, this means improving the AUC, or, more specifically, improving the GAR for a given FAR. In this work, we adopted the GAR as the metric for the fixed FAR. The thresholds used for BSSR1 and MCYT were 10^{-4} and 10^{-1}, respectively. The same threshold could not be applied when taking into account the nature of the data, as depicted in Figs. 2 and 3.

In this section, we present three different sets of experiments. As stated in the Introduction, biometric fusion can be approached as a classifier combination problem or as a pattern classification problem. First, we report all the experiments carried out based on the classifier combination approach. Then, the remaining experiments are related to the pattern classification strategy. First we apply the classical two-class SVM and then the proposed one-class SVM. The objective of all these experiments is to provide a good basis for comparison.

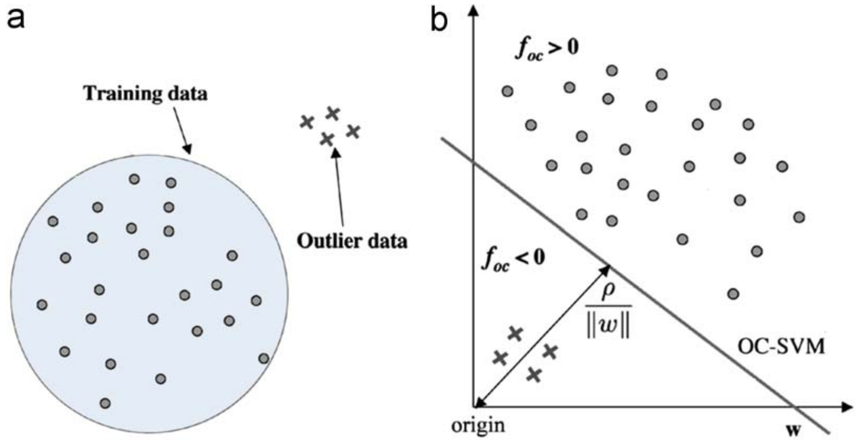

Fig. 4. (a) Distribution of the training data and some outliers and (b) the projection of the data onto a feature space after using a suitable kernel.

Table 1
Biometric fusion using different classifier combination rules.

Strategy	BSSR1 Normalization			MCYT
	Min–Max	*Z*-Score	Col-Norm	
Sum	0.971	0.969	0.930	0.998
Weighted sum	**0.994**	0.974	0.953	**0.999**
Product	0.916	0.742	0.765	0.991
Mean	0.974	0.965	0.789	0.998
Min	0.914	0.866	0.914	0.997
Max	0.852	0.742	0.742	0.989

GAR for FAR fixed at 10^{-4} and 10^{-1} for BSSR1 and MCYT, respectively.

4.1. *Classifier combination*

Typically, matcher scores vary from one system to another in scale, distribution, and meaning. Consequently, matcher scores for each modality must first be normalized to the same interval. In our experiments using BSSR1, we use three well-known normalization methods, namely, *Z*-Score, Min–Max, and Column-Norm [10,22]. For our experiments using MCYT, we did not apply any normalization technique because the data were already normalized.

After normalization, we tested the following fusion rules: sum, weighted sum, product, max, min, and mean.

