Towards anomaly detection for increased security in multibiometric systems: spoofing-resistant 1-median fusion eliminating outliers

Towards Anomaly Detection for Increased Security in Multibiometric Systems: Spoofing-resistant 1-Median Fusion Eliminating Outliers

Peter Wild, Petru Radu, Lulu Chen and James Ferryman
Computational Vision Group, School of Systems Engineering, University of Reading, United Kingdom
{p.wild | p.radu | l.chen | j.m.ferryman}@reading.ac.uk

Abstract

Multibiometrics aims at improving biometric security in presence of spoofing attempts, but exposes a larger availability of points of attack. Standard fusion rules have been shown to be highly sensitive to spoofing attempts – even in case of a single fake instance only. This paper presents a novel spoofing-resistant fusion scheme proposing the detection and elimination of anomalous fusion input in an ensemble of evidence with liveness information. This approach aims at making multibiometric systems more resistant to presentation attacks by modeling the typical behaviour of human surveillance operators detecting anomalies as employed in many decision support systems. It is shown to improve security, while retaining the high accuracy level of standard fusion approaches on the latest Fingerprint Liveness Detection Competition (LivDet) 2013 dataset.

1. Introduction

Multibiometrics involves the use of biometric fusion techniques to combine evidence from multiple sources [16], including sensors, modes, algorithms or instances, aiming to (1) reduce biometric false acceptance and false rejection error rates, (2) minimize failure-to-acquire/capture/enrol rates (3) increase throughput (e.g., using fast indexing methods), and (4) counter spoofing attempts for single biometric characteristics. However, depending on the type of fusion applied, multibiometrics may cause increased exposure to attacks involving single specific characteristics. A spoof attack is an attempt to circumvent the system with the presenter of the biometric simulating the trait of a different claimed identity, e.g. by using an artificial gelatin finger made from latent fingerprints. Even a single spoofed instance may lead to a false accept in multimodal (combining modalities, like face and fingerprint) and more generic multibiometric systems [15, 14, 3, 1]. Despite spoofing sensitivity, it is reasonable to assume that it is typically more difficult for an attacker to obtain multiple biometric samples of a target identity to be claimed. It is therefore important to improve the trade-off between cost and security to limit drawbacks [10, 9, 13]. This is the aim of this paper.

In order to address the spoofing issue, liveness detection methods [4, 5] have been proposed using additional data to assess if the biometric is authentic. In multibiometric configuration, obtained liveness measures may be considered separately for each biometric evidence (modality, sensor or unit) or jointly in the fusion scheme. In the first case, spoof-detected evidence is typically excluded from fusion [10]. If liveness detection is unavailable, the relative robustness of several score-level fusion rules can be used to choose the most robust fusion rule [1], but in either case information may not be exploited in the most efficient manner. This paper focuses on the second type of targeting spoofing attempts in the fusion scheme under the assumption that a subset of combined biometric evidence is spoofed. Presented methods aim at detecting such anomalies at score-level taking optionally additional liveness of individual evidence into account (see Fig. 1).

The contributions of this paper are as follows: (1) a thorough evaluation of spoofing impact on fingerprint fusion, (2) the investigation of lightweight parameterless (no training) 1-median fusion using score-values only for increased robustness versus spoofing attacks, (3) the proposal and evaluation of a novel fusion scheme based on 1-median filtering combining scores with liveness metadata for high robustness versus non-zero-effort (i.e. with access to fake biometric samples, in contrast to zero-effort without spoofing) impostor attacks, using the latest Fingerprint Liveness Detection Competition (LivDet) 2013 dataset.

This paper is organized as follows: Section 2 introduces related work on spoofing impact on multibiometric system security. Section 3 describes the proposed fusion scheme presenting both, a score-only fusion method and a fusion method taking additional meta-information into account. The system under test, the adopted dataset, experimental configuration and obtained results showing the effectiveness of the proposed technique are highlighted in Section 4. Section 5 forms the conclusion.

