ROBUST AUTHENTICATION USING LIKELIHOOD RATIO BASED SCORE FUSION OF VOICE AND FACE

Messaoud Bengherabi, Lamia Mezai, Farid Harizi
Division Architecture des Systèmes et Multimédia, Centre de Développement des Technologies Avancées
Cité 20 Aout, BP 11, Baba Hassen, Algiers, Algeria

Abderrazak Guessoum
Université Saad Dahlab de Blida, Route De Soumaa BP 270 BLIDA, Algeria

Mohamed Cheriet
École des Technologies Supérieur, 1100, Rue Notre-Dame Ouest, Montréal H3C1K3, Canada

Keywords: Biometrics, Score fusion, Face, Voice, Likelihood ratio, GMM.

Abstract: With the increased use of biometrics for identity verification, there has been a similar increase in the use of multimodal fusion to overcome the limitations of unimodal biometric systems. While there are several types of fusion (e.g. decision level, score level, feature level, sensor level), research has shown that score level fusion is the most effective in delivering increased accuracy. Recently a promising framework for optimal combination of match scores based on the likelihood ratio test is proposed; where the distributions of genuine and impostor match scores are modelled as finite Gaussian mixture model. In this paper, we examine the performance of combining face and voice biometrics at the score level using the LR classifier. Our experiments on the publicly available scores of the XM2VTS Benchmark database show a consistent improvement in performance compared to the famous efficient sum rule preceded by Min–Max, z-score and tanh score normalization techniques.

1 INTRODUCTION AND MOTIVATION

Nowadays, biometric verification systems based on face images and/or speech signals have been shown to be quite effective in various security applications such as local or distant secure access, identity check at an airport, , forensics ...etc. However, their performance easily degrades in the presence of a mismatch between training and testing conditions. For speech based systems this is usually in the form of channel distortion and/or ambient noise; for face based systems it can be in the form of a change in the illumination direction, varying pose, occlusion, non-uniform background, etc. In order to achieve better recognition performance and to overcome other limitations of unimodal biometric systems; information fusion from multiple biometric systems has already been the subject of an intensive research (Ross et al., 2006), (Toh et al., 2004). Multibiometric systems are categorized into four system architectures according to the strategies used for information fusion: at the sensor, feature extraction, matching score and decision levels (Ross and Jain, 2003).

The score level fusion is generally preferred because of its good performance and simplicity (Alsaade, 2008). Combining match scores is a challenging task because the scores of different matchers don't have the same nature and scale. According to (Nandakumar et al., 2007), score fusion techniques can be divided into the following three categories: transformation-based score fusion (Jain et al., 2005), (Snelick, et al., 2005), classifier-based score fusion (Ma et al., 2005), (Fierrez-Aguilar et al., 2003) and density-based score fusion (Dass et al., 2005), (Nandakumar, 2008), the last

Bengherabi M., Mezai L., Harizi F., Guessoum A. and Cheriet M. (2009).
ROBUST AUTHENTICATION USING LIKELIHOOD RATIO BASED SCORE FUSION OF VOICE AND FACE.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 57-61
DOI: 10.5220/0002232200570061

category is based on the likelihood ratio test and it requires explicit estimation of genuine and impostor match score densities. Density based approach followed by a classifier based on the Neyman-Pearson theorem (Lehmann and Romano, 2005) has the advantage that it directly achieves optimal performance at any desired operating point (FAR), provided the score densities are estimated accurately.

The authors in (Nandakumar et al., 2007) highlight that finite Gaussian mixture model (GMM) is quite effective in modelling the genuine and impostor score densities and the likelihood ratio based fusion rule with GMM-based density estimation achieves consistently low verification errors rates without the need for parameter tuning by the system designer, and they conclude their work by saying “while other fusion schemes such as sum rule and SVM can provide performance comparable to that of LR fusion, these approaches require careful selection of parameters (e.g., score normalization and fusion weights in sum rule, type of kernel and kernel parameters in SVM) on a case-by-case basis”.

However, their tests on the XM2VTS database (Poh and Bengio, 2006) were restricted only to the fusion of the best face and voice matchers, although we have a total of 8 matchers (5 for face and 3 for voice) yielding to a total of 15 bimodal combinations.

