Expert Systems with Applications 38 (2011) 58–63

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Likelihood ratio based features for a trained biometric score fusion

Loris Nanni [a,*], Alessandra Lumini [a], Sheryl Brahnam [b]

[a] Department of Electronic, Informatics and Systems (DEIS), Università di Bologna, Via Venezia 52, 47023 Cesena, Italy
[b] Computer Information Systems, Missouri State University, 901 S. National, Springfield, MO 65804, USA

ARTICLE INFO

Keywords:
Likelihood ratio
Mixture of Gaussians
Support vector machine
Biometric score fusion

ABSTRACT

In this work, we present a novel trained method for combining biometric matchers at the score level. The new method is based on a combination of machine learning classifiers trained using the match scores from different biometric approaches as features. The parameters of a finite Gaussian mixture model are used for modelling the genuine and impostor score densities during the fusion step.

Several tests on different biometric verification systems (related to fingerprints, palms, fingers, hand geometry and faces) show that the new method outperforms other trained and non-trained approaches for combining biometric matchers.

We have tested some different classifiers, support vector machines, AdaBoost of neural networks, and their random subspace versions, demonstrating that the choice for the proposed method is the Random Subspace of AdaBoost.

1. Introduction

One recent focus of interest in biometrics research is the successful combination of different sources of information resulting in the so-called multi-biometric. Unibiometric systems, which are based on a single source of information may suffer from limitations such as the lack of uniqueness and non-universality of the chosen biometric trait, noisy data and spoof attacks (Ross, Nandakumar, & Jain, 2006). Multi-biometric systems fuse information from multiple biometric sources in order to achieve better recognition performance and to overcome other limitations of unibiometric systems (Brunelli & Falavigna, 1995; Prabhakar & Jain, 2002; Toh, Jiang, & Yau, 2004). A sound theoretical framework for combining classifiers with application to biometric verification is described in Bigun, Bigun, Duc, and Fischer (1997), where an algorithm functioning as a supervisor in a multi expert decision making machine is proposed which uses the Bayes theory in order to estimate the biases of individual expert opinions (the scores of each unibiometric system). Machine learning approaches have also been applied for combining biometric classifiers (Fierrez-Aguilar, Garcia-Romero, Ortega-Garcia, & Gonzalez-Rodriguez, 2004).

A first study on the combination of different fingerprint systems submitted to FVC2004 is carried out in Fierrez-Aguilar, Nanni, Ortega-Garcia, Cappelli, and Maltoni (2005) and Maio and Nanni (2006) where the benefits and limits of the resulting multiple classifier approaches have been analysed. In these works, it is shown that combining systems that are based on heterogeneous matching strategies permits a reduction of the Equal Error Rate with respect to the best unibiometric system. In Lumini and Nanni (2006), a very effective multi-biometric system based of the combination of different fingerprint systems and an Iris matcher is proposed.

In Nanni and Lumini (2008), starting from the similarity scores obtained by two biometric matchers (Face and Iris), a set of eight “original” features are extracted to discriminate between genuine and impostor classes. Moreover several new “artificial” features are generated by combining one or more original ones, by means of some mathematical operators. The resulting system, based on the original and a selection of the artificial features has experimentally demonstrated to give a very good verification performance.

In multi-biometric systems fusion performed at the score level is generally preferred (Prabhakar & Jain, 2002) to fusion at the feature and decision levels; the score fusion techniques proposed in the literature can be divided into three categories (following the taxonomy used in Nandakumar, Chen, Dass, & Jain (2008)):

- Transformation-based score fusion: The match scores are first normalized (transformed) to a common domain and then combined. The main drawback is that these methods are data-dependent and require extensive empirical evaluation (Jain, Nandakumar, & Ross, 2005; Snelick, Uludag, Mink, Indovina, & Jain, 2005; Toh et al., 2004).
- Classifier-based score fusion: Scores from multiple matchers are used to train a classifier that discriminates between genuine and impostor (Brunelli & Falavigna, 1995; Fierrez-Aguilar, Ortega-Garcia, Gonzalez-Rodriguez, & Bigun, 2005; Ma, Cukic, & Singh, 2005) features.