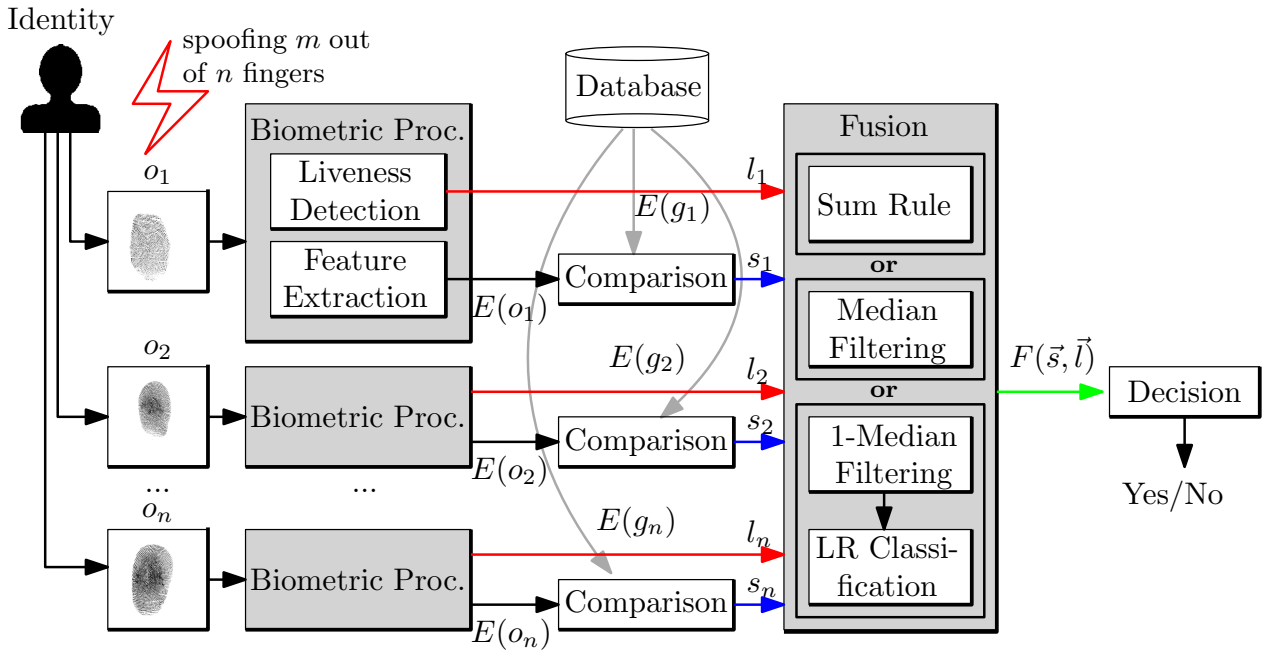

Figure 1. Proposed fusion architecture combining comparison scores s_i and liveness values l_i from observations o_i and gallery samples g_i.

2. Related Work

The task of combining scores from presentation attacks and scores from biometric modalities or multiple instances is still not standardized, and an active research topic [6, 9].

Several previous research projects have indicated the vulnerability of multimodal biometric systems against single or multiple biometric spoofing attacks [15, 14, 3, 1]. Akhtar et al. [2] conducted the first study to examine the vulnerability of both parallel and serial fusion in bimodal (face and fingerprint) configuration against spoofing attacks. Their results indicated that score-level fusion methods from the literature are not robust to spoofing attacks, because both fusion modes can be fooled by a single biometric. However, serial fusion gave better results, attaining a better trade-off between performance, verification time, user acceptability and robustness.

To enhance the security of a biometric fusion system, several fusion schemes have recently been proposed. Rodrigues et al. [15] first addressed the security issue of spoofing attacks against a multimodal biometric system. They introduced two fusion schemes, by adding security parameters of each unimodal biometric system with match scores and quality scores. The two schemes were based on extended likelihood ratio (LLR) and fuzzy logic. Fuzzy logic gave better overall performance in their experiment. An increased robustness against spoofing attacks, compared with traditional fusion rules, was indicated in their experiment results.

