Increase the security of multibiometric systems by incorporating a spoofing detection algorithm in the fusion mechanism

Emanuela Marasco*[1], Peter Johnson[2], Carlo Sansone[1] and Stephanie Schuckers[2]

[1] Dipartimento di Informatica e Sistemistica,
Università degli Studi di Napoli Federico II
Via Claudio, 21 I-80125 Napoli, Italy
{emanuela.marasco,carlosan}@unina.it
[2] Department of Electrical and Computer Engineering,
Clarkson University
PO Box 5720 Potsdam, NY 13699
{sschucke,johnsopa}@clarkson.edu

Abstract. The use of multimodal biometric systems has been encouraged by the threat of spoofing, where an impostor fakes a biometric trait. The reason lies on the assumption that, an impostor must fake all the fused modalities to be accepted. Recent studies showed that there is a vulnerability of the existing fusion schemes in presence of attacks where only a subset of the fused modalities is spoofed. In this paper, we demonstrated that, by incorporating a liveness detection algorithm in the fusion scheme, the multimodal system results robust in presence of spoof attacks involving only a subset of the fused modalities. The experiments were carried out by analyzing different fusion rules on the Biosecure multimodal database.

1 Introduction

A biological measurement can be qualified as a biometric if it satisfies basic requisite like universality, permanence, distinctiveness, circumvention. The last property concerns the possibility of a non-client being falsely accepted, typically by spoofing the biometric trait of an authorized user [1]. Previous works have shown that it is possible to spoof a variety of fingerprint technologies using spoof fingers made with materials as *Silicon*, *Play-Doh*, *Clay* and *Gelatin* (gummy finger) [2].

Multibiometric systems improve the reliability of the biometric authentication by exploiting multiple sources, such as different biometric traits, multiple samples, multiple algorithms. They are able to improve the recognition accuracy, to increase the population coverage, to offer user choice and to make biometric

* Emanuela Marasco is currently a post-doctoral candidate at the Lane Department of Computer Science and Electrical Engineering, West Virginia University, WV (USA).

authentication systems more robust to spoofing [3]. Several works in the literature on biometrics demonstrate the efficiency of the multimodal fusion to enhance the recognition accuracy of the unimodal biometric systems [4].

From a security perspective, a multimodal system appears more protected than its unimodal components. The reason is that, one assumes that an impostor must fake all the fused modalities to be accepted and spoofing multiple modalities is harder than spoofing only one [5]. However, a hacker may fake only a subset of the fused biometric traits. Recently, researchers demonstrated that the existing multimodal systems can be deceived also when only a subset of the fused modalities is spoofed [6]. Rodrigues *et al.* proposed an approach to measure the security of a multimodal system, where the contribution provided by each single modality matcher is weighted based on the ease to spoof that biometric trait. For example, the probability of success associated to a spoof attack is high in presence of a sample which gives a low match score. Johnson *et al.* [7] explored the multimodal vulnerability of the score level fusion strategies in a scenario where partial spoofing has occurred.

The goal of this paper is to propose an approach, based on liveness detection techniques, which can improve the security of multimodal biometric systems in presence of spoof attacks involving one fingerprint modality. We have analyzed the performance of different multibiometric systems in presence of partial spoofing when an effective spoofing detection algorithm is incorporated in the fusion mechanism. Our experiments showed that the proposed technique aids to increase the robustness of such systems with respect to the spoofing. In our approach the integration involves match scores, and the spoof attack is detected separately for each modality matcher before fusion. Thus, when a fake sample is detected by the algorithm, the unimodal output does not give any contribution in the fusion which results in a more secure decision.

The current analysis is carried out as a simulation to assess performance of multibiometric systems in presence of spoof attacks. The simulation makes the assumption that live match scores have a similar distribution with respect to spoof match scores. In future work, actual spoof data is needed to assess the performance in a real-world system. However, this simulation can provide a framework for assessing novel algorithms, as well as their relative performance.