* Corresponding author. Tel.: +39 0547 339121; fax: +39 0547 338890.
E-mail addresses: loris.nanni@unibo.it (L. Nanni), alessandra.lumini@unibo.it (A. Lumini), sbrahnam@missouristate.edu (S. Brahnam).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.06.006

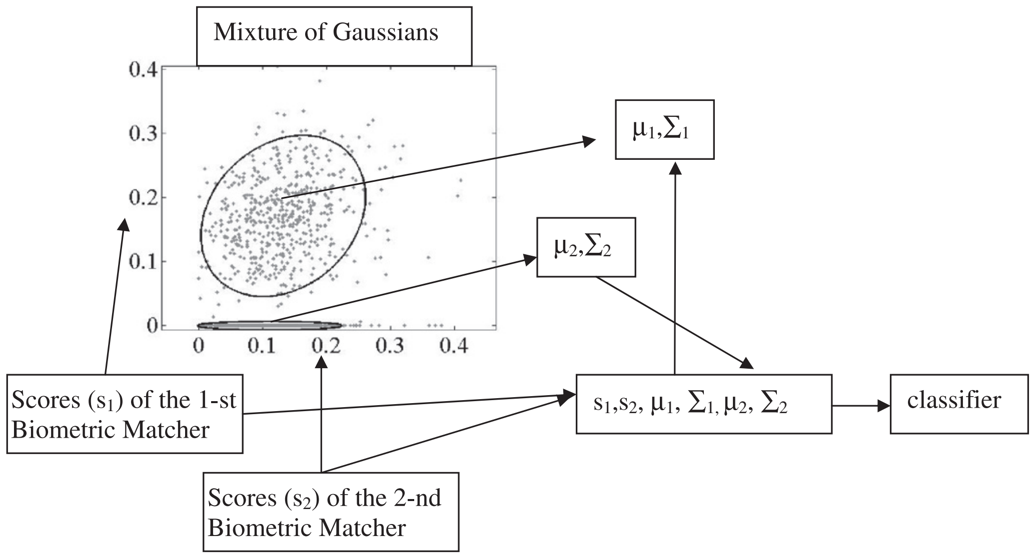

Fig. 1. Biometric fusion system proposed in this work (in the case of the fusion of two matchers).

- Density-based score fusion: This approach is based on the likelihood ratio test, and it requires the densities estimation of genuine and impostor match scores (Griffin, 2004). A comparison of eight biometric fusion techniques conducted by NIST (Ulery, Hicklin, Watson, Fellner, & Hallinan, 2006) with data from 187.000 subjects concluded that Likelihood Ratio was the most accurate method, but it was complex to implement (their density estimation was based on the use of kernel density estimator (KDE) (Lehmann & Romano, 2005)).

In Nandakumar et al. (2008), it is shown that a mixture of Gaussians (MoG) is quite effective in modelling the genuine and impostor score densities, and it is easier to implement than KDE. Their results based on NIST fusion data (National Institute of Standards & Technology, 2004) show that MoG outperforms both the standard Sum Rule (Kittler, Hatef, Duin, & Matas, 1998) and the support vector machine (Duda, Hart, & Stork, 2000) based trained fusion.

In this work, we propose a supervised fusion where the classifiers are trained using as features the match scores and the parameters of the finite Gaussian mixture model that are used for modelling the genuine and impostor score densities of the training data.

Experimental results are reported for two different state-of-the-art classifiers: the support vector machine (SVM) and the AdaBoost of neural network (ADA), and for each classifier their random subspace version has also been tested.

Several tests using different biometric characteristics (fingerprints, the palm, fingers, hand geometry, and the face) show that our method (mainly the one based on the Random Subspace of ADA) outperforms other trained and non-trained approaches for combining biometric matchers.

This paper is organized as follows. In Section 2, the details of the new feature extraction approach is presented. In Section 3, some experimental results are presented and discussed. Finally, we draw conclusions in Section 4.

2. System overview

According to the Neyman–Pearson theorem (Lehmann & Romano, 2005), the optimal test for assigning a score vector **x** to the class genuine or impostor is the likelihood ratio test given by $f_{gen}(\mathbf{x})/f_{imp}(\mathbf{x})$, where $f_{gen}(\mathbf{x})$ and $f_{imp}(\mathbf{x})$ are the densities of the genuine training data and of the impostor training data.

It is well known that the Gaussian density is not appropriate for modelling biometric match scores; to obtain a more reliable density method, the normal distribution can be extended to a mixture of Gaussians (MoG)[1] (Figueiredo & Jain, 2002) (i.e., the linear combination of normal distributions). The main drawback of MoG is that it requires far more data for training (Li & Barron, 1999; Rakhlin, Panchenko, & Mukherjee, 2005). In this paper, the mixture is estimates using the EM algorithm (Nanni, 2006) and a number of Gaussians. K is automatically calculated by means of the minimum message length criterion.