In this paper, we examine the performance of combining face and voice biometrics at the score level using the LR classifier and a finite Gaussian mixture model (GMM) in modelling the genuine and impostor score densities. The tests are done for all the 15 possible combinations with different GMM model orders. The results are compared with the famous efficient sum rule preceded by Min–Max, z-score and tanh score normalization techniques.

This paper is organized as follows: in section 2 we review the likelihood ratio based score fusion using GMM. Section 3 is dedicated to the elaboration and analysis of experimental results. Finally in the last section we conclude this work and highlight a possible perspective.

2 OVERVIEW OF LIKELIHOOD RATIO BASED SCORE FUSION

The Likelihood Ratio Test (LRT) has been used in fusion by many researchers (Nandakumar et al., 2007). Let be a random variable denoting the match score provided by a matcher. Let the distribution function for the genuine scores be denoted as $P_{gen}(s)$ (i.e., P(S≤s|S is genuine)=$Pgen(s)$) with the corresponding density function $p_{gen}(s)$. Similarly, let the distribution function for the impostor scores be denoted as $Pimp(s)$ with the corresponding density function $p_{imp}(s)$. Suppose we need to decide between the genuine and impostor classes (to verify a claimed identity) based on the observed match score s. The likelihood ratio criterion can be expressed as:

$$L(s) = \frac{p_{gen}(s)}{p_{imp}(s)} > threshold \tag{1}$$

The likelihood ratio method avoids the priors of the genuine and the impostor required by the Bayesian decision method, which are hard to estimate or to guess in reality. Instead it calculates the ratio and then thresholds it according to certain performance criterion such as false accept rate (FAR) or false reject rate (FRR). It can be formally proved that the likelihood ratio criterion is optimal in the Neyman-Pearson sense, i.e., when the FAR is fixed, the likelihood ratio criterion minimizes the FRR, and vice versa.

Assuming that both the genuine class and the impostor class have a mixture of Gaussians distribution, as expressed by

$$p(s) = \sum_{j=1}^{M} p_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \exp\left(-\frac{(s-\mu_j)^T |\Sigma_j^{-1}| (s-\mu_j)}{2} \right) \tag{2}$$

where s is match score vector, d is its dimensionality, μ is the mean vector, Σ is the covariance matrix and M is the number of mixture. Introducing logarithm, the criterion in Eq(1) can be rewritten:

$$ln[L(s)] = ln[p_{gen}(s)] - ln[p_{imp}(s)] > threshold \tag{3}$$

In our study, the operating threshold used for performance comparison is the equal error rate obtained when the false accept rate (FAR) is equal to the false reject rate (FRR). From equation (3) we can remark that for a single multidimensional Gaussian the logarithm essentially reduces the probability measure to the difference between the two squared Mahalanobis distances in the genuine and the impostor class.

3 EXPERIMENTAL RESULTS

3.1 The XM2VTS Database and the Lausanne Protocols

The performance of likelihood ratio based fusion rule was evaluated on the score of the XM2VTS Benchmark database available from the website (http://personal.ee.surrey.ac.uk/Personal/Norman.Poh/), (Poh and Bengio, 2006). This database contains synchronised video and speech data from 295 subjects, recorded during four sessions taken at one month intervals. The database is divided into three sets: a training set, an evaluation set and a test set. The training set (LP Train) was used to build client models, while the evaluation set (LP Eval) was used to compute the decision thresholds used by classifiers. Finally, the test set (LP Test) was used to estimate the performance.

The 295 subjects were divided into a set of 200 clients, 25 evaluation impostors and 70 test impostors. There exist two configurations or two different partitioning approaches of the training and evaluation sets. They are called Lausanne Protocol I and II, denoted as LP1 and LP2 (Poh and Bengio, 2006), the description of the Lausanne Protocol is shown in Table 1. In this paper, we have used the Lausanne Protocol I (LP1).

3.2 Test Protocol

We have used combination of classifiers and face and speech features like in (Poh and Bengio, 2006). So we have 15 possible combinations. In the fusion based on Likelihood ratio, we have varied the number of mixtures to estimate the density of impostor and genuine. We have used 1, 2, 4 and 8 mixtures. The simple sum rule preceded with the min-max and tanh normalization methods (Snelick et al., 2005) is used for the aim of comparison.