The paper is organized as follows. Section 2 presents an overview of our approach, together with the combination rules we considered for our study and the liveness detection algorithm exploited in the fusion. Section 3 describes the adopted dataset and the experiments carried out on it, which show the effectiveness of the proposed technique. Section 4 draws our conclusions.

2 Our approach

In the current approach, we have analyzed the performance of different multibiometric systems in presence of spoof attacks involving one fingerprint modality, when an effective spoofing detection algorithm is incorporated in the fusion mechanism. The final multimodal decision is made by considering that, when a

spoofed sample is detected by the algorithm, the corresponding matcher does not give any contribution in the fusion scheme.

2.1 Score fusion rules

When designing a multibiometric system, several factors should be considered. These concern the choice and the number of biometric traits, the level of integration and the mechanism adopted to consolidate the information provided by multiple traits. Fusion at match score level is often chosen since it is easy to access and combine the scores presented by different modalities. The operators which do not contain parameters to be tuned, are known as *fixed* combiners [8]. Based on experimental results, researchers agree that *fixed* rules usually perform well for ensemble of classifiers having similar performance, while *trained* rules handle better matchers having different accuracies. When fusing different modalities, individual matchers often exhibit different performance, thus for this problem *trained* rules should perform better than *fixed* rules [9].

Transformation-based fusion The match scores provided by different matchers are firstly transformed into a common domain (*score normalization*), then they are combined using a fusion rule. It has been shown that the simple sum rule gives very good accuracy [9]. The technique adopted in our fusion framework is the *min-max*, which retains the original distribution of scores except a scaling factor and transform the scores to a common range from zero to one, based on the minimum and the maximum score values. Given a set of matching scores s_k, $k = 1 \dots K$, the normalized scores are given by (1).

$$s_k = \frac{s_k - min}{max - min} \tag{1}$$

The operator employed for the current analysis is the simple score sum, defined by (2)

$$s_{sum} = \sum_{k=1}^{N} \frac{1}{N} s_k \tag{2}$$

Density-based fusion The match scores are considered as random variables, whose class conditional densities are not *a priori* known [10]. So, this approach requires an explicit estimation of density functions from the training data [5]. The model is built by estimating density functions for the genuine and impostor score distributions [11]. A recent method, proposed by Nandakumar et al. in [12], is the framework based on the Likelihood Ratio test, where the scores are modeled as mixture of Gaussians and a statistical test $\Psi(\mathbf{s})$ is performed to discriminate between genuine and impostor classes. The Gaussian Mixture Model (GMM) lets to obtain reliable estimations of the distributions, even if the amount of data needed for it increases as the number of considered biometrics increases. This framework produces high recognition rates at a chosen operating

point (in terms of False Acceptance Rate), when it is possible to perform accurate estimations of the genuine and impostor score densities.
Let $\mathbf{s} = [s_1, s_2, ..., s_K]$ denote the scores emitted by multiple matchers, with s_k representing the match score of the k_{th} matcher, $k = 1, ..., K$.

$$\Psi(\mathbf{s}) = \begin{cases} 1, & \text{when } LR(\mathbf{s}) \geq \eta \\ \\ 0, & \text{when } LR(\mathbf{s}) < \eta \end{cases} \tag{3}$$

where $\mathbf{s} = [s_1, s_2, ...s_K]$ is an observed set of K match scores that is assigned to the genuine class if $LR(\mathbf{s})$ is greater than a fixed threshold η, with $\eta \geq 0$.

2.2 Spoofing detector

Our multimodal fusion approach is evaluated assuming that *fake-live* match scores are similarly distributed as *live-live* match scores. For each modality, the spoof attack was simulated by substituting a genuine match score in place of an impostor match score. The multimodal system considered in this paper is composed by face and fingerprint traits and it is analyzed under normal operation (i.e., without spoofing), and when only a fingerprint trait is spoofed.