The estimates of $f_{gen}(\mathbf{x})$ and $f_{imp}(\mathbf{x})$ are obtained as a mixture of Gaussians; the probability distribution for a d-dimensional object $\mathbf{x}$ is given by:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\},$$

where $\boldsymbol{\mu}$ is the mean and $\boldsymbol{\Sigma}$ is the covariance matrix of the training set, and by

$$f_{gen}(\mathbf{x}) = \sum_i \psi P_{gen,i} \times f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}), \psi,$$

where $p_{gen,i}$ is the weight assigned to the ith mixture component.

Given a set of K Gaussians, we obtain $K \times 2$ parameters $(\boldsymbol{\mu}_1, \sum_1, \ldots, \boldsymbol{\mu}_K, \sum_K)$. These parameters (concatenated with the biometric scores) are used to train a given classifier.

In Fig. 1, our system is detailed.

3. Experiments and discussion

As classifier we have tested the support vector machine (SVM) and the AdaBoost of neural networks (ADA). Moreover, for each classifier, we have tested also their random subspace (RS) version (Ho, 1998). We use an AdaBoost.M1 with 50 iterations of a feed-forward back-propagation network. As SVM we report results on a Linear SVM and a radial basis function SVM. The Random Subspace Method modifies the training data set (generating NK new

[1] The MATLAB code for this algorithm is available at http://www.lx.it.pt/~mtf/mixturecode.zip [bestk, bestpp, bestmu, bestcov, dl, countf] = mixtures4(DATA, 1, 15, 1e-5, 1e-4, 0).

training sets containing only *NFe* of the original features; in this paper *NK* = 25 and *NFe* = 50%). It builds classifiers on these modified training sets, and then combines them into a final decision rule (in this paper the Sum Rule is used (Kittler et al., 1998)).

Experiments have been conducted on several datasets:

- The four fingerprint databases from FVC 2004 DBs (Maio & Nanni, 2006), each containing 800 images from 100 individuals (DB1–DB3 are obtained using different sensors, while DB4 is obtained using an artificial generator (Cappelli, 2004)).
- A Palm database that contains 1000 inkless right-hand images from a digital camera, seven samples from each user, for 100 users. From this dataset several biometric characteristics are extracted (Palm, Hand Geometry, Middle Finger, and Ring Finger). The palm is extracted using a method similar to that proposed in Connie, Jin, Ong, and Ling (2005). The images of

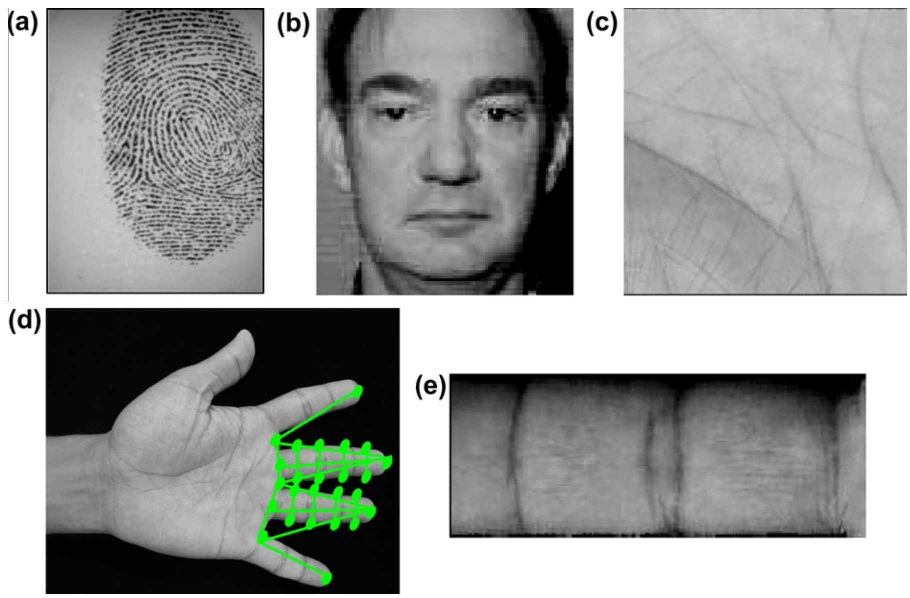

Fig. 2. Some samples from the dataset used in this work, (a) fingerprint; (b) face; (c) palm; (d) Hand geometry features (the length of the green lines that link two green balls are the extracted features); (e) finger. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