More recently, Rattani et al. [13] claimed that image quality, liveness measurement and match scores are normally influenced by the sensor. The authors developed a learning-based fusion framework using a graphical model by adding impact parameters for the sensor. Marasco et al. [10] directly incorporated liveness detection with match scores in their fusion scheme. The liveness detection is implemented separately for each modality before fusion. If a spoofing attempt is indicated, the current modality matching score will be ignored in the fusion. In another work, Marasco et al. [9] analysed three different types of fusion schemes (sequential, classifier and Bayesian Belief Network (BBN)) combining match scores and liveness measures (degree of liveness). In testing on the LivDet2009 dataset, the BBN fusion scheme gave the best accuracy. However, the results also indicated that the accuracy is affected by the liveness detection performance.

Research on fusion between match scores and liveness factors has only recently started. To our knowledge, the evaluation has only been done with a single spoofing sample. In the proposed scheme, the evaluation is undertaken with sets of spoofing samples of various sizes, taken from a standard dataset.

3. Proposed Fusion Scheme

The idea of the proposed fusion scheme at score-level [16] is to consider spoofing attempts in the fusion module. Therefore, this work extends the framework of Rodrigues et al. [15] to liveness metadata (see Fig. 1). The biometric system's task is to decide, whether given a vector of biometric observation samples $\vec{o} = (o_1, \ldots o_n)$ an identity claim referring to enrolled gallery samples $\vec{g} =$

$(g_1, \ldots, g_n)$ for this individual (e.g. fingers, eyes, etc.) is true (belonging to the class *genuine*) or false (belonging to the class *impostor*). Let i be the current index and $E(o_i), E(g_i)$ refer to extracted features of samples. Further, $s_i = C(E(o_i), E(g_i)) \in [0,1]$ denote comparison scores indicating the degree how much o_i, g_i resemble (assuming lower values represent more likeliness) and $l_i = L(o_i) \in [0,1]$ denote liveness scores (assuming higher values represent spoofs). The task of the fusion module F is to produce a decision score based on the vectors of comparison scores $\vec{s} = (s_1, \ldots, s_n)$ and liveness values $\vec{l} = (l_1, \ldots, l_n)$ used for verification V based on a threshold η:

$$V(\vec{o}, \vec{g}) := \begin{cases} genuine, & if\ F(\vec{s}, \vec{l}) \geq \eta; \\ impostor, & else. \end{cases} \quad (1)$$

The aim is to find a method F which is not affected in performance if m out of the n instances of $\vec{o}$ are spoofed.

3.1. Median Filtering in Score Fusion

For an integration of counter-spoofing techniques into biometric fusion, this work investigates fixed score rules following Kittler et al.'s classical framework [7], where liveness $\vec{l}$ is not considered, see *sum* and *median* rules:

$$F_{sum}(\vec{s}) := \frac{1}{n} \sum_{i=1}^{n} s_i; \quad F_{median}(\vec{s}) := \underset{i=1}{\overset{n}{med}}\, s_i. \quad (2)$$

Being claimed to be outperformed by more effective sum and product rules, the median rule has widely been neglected in score-level fusion in general [16] and so far not been investigated for counter-spoofing [1, 14]. This paper introduces a variation of the median rule, called *median-filter* for higher spoofing-resistance:

$$F_{mf}(\vec{s}) := \frac{1}{\sum_{i=1}^{n} M(\vec{s}, s_i)} \sum_{i=1}^{n} M(\vec{s}, s_i) s_i. \quad (3)$$

$$M(\vec{s}, s_i) := \begin{cases} 1, & if\ \left| s_i - \underset{j=1}{\overset{n}{med}}\, s_j \right| < \phi; \\ 0, & else. \end{cases} \quad (4)$$

Parameter ϕ is either a fixed (trained) or score-dependent threshold (our experiments employ the standard deviation of scores: $\phi = 2\sigma, 3\sigma$). The median filter uses the median of the score-set to remove outliers and local average to find a better representative. This fusion rule aims at: (1) higher resistance to outliers, as generally single spoofed scores represent scores following a different (more genuine-like) distribution, (2) better representation of non-outliers using a filter, (3) easy integration and (4) no need for training.