In all the scenarios, a fingerprint liveness detection is integrated in the fusion scheme. In this investigation, we incorporate the performance, known by the literature, of a liveness algorithm which combines perspiration- and morphology-based static features [13]. The classification performance of the adopted algorithm was evaluated by using the parameters of the Liveness Detection Competition 2009 (LivDet09) [14], defined as follows:

- *Ferrlive*: rate of misclassified live fingerprints.
- *Ferrfake*: rate of misclassified fake fingerprints.

In particular, the values of $Ferrlive$ and $Ferrfake$ were averaged on the three databases (*Biometrika*, *CrossMatch* and *Identix*) that compose the LivDet09 data. On such data, the liveness algorithm exploited in our approach presented an average $Ferrlive$ of 12.60% and an average $Ferrfake$ of 12.30% [13]. We used the performance obtained on these three different databases taken from LivDet09 as an estimate of the actual performance of the algorithm on the database used in this paper.

When a spoofed modality is detected by the incorporated algorithm, it will not give any contribution to the final decision. In particular, (100%-$Ferrfake$) indicates the percentage of correctly detected live-spoof match scores to be excluded from the combination, and $Ferrlive$ indicates the percentage of wrongly detected live-live match scores to be excluded from the combination. In the proposed approach, live genuine match scores are employed for a real live genuine scenario, where FRR can be assessed, and also in place of impostor match scores in order to simulate spoofing.

- In the score sum scheme, this is realized by resetting a percentage of (100%-$Ferrfake$) impostor scores substituted by genuine scores, and a percentage of $Ferrlive$ genuine scores before performing the sum.

– In the likelihood ratio scheme, the detected spoofed modality can be marginalized by employing, for the (100%-$Errfake$) of the fake samples and for the $Errlive$ of the live samples, the join density functions involving only the live modalities.

3 Experimental Results

3.1 Dataset

The performance of the proposed strategy was evaluated on a subset of the BioSecure multimodal database. This database contains 51 subjects in the Development Set (training) and 156 different subjects in the Evaluation Set (testing). For each subject, four biometric samples are available over two sessions: session 1 and session 2. The first sample of each subject in the first session was used to compose the gallery database while the second sample of the first session and the two samples of the second session were used as probes (P_1, P_2, P_3). For the purpose of this study, we have employed one face and three fingerprint modalities, denoted as *fnf*, *fo1*, *fo2* and *fo3*, respectively [15]. The scores used in our experiments are the output of the matching between the first available sample and the second one for each subject. Our second dataset consists in an unbalanced population composed by 516 genuine and 24,180 (156*155) impostor match scores. The details are reported in Tables 1 and 2.

Table 1. The Biosecure database: Development Set

Biometric	Subjects	Samples	Scores
Face	51	4 per subject	Gen 204×3 Imp $51 \times 50 \times 16$
Fingerprint	51	4 per subject	Gen $(204 \times 3) \times 3$ Imp $(51 \times 50 \times 16) \times 3$

Table 2. The Biosecure database: Evaluation Set

Biometric	Subjects	Samples	Scores
Face	156	4 per subject	Gen 624×3 Imp $156 \times 155 \times 16$
Fingerprint	156	4 per subject	Gen $(624 \times 3) \times 3$ Imp $(156 \times 155 \times 16) \times 3$

3.2 Results

The evaluation of the multibiometric system is carried out by adopting the metric denoted as Spoof False Accept Rate (SFAR) which corresponds to a percentage of times a spoof attack results in success. In this paper, a successful spoof attack is when the sum of match score (in the case of sum rule) is above the threshold when a partial spoof attack has occurred (substitution of genuine score for imposter scores). Such a metric has been introduced in [7] to distinguish from traditional FAR. The complete performance curve which represents the full capabilities of the system at different operating points, is given by the Detection Error Tradeoff (DET) in which FAR is a function of FRR/SFAR obtained using logarithmic scales on both axes.

Table 3 reports our results averaged on 20 iterations where for each iteration the fake samples detected by the algorithm has been randomly varied.