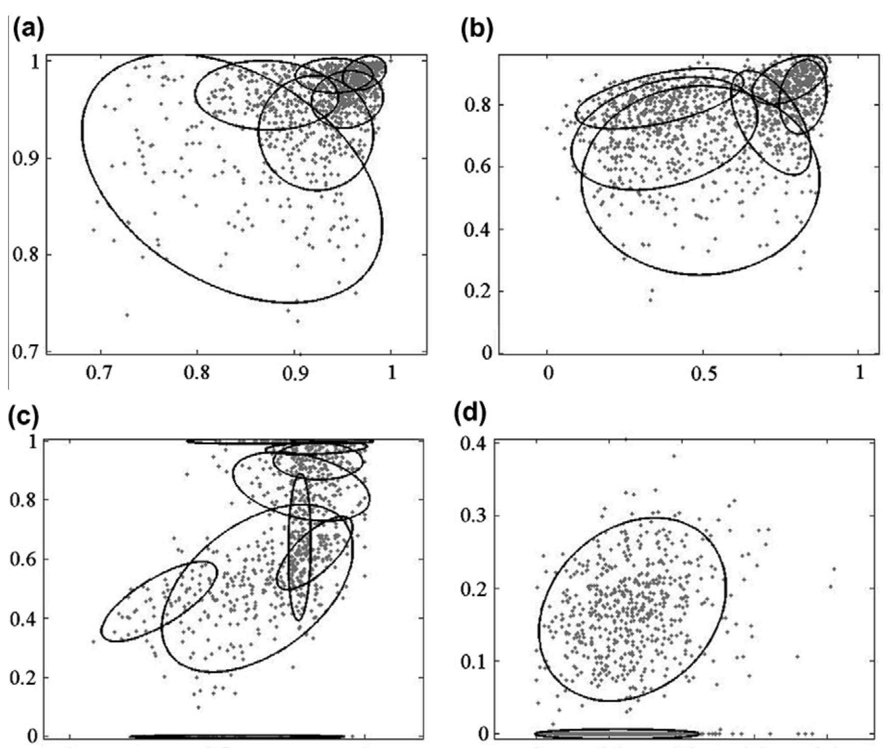

Fig. 3. Real examples of MoG: (a) PALMFINGER-Genuine; (b) PALMFINGER-Impostor; (c) DB3-Genuine; (d) DB3-Impostor.