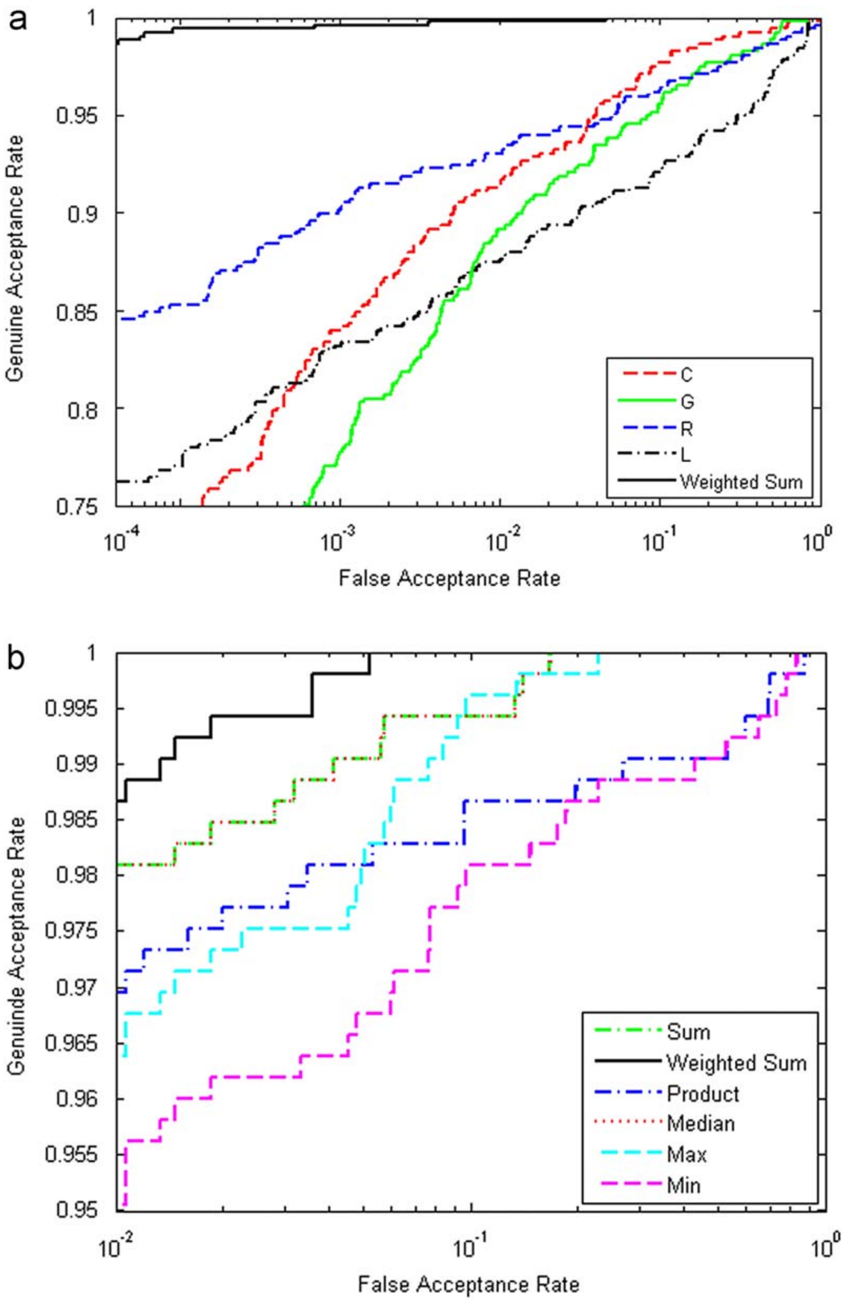

Fig. 5. ROC of the weighted sum rule: (a) BSSR1 and (b) MCYT.

Table 1 compares all these combination rules using the normalization strategies as outlined above.

From Table 1, we see that the Min–Max normalization method usually leads to the best performance. The same findings have been reported by Snelick et al. [22]. We can also observe that the weighted sum is the best combination rule for both databases. Some justification for this behavior can be found in the work of Jain and Ross [9] and of Fierrez et al. [4]. They argue, and show through experimentation, that the performance of biometric systems can be further improved by learning user-specific parameters. In such a case, the parameters would be the weights used by each matcher, which indicate the importance of matching scores provided by each biometric trait. One drawback of this approach compared to other combination rules is the need to find the best weights through some kind of search.

In our experiments we have used an exhaustive search to define the best weights for the weighted sum approach. The best weights were 0.3C, 0.4G, 0.2R, 0.1L, and 0.4N, 0.0Q, 0.6S, for BSSR1 and MCYT, respectively. In the case of BSSR1, this emphasizes that all the matchers are important for producing a more reliable combination. In the case of MCYT, on the other hand, Q is much weaker, as we