Table 3. Results on Biosecure database in different scenarios where one fingerprint modality is spoofed

Fusion rule	Fused modalities	Spoofed modality	EER No spoof	SFAR 1 spoofed modality	SFAR with algorithm	FRR
sum	3 fg + 1 face	fo1	0.32%	56.36%	28.98%	0.32%
sum	3 fg + 1 face	fo2	0.32%	18.67%	5.67%	0.32%
sum	3 fg + 1 face	fo3	0.32%	9.01%	0.46%	0.32%
avg sum	4 mod	1 fg	0.32%	28.01%	11.70%	0.32%
LR	3 fg + 1 face	fo1	0.004%	91.37%	11.12%	0.004%
LR	3 fg + 1 face	fo2	0.004%	62.47%	0.18%	0.004%
LR	3 fg + 1 face	fo3	0.004%	56.83%	8.81%	0.004%
avg LR	4 mod	1 fg	0.004%	70.22%	6.70%	0.004%

In a multimodal system based on the sum of scores with four modalities, three fingerprints and one face, the EER point fixed on the curve without spoofing corresponds to 0.32%, while for this value of FRR, when the fingerprint $fo3$ is spoofed, SFAR becomes equal to 9.01% (see Fig.1); while incorporating in the fusion the fingerprint liveness detection algorithm, SFAR significantly decreases to a value of 0.46%. See note for Figure 1.

In a multimodal system based on the likelihood ratio involving three fingerprint and one face modalities, the EER point fixed on the curve without spoofing, corresponds to 0.004%, while for this value of FRR, when the fingerprint $fo1$ is spoofed, SFAR becomes equal to 91.37% (see Fig.2 notes); while incorporating in the fusion the fingerprint liveness detection algorithm, SFAR significantly decreases to a value of 11.12%. When $fo2$ is the fingerprint spoofed, SFAR increases to 62.47%, but the error rate can be reduced by introducing the algorithm until a percentage of 0.17%.

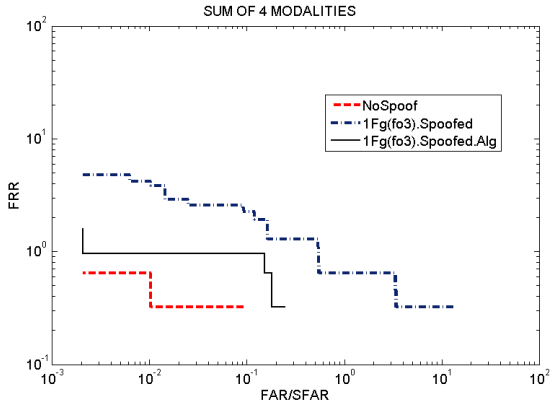

Fig. 1. DET curve of the score sum of three fingerprint and one face modalities taken from Biosecure database over 20 iterations, where one fingerprint is spoofed. Both vertical and horizontal axis of the plot is logarithmically scaled.

3.3 Finding the best error rate at spoof detection level

We have extended our investigation by experimentally analyzing how multimodal performance can improve when reducing $Ferrlive$ and $Errfake$, starting from the values of 12.60% and 12.30% respectively, we used in our previous experiments (see Fig.3 and Fig.4).

This step aids to understand which is the best trade-off between the error rate required to a liveness detection algorithm and the fusion performance achieved after incorporating it in the combination scheme. Results are reported in Table 4. The benefits obtained by incorporating the algorithm in the fusion mechanism

Table 4. Results on Biosecure database by varying the error rate of the liveness detection algorithm. In the plot x%-x% indicates the percentage of $Ferrlive$ and $Errfake$.