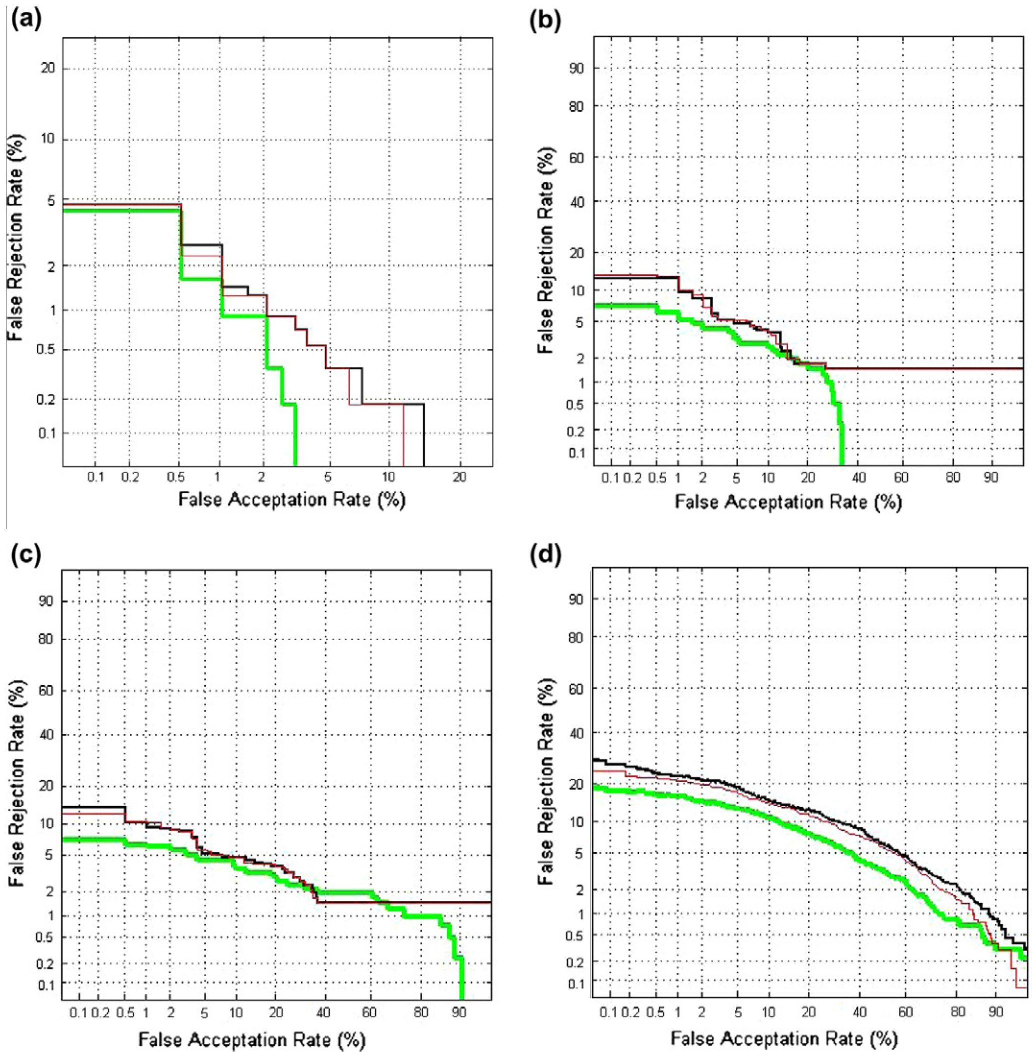

Fig. 4. DET-curves: (a) FVC2004-DB2; (b) PALMFINGER; (c) HAND; (d) FACE.

the Palm and of the Finger have been resized to the same dimension of 100×100 before processing.
- A Face database, the Notre-Dame Dataset[2] collection D (http://www.nd.edu/~cvrl/), that contains a total of 275 different persons who participated in one or more sessions. Two four-week sessions were conducted for data collection with approximately a six weeks time lapse between the two.

In Fig. 2, we show some samples from the datasets. According to the very difficult FVC2002 testing protocol, the following matching attempts are calculated:

- *Genuine recognition attempts*: The template of each impression is matched against the remaining impressions of the same individual, while avoiding symmetric matches.
- *Impostor recognition attempts*: The template of the first impression is matched against the first impressions of the remaining individuals while avoiding symmetric matches.

The performance have been measured by means of the equal error rate (EER) (Maio, Maltoni, Jain, & Prabhakar, 2003). Moreover, in order to confirm the benefit of the our method, the DET curve has been also considered.The DET curve (Martin et al., 1997) is a two-dimensional measure of classification performance that plots the probability of false acceptation against the rate of false rejection.

Now we report the matchers involved in the fusions tested in this paper:

- In the FVC2004 DBs, we use the winner of the competition and the third best matcher (the second best matcher has as scores mainly the values 0 or 1, and hence it is not well suited for the fusion).
- The Palm matcher, the Finger matcher and the first Face matcher is the Euclidean distance on the 100 Discrete Cosine Coefficients with higher variance. The pre-processing stage used in Connie et al. (2005) is performed to normalize the images in order to smoothen the noise and lighting effect.
- The second Face matcher is the Euclidean distance on the Locally Binary Patterns features (the histograms of 10 bins, 18 bins and 26 bins are concatenated as in Nanni & Lumini (2007)).
- The Hand Geometry matcher is the Euclidean distance where the features are the length of the lines that link two datum points (see Fig. 2).

[2] http://www.nd.edu/~cvrl/.

Table 1
Number of mixtures found for the genuine data and for the impostor data.

	FVC2004				PALMFINGER	HAND	FACE
	DB1	DB2	DB3	DB4			
GENUINE	8	9	9	6	6	3	7
IMPOSTOR	7	5	2	6	6	5	9

Table 2
EER obtained when the training set contains 80% of the users.