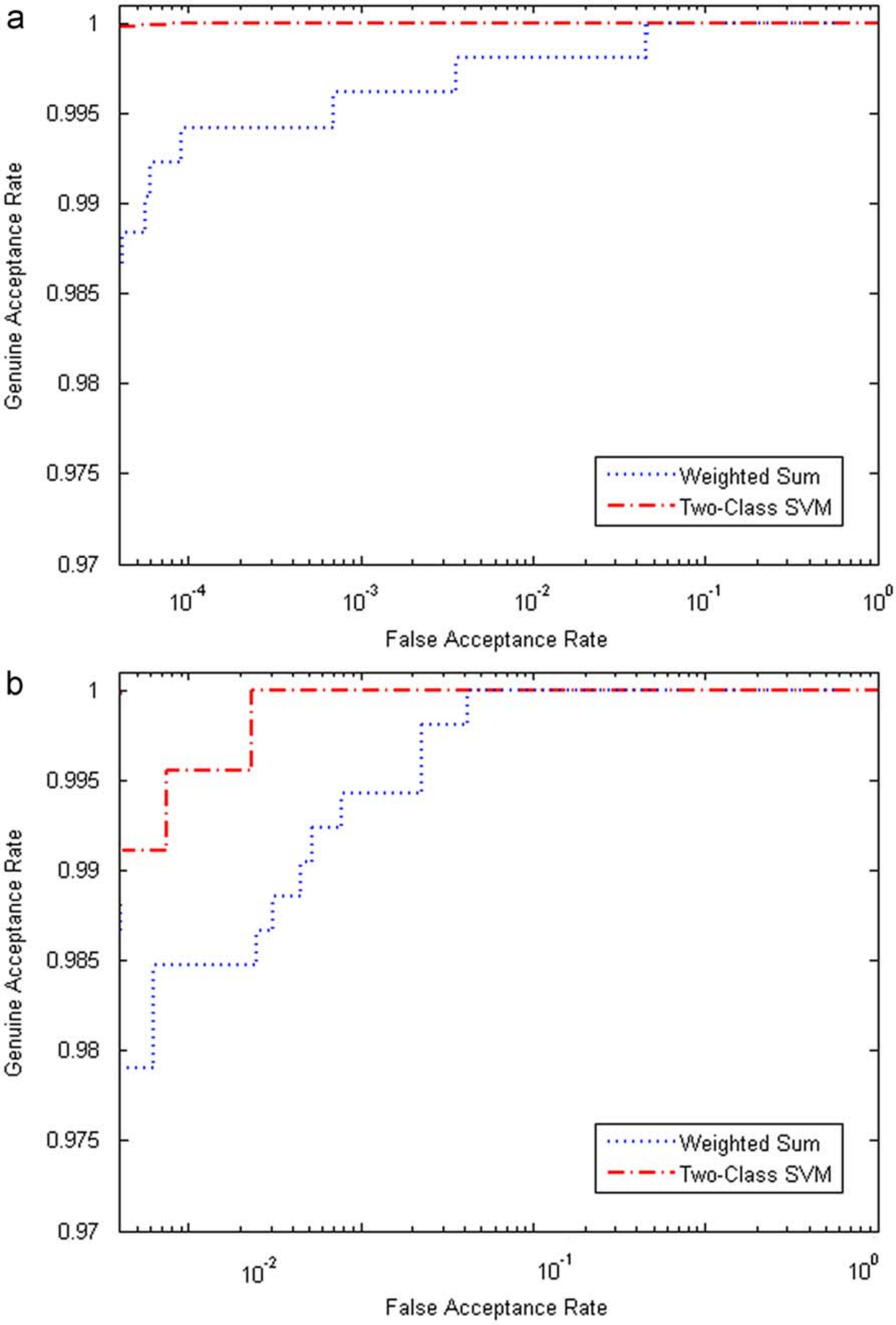

Fig. 6. Performance of the standard two-class SVM: (a) BSSR1 and (b) MCYT.

can observe from Fig. 3, and therefore was not selected for the pool. Fig. 5a shows the ROC for the weighted sum rule using Min–Max normalization, while Fig. 5b depicts the ROC for MCYT using the same combination rule.

4.2. Two-class SVM

So far, we have seen that most of the combination rules can yield interesting improvements. In this section we report the experiments carried out using the standard two-class SVM. We demonstrate that the results achieved by the weighted sum rule can be further improved. Moreover, the results reported here will help us to better assess the results produced by the one-class SVM.