Fusion rule	Fused modalities	Modality spoofed	SFAR 1%-1%	SFAR 2%-2%	SFAR 5%-5%	SFAR 7%-7%	SFAR 9%-9%	SFAR 12.60%-12.30%
sum	3 fg + 1 face	fo1	3.87%	5.83%	12.34%	18.93%	26.84%	28.98%
sum	3 fg + 1 face	fo2	0.99%	1.95%	2.34%	3.64%	4.78%	5.67%
sum	3 fg + 1 face	fo3	0.05%	0.08%	0.10%	0.15%	0.38%	0.46%
avg LR	4 mod	1 fg	1.64%	2.62%	4.93%	7.53%	10.67%	11.70%
LR	3 fg + 1 face	fo1	1.02%	1.03%	4.71%	4.81%	8.38%	11.11%
LR	3 fg + 1 face	fo2	0.11%	0.13%	0.13%	0.16%	0.18%	0.18%
LR	3 fg + 1 face	fo3	1.51%	2.28%	3.93%	5.49%	6.79%	8.81%
avg LR	4 mod	1 fg	0.88%	1.15%	2.92%	3.48%	5.12%	6.70%

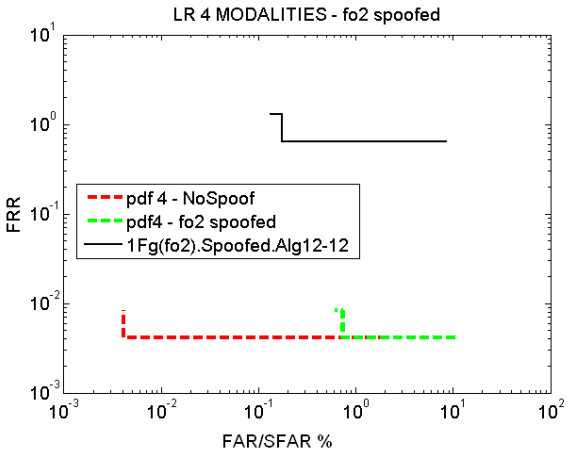

Fig. 2. DET curve of the likelihood ratio involving three fingerprint and one face modalities taken from Biosecure database over 20 iterations, where one fingerprint is spoofed.

change by varying the fusion rule. Regarding the LR-based mechanism, the benefits obtained by incorporating the algorithm in the fusion are more significant; in particular, $SFAR$ can be reduced to the value of 0.88% when the spoofing is detected by an algorithm with $Ferrlive$ and $Errfake$ both equal to 1.00%.

4 Conclusions and future directions

In this paper, we have analyzed the performance of the most efficient fusion approaches at score level under spoof attacks which involve only one fingerprint modality. We have considered a multimodal biometric system in presence of a worst case spoof attack, where the *fake-live* match score distribution is assumed to coincide with the *live-live* match score distribution. Previous works and the results here showed that, when only a subset of the fused modalities is spoofed, multimodal systems can be deceived. Our experiments also demonstrated that a more robust fusion can be realized by incorporating a fingerprint liveness detection algorithm in the combination scheme. Further, we have reduced the error at spoof detection level and found the best trade-off between the optimal $Ferrlive$ and $Errfake$ values and the multimodal performance.

This paper considers the case where spoofing is simulated by substituting with genuine scores. One limitation of the proposed approach lies on the assumption that spoof match scores are distributed as live match scores. Since spoofing is difficult, it may be that the spoof match score distribution has a mean match score which is lower. Therefore, this simulation could be considered as a worst case scenario. Incorporating a spoofing detection, even if it improves

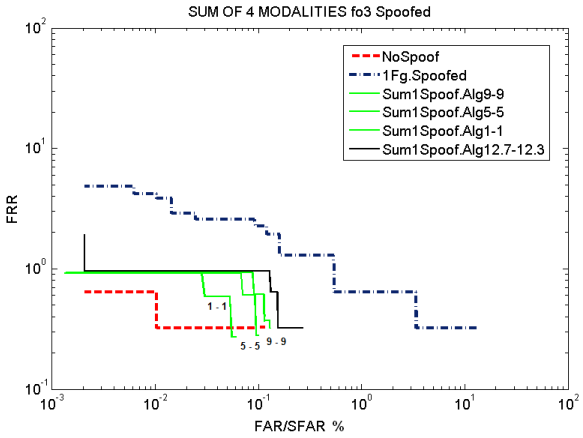

Fig. 3. Performance of the score sum between two fingerprint modalities when one fingerprint is spoofed by varying the $Ferrlive$ and $Errfake$ of the liveness detection algorithm incorporated in the fusion.