Method	FVC2004				PALMFINGER	HAND	FACE
	DB1	DB2	DB3	DB4			
1st MATCHER	1.89	2.74	0.58	0.56	8.9	8.9	14.39
2nd MATCHER	3.91	2.3	1.38	0.58	9.4	9.4	24.25
3rd MATCHER	–	–	–	–	–	11.3	–
4th MATCHER	–	–	–	–	–	13.2	–
Sum Rule	1.75	1.2	0.7	0.46	7.4	7.6	13.99
ADA	1.66	1.11	0.73	0.35	25.3	7.5	14.17
SVM	1.64	1.16	0.53	0.46	7.1	7.4	14.12
LR	3.62	2.69	1.21	0.73	10.5	19	11.50
ADA-LR	2.08	**0.88**	0.53	0.54	6.2	7.3	12.71
SVM-LR	1.64	0.93	**0.44**	0.39	5.4	5.7	15.78
RS-ADA	1.61	1.09	0.58	**0.33**	5.4	6.6	**11.43**
RS-SVM	**1.43**	1	0.58	0.42	**5.3**	**5.2**	16.43

Bold values indicates the lowest ERR in each database.

For each fusion (see Tables 2 and 3, and Figs. 3 and 4), we report which matchers are involved:

Table 3
EER obtained when the training set contains the 50% of the users.

Method	FVC2004				PALMFINGER	HAND	FACE
	DB1	DB2	DB3	DB4			
1st MATCHER	2.60	3.35	1.04	0.79	8.24	8.24	14.30
2nd MATCHER	4.47	2.53	1.56	0.55	10.62	10.62	25.93
3rd MATCHER	–	–	–	–	–	7.09	–
4th MATCHER	–	–	–	–	–	9.86	–
Sum Rule	2.23	1.49	0.83	0.50	6.53	4.99	13.74
ADA	**2.11**	1.41	0.61	0.47	26.53	5.44	16.20
SVM	2.15	1.56	0.69	0.52	6.57	4.79	13.53
LR	3.72	2.95	1.26	1.11	12.42	13.39	**11.45**
ADA-LR	2.30	1.31	0.66	0.50	7.00	4.71	13.35
SVM-LR	2.20	**0.99**	0.63	0.49	5.42	4.16	12.23
RS-ADA	**2.11**	1.01	**0.54**	0.45	5.73	3.98	11.70
RS-SVM	**2.11**	**0.99**	0.58	**0.43**	**5.16**	**3.86**	11.80

Bold values indicates the lowest ERR in each database.

- FVC2004 DB1-4: the two FVC2004 matchers are combined.
- PALMFINGER: the Palm matcher and the middle finger matchers are combined.
- HAND: the Palm matcher, the middle finger matcher, the ring finger and the hand geometry matchers are combined.
- FACE: the two Face Matchers are combined.

In Table 1, we report the number of mixture found for the genuine data and for the impostor data in the seven datasets used in this work. In Fig. 2 we show some real examples of MoG.