Since this is a pattern classification approach, the database should be divided into three parts: training, validation, and testing. For BSSR1, we used 217 genuine and 240,000 impostors for testing, and 300 genuine and 20,000 impostors for training and validation. For MCYT, we used 225 genuine and 450 impostors for testing, and 300 genuine and 300 impostors for training and validation. For validation, we used a k-fold cross-validation ($k = 10$) because of the small number of genuine samples. The kernel used in these experiments was the RBF and the parameters C and γ were determined through a grid search.

We have evaluated the impact of increasing the number of both the genuine and the impostor samples. For the genuine samples, the size of the database ranges from 100 to 300, while for the impostor samples, it ranges from 100 to 20,000 (in the case of BSSR1). We have noticed that increasing the number of genuine samples improves performance slightly, but the greatest improvement was achieved when the number of impostors was increased. The experiments revealed, however, that after 10,000 impostors there was no further improvement in the results. Fig. 6 shows the performance of the standard two-class SVM and compares it to the weighted sum as well.

Our findings corroborate with other results reported in the literature for different biometric databases. For example, Fierrez-Aguilar et al. [5] successfully used an SVM to compute a multimodal combined score using face, fingerprint, and on-line signature biometrics. Jiang and Su [11] demonstrate that the fusion using SVM surpassed methods such as Fisher linear discriminant analysis and the weighted sum method.

With regard to the importance of the matchers, we have noticed the same behavior here as with the weighted sum approach. We performed exhaustive feature selection (feasible because we have only four features) and the best results were produced when the four matchers were available.

4.3. One-class SVM

For the experiments regarding one-class classification, the BSSR1 database was divided into 20,000 impostors for training, 2000 impostors for validation, and 517 genuine plus 240,000 impostors for testing. The MCYT database was divided into 400 impostors for training, 350 impostors for validation, and 525 genuine plus 250 impostors for testing.

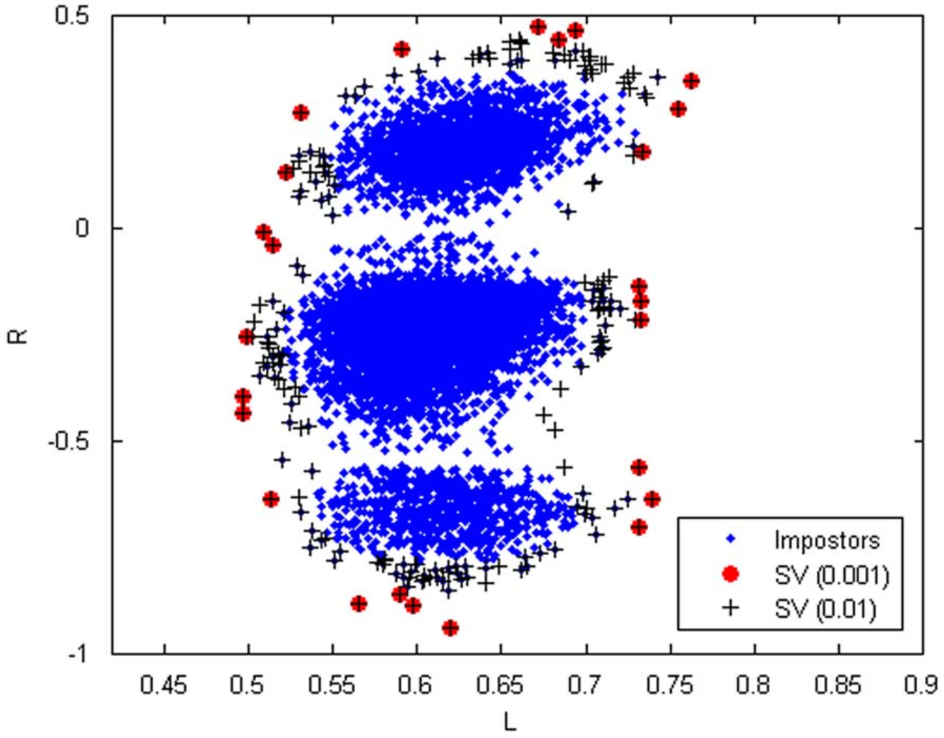

Fig. 7. Different boundaries built by different ν parameters.