FAR under spoof attacks, could have a significant impact on the FRR, as we showed in the case of likelihood ratio-based scheme.

As a future step in this research, the experiments will be extended to additional multimodal databases. A number of fusion algorithms will also be collected and compared using the methods outlined in this paper. Moreover, the performance of the proposed approach will be evaluated by employing real spoofed data.

References

1. A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transaction on Circuits and Systems for Video*, 14(1):4–20, January 2004.
2. K. Yamada T. Matsumoto, H. Matsumoto and S. Hoshino. Impact of artificial gummy fingers on fingerprint systems. *Optical Security and Counterfait Deterrence Techniques IV*, 4677:275–289, January 2002.
3. J. Kittler, Y. P. Li, J. Matas, and M. U. R. Sanchez. Combining evidence in multimodal personal identity recognition systems. *International Conference on Audio- and Video-based Biometric Person Authentication*, 1997.
4. A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters 24*, pages 2115–2125, 2003.
5. A. Ross and A. Jain. *Handbook in MultiBiometrics*. Springer, 2008.
6. R. N. Rodrigues, N. Kamat, and V. Govindaraju. Evaluation of biometric spoofing in a multimodal system. *IEEE International Conference on Biometrics (BTAS)*, 2010.
7. P. A. Johnson, B. Tan, and S. Schuckers. Multimodal fusion vulnerability to non-zero effort (spoof) imposters. *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2010.

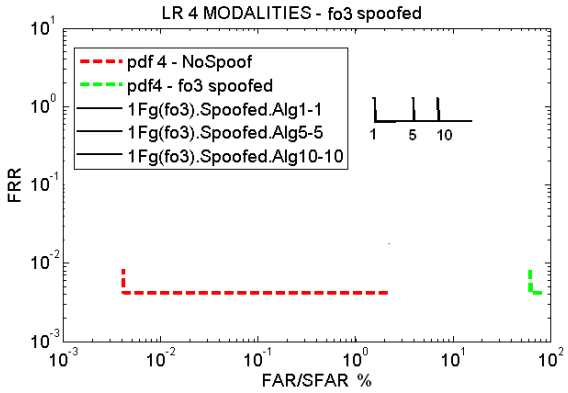

Fig. 4. Performance of the likelihood ratio when one fingerprint is spoofed by varying the $Ferrlive$ and $Ferrfake$ of the liveness detection algorithm incorporated in the fusion.

8. N. Poh. *École Polytechnique Fédéral de Lausanne*. Multi-system biometric authentication : optimal fusion and user-specific information, 2006.
9. F. Roli, J. Kittler, G. Fumera, and D. Muntoni. An experimental comparison of classifier fusion rules for multimodal personal identity verification systems. *Proc. Multiple Classifier Systems, Sringer-Verlag*, 2364:325–336, 2002.
10. S. Dass, K. Nandakumar, and A. Jain. A principled approach to score level fusion in multimodal biometric systems. *Fifth AVBPA*, pages 1049–1058, July 2005.
11. M. Vatsa, R. Singh, A. Noore, and A. Ross. On the dynamic selection of biometric fusion algorithms. *IEEE Transaction on Information Forensics and Security*, 5(3):470–479, 2010.
12. K. Nandakumar, Y. Chen, S. Dass, and A. Jain. Likelihood ratio-based biometric score fusion. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 30(2):342–347, February 2008.
13. E. Marasco and C. Sansone. An anti-spoofing technique using multiple textural features in fingerprint scanners. *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMs)*, pages 8–14, 2010.
14. G. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers. First international fingerprint liveness detection competition - livdet 2009. *Lecture Notes in Computer Science*, 5716:12–23, August 2009.
15. N. Poh, T. Bourlai, and J. Kittler. A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms. *Pattern Recognition*, 43:1094–1105, 2010.