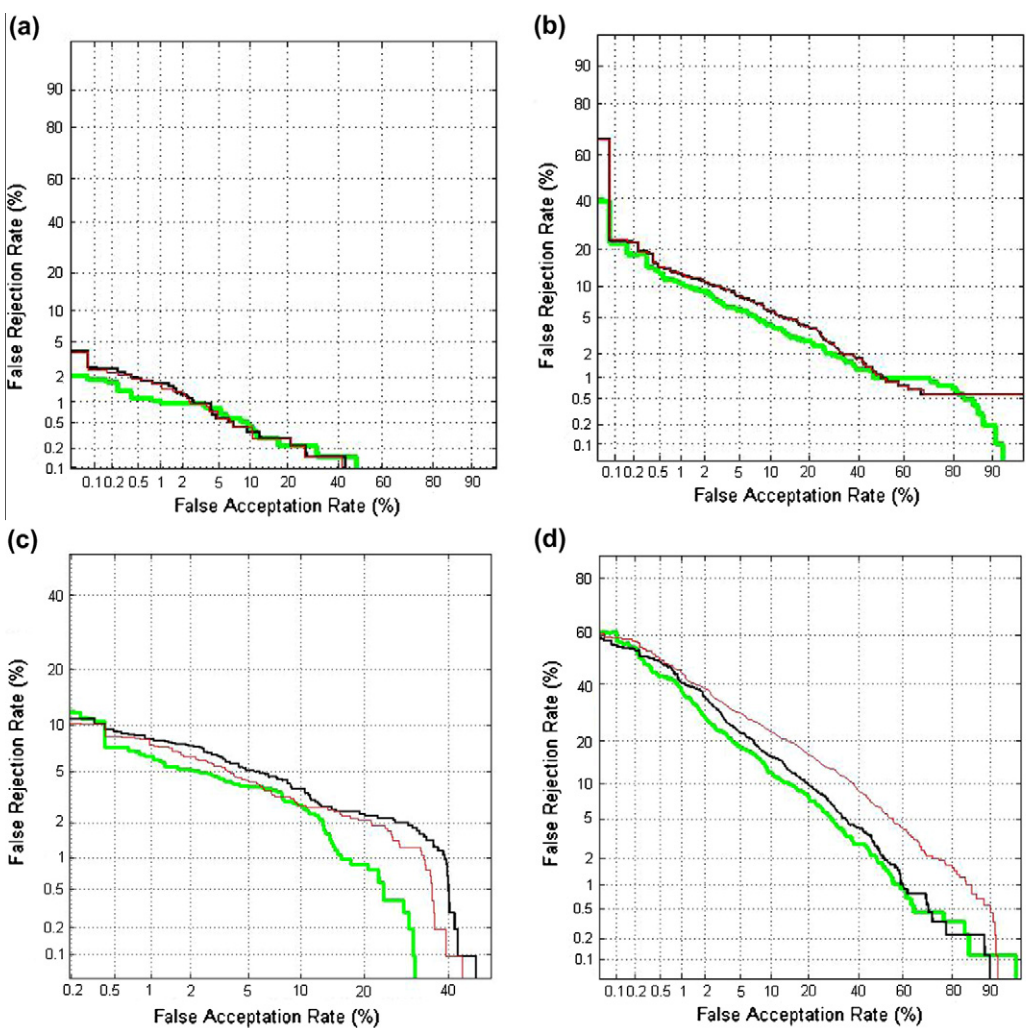

Fig. 5. DET-curves: (a) FVC2004-DB2; (b) PALMFINGER; (c) HAND; (d) FACE.

In Tables 2 and 3, we compare several methods on the tested databases varying the dimension of the training set. In Table 2, the training set contains 80% of the users. In Table 3, the training set contains 50% of the users. For the testing set we consider only the matches where the users that belong to the training set are not present. We randomly divide the users in the training and testing sets ten times, and we report the average EER.

We compare the following state-of-the-art methods:

- ADA, the trained fusion where ADA is trained considering only the match scores.
- SVM, the trained fusion where the Linear SVM is trained considering only the match scores.
- LR, the method proposed in Nandakumar et al. (2008).
- ADA-LR, ADA trained using the features described in Section 2.
- SVM-LR, SVM trained using the features described in Section 2.
- RS-ADA, RS of ADA trained using the features described in Section 2.
- RS-SVM, RS of Radial Basis Function SVM (Gamma = 1, Cost of the constrain violation = 100) trained using the features described in Section 2.

We want to stress that in Table 2, the best performance is always obtained using the classifiers trained using the features described in Section 2. Among the seven tests, the best results are obtained by RS-ADA; it always outperforms the standard methods. Moreover, only in the FACE fusion test does LR works better than SVM. In our opinion, this is due to the fact that LR needs a very large dataset (as the datasets used in Nandakumar et al. (2008)) for training.

Finally, when SVM is trained considering only the match scores, we obtain the best performance using the Linear SVM. When we use the features proposed in this paper, we use the Radial Basis Function SVM. Notice that we report the performance obtained by Radial Basis Function SVM with the same parameters in all the seven fusion tests.

In Fig. 4, the DET-Curve of a single run of RS-ADA (green line), SVM (black line) and SUM (red line) are reported.

In Table 3, where we use a reduced training set, our methods obtain the best performance. In Fig. 5, the DET-curves obtained using the reduced training set are reported.

4. Conclusions

In this work, we have presented a feature extraction approach for the fusion of match scores in a multi-biometric system based on the likelihood ratio test. We show that densities estimated using a mixture of Gaussian models can be used to train a machine learning classifier.

Based on these experiments, our conclusions are the following:

- The likelihood ratio based feature coupled with a Random Subspace of AdaBoost of neural networks achieves a low Equal Error Rate in several tests without parameter tuning for each dataset.
- Both SVM and LR work well in some datasets and not so well in other datasets; however, our best proposed method works well in all the tested datasets.

As future work we want to study whether the incorporation of the Biometric sample quality information (as in Nandakumar et al. (2008)) within the likelihood ratio based fusion framework, improves performance in the proposed systems.

References

Bigun, E. S., Bigun, J., Duc, B., & Fischer, S. (1997). Expert conciliation for multi modal person authentication systems by Bayesian statistics. In *Proceedings of AVBPA* (pp. 291–300).

Brunelli, R., & Falavigna, D. (1995). Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17*(10), 955–966.