The min-max normalization method maps the score to the [0, 1] range, the quantities *Smax* and Smin specify the end points of the score range (Snelick et al., 2005) and *Sn* (the normalized score) is given by:

$$S_n = \frac{S - S_{min}}{S_{max} - S_{min}} \tag{4}$$

where *Smin*=min(*s1, ..., sK*) and *Smax*=max(*s1, ..., sK*).

On other hand the hyperbolic tangent (Tanh) is a robust statistical method which maps the scores to the [0, 1] range (Snelick et al., 2005):

$$S_n = \frac{1}{2}\left[tanh\left(0.01 \frac{\left(S - mean\left(S_{gen}\right)\right)}{std\left(S_{gen}\right)} \right) + 1 \right] \tag{5}$$

where *std* stands for the standard deviation and gen for the genuine scores (it was proven via experiments that it is better to use the genuine scores rather than both the genuine and impostor scores). We have also compared the Likelihood ratio with the work of (Poh and Bengio, 2006) in which, he have used the simple sum rule with z-score normalization.

3.3 Performance Evaluation

The Half Total Error Rate (HTER) (Poh and Bengio, 2006) of the likelihood ratio based fusion, simple sum rule using min-max and z-score and tanh normalization techniques is used to compare the performance of the different fusion techniques. Note that the HTER is defined as:

Δ^* is the optimal threshold that minimizes the Error Equal Rate (EER) on a development set. It can be

$$HTER\left(\Delta^*\right) = \frac{FAR\left(\Delta^*\right) + FRR\left(\Delta^*\right)}{2} \tag{6}$$

Table 1: Description of Lausanne Protocols.

		Lausanne Protocol I			Lausanne Protocol II		
		Number of subjects	Number of recording per subject	Number of Scores	Number of subjects	Number of recording per subject	Number of Scores
Training set	Clients	200	3	600	200	4	800
	Impostors	/	/	/	/	/	/
Evaluation set	Clients	200	3	600	200	2	400
	Impostors	25	8	40000	25	8	40000
Test set	Clients	200	2	400	200	2	400
	Impostors	70	8	112000	70	8	112000

Table 2: Comparison of the HTER between the likelihood ratio based fusion and the simple sum rule.

No.	Fusion candidates	Face	Voice	(log-likelihood ratio)				Simple sum rule		
				1	2	4	8	zscore	Min-max	Tanh
1	(FH,MLP)(LFCC,GMM)	1,883	1,148	1,108	0,426	0,565	**0,297**	0,795	0,862	0,737
2	(FH,MLP)(PAC,GMM)	1,883	6,208	1,441	1,097	**0,992**	1,079	1,133	1,161	1,026
3	(FH,MLP)(SSC,GMM)	1,883	4,494	1,339	1,054	0,962	0,963	0,868	1,072	**0,778**
4	(DCTs,GMM)(LFCC,GMM)	4,250	1,148	0,574	0,571	0,575	0,568	0,526	**0,492**	0,583
5	(DCTs,GMM)(PAC,GMM)	4,250	6,208	1,417	**1,331**	1,428	1,422	1,436	1,417	1,376
6	(DCTs,GMM)(SSC,GMM)	4,250	4,494	1,201	1,197	1,152	1,155	1,144	1,218	**1,132**
7	(DCTb,GMM)(LFCC,GMM)	**1,734**	**1,148**	0,499	**0,476**	0,479	0,486	0,553	0,503	**0,467**
8	(DCTb,GMM)(PAC,GMM)	1,734	6,208	1,106	1,087	1,068	**1,066**	1,127	1,093	1,661
9	(DCTb,GMM)(SSC,GMM)	1,734	4,494	0,764	0,747	0,849	0,841	0,747	**0,720**	0,733
10	(DCTs,MLP)(LFCC,GMM)	3,363	1,148	1,193	0,574	0,597	**0,575**	0,841	0,972	0,728
11	(DCTs,MLP)(PAC,GMM)	3,363	6,208	1,982	1,000	0,894	0,961	1,119	1,413	**0,822**
12	(DCTs,MLP)(SSC,GMM)	3,363	4,494	1,721	1,111	**0,909**	0,965	1,372	1,594	1,036
13	(DCTb,MLP)(LFCC,GMM)	6,225	1,148	1,693	0,719	**0,609**	0,682	1,621	3,278	0,874
14	(DCTb,MLP)(PAC,GMM)	6,225	6,208	3,547	2,579	**2,167**	2,410	3,653	4,121	2,623
15	(DCTb,MLP)(SSC,GMM)	6,225	4,494	3,722	2,038	**1,671**	1,831	2,883	4,329	2,058