As for the two-class classification, the kernel we used was the RBF, where the following two parameters have to be adjusted: the maximum fraction of training error, ν, and the kernel parameter, γ. According to Scholkopf et al. [21], when the offset of the hyperplane for the origin is greater than zero, then the parameter ν can be set to the highest allowable fraction of misclassification of the target class. In our case, we have allowed a 1% of error on the training set, i.e. $\nu = 0.01$.

The impact of assigning different values to ν is shown in Fig. 7. In this experiment, we used two matchers (L and R) as features and two different values for ν: 0.01 and 0.001. Using $\nu = 0.01$, we get a more specialized boundary, while $\nu = 0.001$ produces a more generic boundary.

In order to tune the kernel parameter, the number of support vectors can be minimized by dividing the total number of support vectors by the number of training examples. This gives a leave-one-out bound on the test error of the training data [3]. Another alternative is to maximize the margin of separation from the origin, ($\rho/\|\mathbf{w}\|$ from Fig. 4), which is equivalent to minimizing the radius of the smallest sphere enclosing the data [25]. In any case, the most common way to find the kernel parameter is through a validation set. Because of the

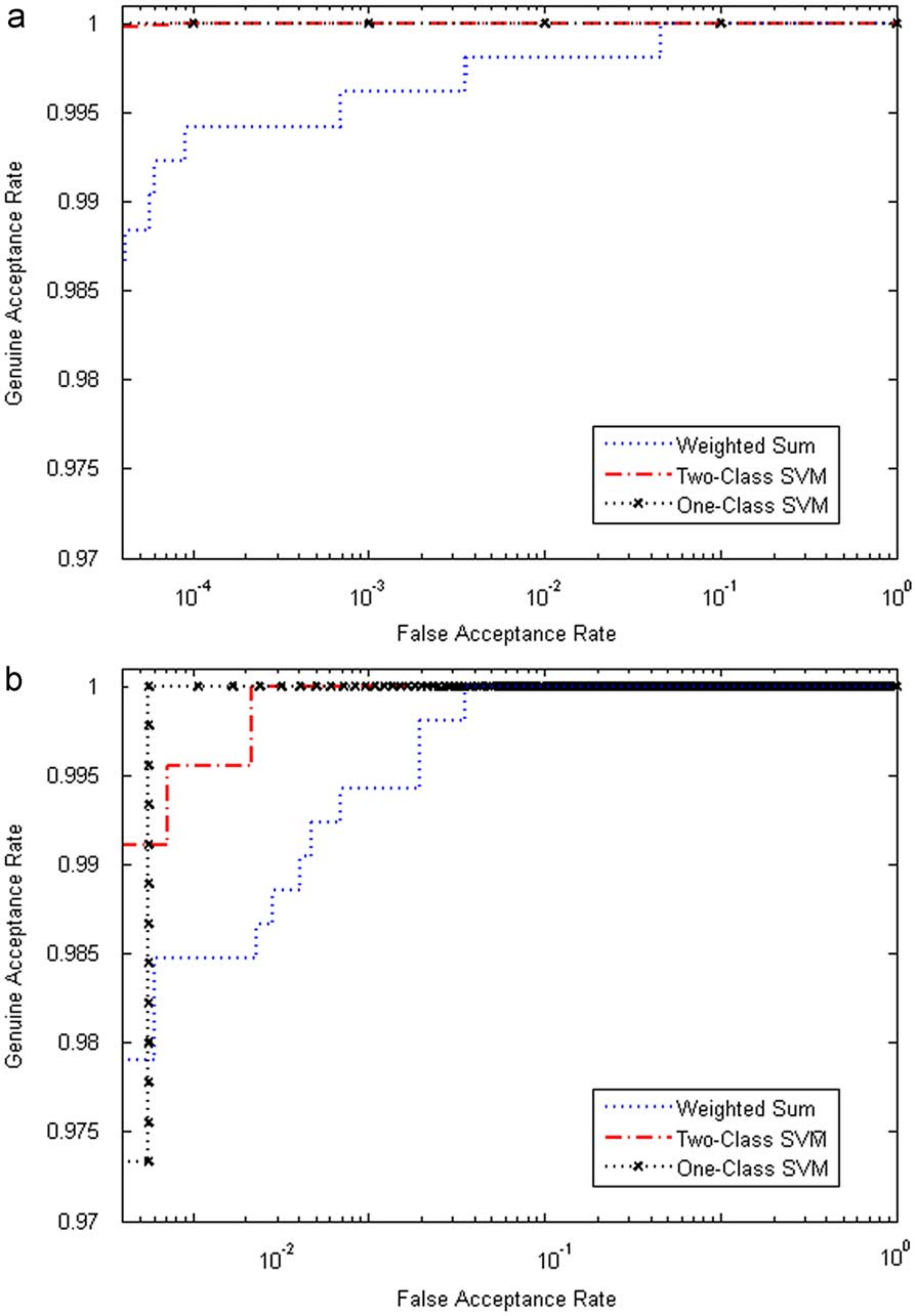

Fig. 8. Performance of the one-class SVM: (a) BSSR1 and (b) MCYT.

nature of the problem, a validation set rarely exists, which makes the task of parameter tuning more difficult.

Taking into account that we have access to few data from the genuine class, we could have used them to build a validation set to tune the parameters. But, to better simulate the difficulties of the one-class classification, we only used samples of impostors in our validation set.

As in the previous simulations with the standard two-class SVM, we also assessed the impact of the size of the database. In the case of BSSR1, we varied the size of the database from 100 to 20,000, and noted only a slight improvement up to 5000. Beyond that we observed no improvement at all. In other words, the 5000 samples represent the cluster of impostors quite well. At the same time, we noted that all 400 samples were quite important for training in the case of MCYT. Fig. 8 shows the ROC for the one-class SVM for both databases.