Cappelli, R. (2004). SFinGe: An approach to synthetic fingerprint generation. In *Proceedings of the international workshop on biometric technologies (BT2004)* (pp. 147–154).

Connie, T., Jin, A. T. B., Ong, M. G. K., & Ling, D. N. C. (2005). An automated palmprint recognition system. *Image and Vision Computing, 23*, 501–515.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (2nd ed.). New York: Wiley.

Fierrez-Aguilar, J., Garcia-Romero, D., Ortega-Garcia, J., & Gonzalez-Rodriguez, J. (2004). Exploiting general knowledge in user-dependent fusion strategies for multimodal biometric verification. In *Proceedings of ICASSP* (pp. 617–620).

Fierrez-Aguilar, J., Nanni, L., Ortega-Garcia, J., Cappelli, R., & Maltoni, D. (2005). Combining multiple matchers for fingerprint verification: A case study in FVC2004. In *Proceedings of the 13th international conference on image analysis and processing (ICIAP2005), Cagliari*.

Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J., & Bigun, J. (2005). Discriminative multimodal biometric authentication based on quality measures. *Pattern Recognition, 38*(5), 777–779.

Figueiredo, M., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(3), 381–396.

Griffin, P. (2004). Optimal biometric fusion for identity verification, Identix Research, Technical Report RDNJ-03-0064.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(8), 832–844.

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

Kittler, J., Hatef, M., Duin, R. P., & Matas, J. G. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(3), 226–239.

Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer.

Li, J. Q., & Barron, A. (1999). Mixture density estimation. In S. A. Solla, T. K. Leen, & K.-R. Muller (Eds.). *Advances, neural information processings systems* (Vol. 12). San Mateo, USA: Morgan Kaufman Publishers.

Lumini, A., & Nanni, L. (2006). When fingerprints are combined with iris – A case study: FVC2004 and CASIA. *International Journal of Network Security, 3*(2), 317–324.

Ma, Y., Cukic, B., & Singh, H. (2005). A classification approach to multi-biometric score fusion. In *Proceedings of fifth international conference on AVBPA, Rye Brook, USA* (pp. 484–493).

Maio, D., Maltoni, D., Jain, A. K., & Prabhakar, S. (2003). *Handbook of fingerprint recognition*. New York: Springer.

Maio, D., & Nanni, L. (2006). Combination of different fingerprint systems: A case study FVC2004. *Sensor Review, 26*(1), 51–57.

Martin, A. et al. (1997). The DET curve in assessment of decision task performance. In *Proceedings of EuroSpeech* (pp. 1895–1898).

Nandakumar, K., Chen, Y., Dass, S. C., & Jain, A. K. (2008). Likelihood ratio based biometric score fusion. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 342–347.

Nanni, L. (2006). Experimental comparison of one-class classifiers for on-line signature verification. *Neurocomputing, 69*(7–9), 869–873.

Nanni, L., & Lumini, A. (2007). RegionBoost learning for 2D + 3D based face recognition. *Pattern Recognition Letters, 28*(15), 2063–2070.

Nanni, L., & Lumini, A. (2008). Over-complete feature generation and feature selection for biometry. *Expert Systems With Applications, 35*(4), 2049–2055.

National Institute of Standards and Technology. (2004). NIST biometric scores set – Release 1. Available at: <http://www.itl.nist.gov/iad/894.03/biometricscores>.

Prabhakar, S., & Jain, A. K. (2002). Decision-level fusion in fingerprint verification. *Pattern Recognition, 35*(4), 861–874.

Rakhlin, A., Panchenko, D., & Mukherjee, S. (2005). Risk bounds for mixture density estimation. *ESAIM: Probability and Statistics, 9*, 220–229.

Ross, A., Nandakumar, K., & Jain, A. K. (2006). *Handbook of multibiometrics*. Springer-Verlag.

Snelick, R., Uludag, U., Mink, A., Indovina, M., & Jain, A. K. (2005). Large scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on PAMI, 27*(3), 450–455.

Toh, K.-A., Jiang, X., & Yau, W.-Y. (2004). Exploiting global and local decisions for multimodal biometrics verification. *IEEE Transactions on Signal Processing, (Supplement on Secure Media), 52*(10), 3059–3072.

Ulery, B., Hicklin, A. R., Watson, C., Fellner, W., & Hallinan, P. (2006). Studies of biometric fusion, NIST, Technical Report IR 7346.