Table 3: Comparison of the average of the HTER between the likelihood ratio based fusion and the simple sum rule.

	(log-likelihood ratio) Number of mixtures				Simple sum rule		
	1	2	4	8	zscore	Min-max	Tanh
Average of HTER of the 15 combinations	1,554	**1,067**	**0,994**	**1,020**	1,616	1,321	1,109

calculated as follows:

$$\Delta^* = \underset{\Delta}{arg\ min}\ EER(\Delta) \tag{7}$$

where

$$EER = \frac{1}{2}\left(FAR(\Delta) + FRR(\Delta)\right) \tag{8}$$

where FAR and FRR designate the false acceptance rate and false rejection rate respectively.

We can notice from table 2, that using LR test with only one Gaussian gives the worst results. This is expected because only one Gaussian is not sufficient to estimate efficiently the score distributions. However a consistent performance improvement is obtained by increasing the number of Gaussians to 4 where the best performance are abstained, good results are obtained with eight Gaussians but it is clear that 8 Gaussians are more than enough to estimate the client and impostor distributions and also this is due to the lack of data.

To summarize Table 2, we have computed the average HTER of the 15 possible matcher combinations, the results are summarized in Table 3. It is so clear from this table the superiority of the LR test using GMM for modelling the genuine and impostor classes. We can conclude that although the sum rule can obtain a better performance with an appropriate normalisation (min-max or tanh in our case) the gain compared to the LR is not significant.

4 CONCLUSIONS

In this paper, we have analyzed the performance of combining face and voice biometrics at the score level using the LR classifier. Our experiments on the publicly available scores of the XM2VTS Benchmark database show a consistent high performance regardless of the score nature of different speech and face matchers. As a perspective of this work is the introduction of user specific information jointly with the LR test and GMM score modelling.

REFERENCES

Alsaade, F. 2008. Score-Level Fusion for Multimodal

Biometrics. *Phd thesis, University of Hertfordshire, England*.

Dass, S. C., Nandakumar, K., Jain, A. K. 2005. A Principled Approach to Score Level Fusion in Multimodal Biometric Systems. *Lecture Notes in Computer Science proceedings of the Audio- and Video-Based Biometric Person Authentication conference AVBPA'*2005, 1049-1058 Springer Berlin / Heidelberg.

Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J. 2003. Fusion Strategies in Multimodal Biometric Verification. *Proceedings of the IEEE International Conference on Multimedia and Expo*, ICME '03, pp 5 – 8.

Jain, A., Nandakumar, K., Ross, A. 2005. Score normalization in multimodal biometric systems. *Pattern Recognition*, vol. 38, No. 12, pp. 2270-2285.

Lehmann, E. L., Romano, J. P. 2005. *Testing Statistical Hypotheses*. Springer.

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*, pp. 484–493.

Nandakumar, K. 2008, Multibiometric Systems: Fusion Strategies and Template Security. *Phd Thesis, Michigan State University, Department of Computer Science and Engineering*.

Nandakumar, K., Chen, Y., Jain, K. 2007. Likelihood Ratio Based Biometric Score Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

Poh N., Bengio, S. 2006. Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometric Authentication. *Pattern Recognition*, vol. 39, no. 2, pp. 223–233.

Ross, A., Jain, A. K. 2003. Information Fusion in Biometrics. *Pattern Recognition Letters*.

Ross, A., Nandakumar, K., Jain, A. K. 2006. *Handbook of Multibiometrics*.Springer-Verlag.

Snelick, R., Uludag, U., Mink, A., Indovina, M., Jain, A. 2005. Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, pp 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)*, vol. 52, no. 10, pp. 3059–3072.