In general, it cannot be expected that the one-class classifier will perform as well as a two-class classifier, because training samples from two classes provides more information to define the decision boundary than sampling only on one side [30]. As stated previously, one-class classification is very useful when only one data class is available. As depicted in Fig. 8, one-class classification yielded the same ROC as the two-class SVM. For BSSR1 database the curves are the same (Fig. 8a) while for MCYT the curves are slightly different before FAR $= 10^{-2}$ (Fig. 8b). In terms of recognition rate, the one-class SVM achieved 99.67% on the test set, as against 99.80% on the two-class SVM, for the BSSR1 database. For the MCYT, both models, one-class and two-class, achieved a 99.9% recognition rate.

As discussed earlier, the one-class classification attempts to describe the target data domain by finding a hypersphere containing most of the target data (Fig. 7). A plausible justification for the good performance of the one-class classification on both databases is that the impostors are roughly distributed within a single spherical region. If the target samples were scattered in several small regions, a spherical boundary to fit the data would enclose a large empty area, which would enhance the chances of accepting outliers [28]. In that case, a two-class SVM would perform much better.

5. Conclusion

In this paper, we have proposed the use of one-class classification with SVM to combine the scores of four different biometric systems of the NIST BSSR1 and three biometric systems of the MCYT. Through a series of experiments, we have demonstrated that the one-class SVM surpasses all the combination rules and compares well with the standard two-class SVM. It is worth noting, though, that the one-class SVM performs well when the target data meet certain constraints, such as when they are roughly distributed within a single spherical region.

Moreover, this strategy is quite suitable when the data are highly unbalanced or when only one class is available for training. In further work, we plan to investigate user-dependent feature selection to form more compact clusters. In this case, the volume of the hypersphere could serve as a criterion during the search. In addition, we plan to evaluate the one-class strategy on the different biometric datasets available in the literature.

Acknowledgment

This research has been supported by The National Council for Scientific and Technological Development (CNPq) Grants 471496/2007-3 and 306358/2008-5.

References

[1] V. Barnett, T. Lewis, Outliers in Statistical Data, second ed., Wiley, New York, 1978.

[2] H.E. Cetingula, E. Erzin, Y. Yemez, A.M. Tekalp, Multimodal speaker/speech recognition using lip motion, lip texture and audio, Signal Processing 86 (12) (2006) 3549–3558.

[3] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and other Kernel-based Learning Methods, Cambridge University Press, Cambridge, 2000.

[4] J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, J. Gonzalez-Rodrigues, Adapted user-dependent multimodal biometric authentication exploiting general information, Pattern Recognition Letters (2005) 2628–2639.

[5] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodrigues, Fusion strategies in multimodal biometric verification, in: International Conference on Multimedia and Expo, vol. 3, 2003, pp. 5–8.

[6] B. Gutschoven, P. Verlinde, Multi-modal identity verification support vector machines, in: Proceedings of the Third International Conference on Information Fusion, 2000.

[7] V. Hodge, J. Austin, A survey of outlier detection methodologies, Artificial Intelligence Review 22 (2) (2004) 85–126.

[8] L. Hong, A. Jain, Integrating faces and fingerprints for personal identification, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (12) (1998) 1295–1307.

[9] A. Jain, A. Ross, Learning user-specific parameters in a multibiometric system, in: International Conference on Image Processing, 2002.

[10] A.K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (12) (2005) 2270–2285.

[11] C.-H. Jiang, G.-D. Su, Information fusion in face and fingerprint identity verification system, in: Third International Conference on Machine Learning and Cybernetics, 2004, pp. 3529–3535.

[12] A.K. Jain, A. Ross, S. Pankanti, Biometrics: a tool for information security, IEEE Transactions on Information Security 1 (2) (2006) 125–143.

[13] R. Luis-Garcia, C. Alberola-López, J. Ruiz-Alzola, O. Aghzout, Biometric identification systems, Signal Processing 83 (12) (2003) 2539–2557.

[14] M. Moya, M. Koch, L. Hostetler, One-class classifier networks for target recognition applications, in: World Congress on Neural Networks, 1993, pp. 797–801.

[15] K. Nandakumar, Y. Chen, S.D. Dass, A.K. Jain, Likelihood ratio-based biometric score fusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2) (2008) 342–347.

[16] National Institute of Standards and Technology, Nist biometric scores set-release 1, 2004 ⟨http://www.itl.nist.gov/iad/894.03/biometricscores⟩.

[17] 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, Q.-I. Moro, Mcyt baseline corpus: a bimodal biometric database, IEE Proceedings—Vision Image and Signal Processing 150 (6) (2003) 395–401.

[18] L. Parra, G. Deco, S. Miesbach, Statistical independence and novelty detection with information preserving non-linear maps, Neural Computation 8 (1996) 260–269.

[19] P. Verlinde, G. Chollet, M. Achcrov, Multi-modal identity verification using expert fusion, Information Fusion 1 (1) (2000) 17–33.

[20] A. Ross, A.K. Jain, Information fusion in biometrics, Pattern Recognition Letters 24 (2003) 2115–2125.

[21] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, R. Williamson, Estimating the support of a high dimensional distribution, Neural Computation 13 (7) (2001) 1443–1472.

[22] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain, Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 450–455.

[23] L. Tarassenko, P. Hayton, M. Brady, Novelty detection for the identification of masses in mammograms, in: Fourth International Conference on Artificial Neural Networks, 1995, pp. 442–447.

[24] D.M.J. Tax, One-class classification, Ph.D. Thesis, TU Delft, 2001.

[25] D.M.J. Tax, R.P.W. Duin, Support vector domain description, Pattern Recognition Letters 20 (1999) 1191–1199.

[26] Q.-A. Tran, X. Li, H. Duan, Efficient performance estimate for one-class support vector machine, Pattern Recognition Letters 26 (2005) 1174–1182.

[27] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, Berlin, 1995.

[28] D. Wang, D. Yeung, E. Tsang, Structured one-class classification, IEEE Transactions on Systems, Man, and Cybernetics, Part B—Cybernetics 36 (6) (2006) 1283–1295.

[29] Y. Wang, T. Tan, A.K. Jain, Combining face and iris biometrics for identity verification, in: Fourth International Conference on AVBPA, 2003, pp. 805–813.

[30] H. Yu, SVMC: single-class classification with support vector machines, in: Proceedings of the International Joint Conferences on Artificial Intelligence, 2003, pp. 567–576.