3.2. Median-filtered Score-and-Liveness Fusion

The addition of further metadata like liveness suggests training to learn the impact of ancillary information. Fig. 2 illustrates observed pairs of liveness and comparison scores for sets of genuine, spoof and (zero-effort) impostor comparisons. Generic 2 class classifiers, as logistic regression (LR) or support vector machines (SVM) can be employed to find the hyperplane $\Psi : \vec{w} \cdot \vec{x} - \vec{b} = 0$ optimally separating these sets:

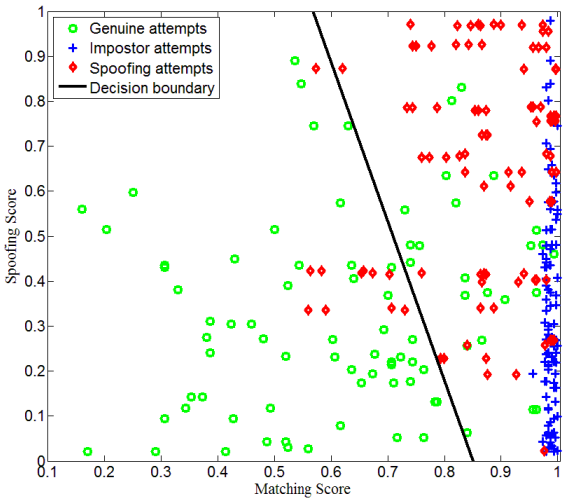

Figure 2. Liveness and comparison scores with trained decision boundary for genuine, impostor and 1-spoof pairs of one finger.

$$D = \{(x_i, y_i) | x_i \in \mathbb{R}^d, y_i \in \{-1, 1\}\}, \quad (5)$$

where $d = 2$ in the scheme combining scores and liveness, thus $x_i = (s_i, l_i)$ is the value-pair and $\{-1, 1\}$ stand for the sets *genuine* and joint *impostor-spoof*. In order to combine multiple evidence in this scenario, again a *median filter* is employed returning a fusion result in $[0,1]^2$:

$$F^2_{mf}(\vec{s}, \vec{l}) := \frac{1}{\sum_{i=1}^{n} M\left(\begin{bmatrix}\vec{s}\\ \vec{l}\end{bmatrix}, \begin{bmatrix}s_i\\ l_i\end{bmatrix}\right)} \sum_{i=1}^{n} M\left(\begin{bmatrix}\vec{s}\\ \vec{l}\end{bmatrix}, \begin{bmatrix}s_i\\ l_i\end{bmatrix}\right) \begin{bmatrix}s_i\\ l_i\end{bmatrix}. \quad (6)$$

$$M\left(\begin{bmatrix}\vec{s}\\ \vec{l}\end{bmatrix}, \begin{bmatrix}s_i\\ l_i\end{bmatrix}\right) := \begin{cases} 1, & if\ \left\| \begin{bmatrix}s_i\\ l_i\end{bmatrix} - \underset{j=1}{\overset{n}{med}} \begin{bmatrix}s_j\\ l_j\end{bmatrix} \right\| < \phi; \\ 0, & else. \end{cases} \quad (7)$$

In this case, the 1-median (the geometric median) in Euclidean space is employed, the point minimizing the sum of distances to the sample points. The fusion result as a score-liveness pair can be classified using LR or SVM returning the (signed) distance to the dividing hyperplane as result:

$$F_{mf}(\vec{s}, \vec{l}) := dist(F^2_{mf}(\vec{s}, \vec{l}), \Psi). \quad (8)$$

This way, threshold variation is equal to moving the hyperplane separating the two (genuine and impostor) joint score- and liveness-distributions. In experiments LR classification is used.

Table 1. EER/SEER (in %) and d-Prime results of four-finger fusion on LivDet2013 varying the number of spoofed fingers.

Method	(S)EER					d-Prime				
	0-spoof	1-spoof	2-spoof	3-spoof	4-spoof	0-spoof	1-spoof	2-spoof	3-spoof	4-spoof
Sum rule	0.14	1.91	3.42	5.83	7.52	2.48	2.40	2.27	2.10	1.94
Median rule	1.56	1.23	2.75	5.05	7.5	2.43	2.41	2.27	2.07	1.87
Median filter	1.24	1.29	2.89	5.60	7.76	2.55	2.52	2.34	2.12	1.93
1-Median filter + LR	1.69	1.78	1.78	1.78	1.78	2.89	2.89	2.89	2.89	2.89

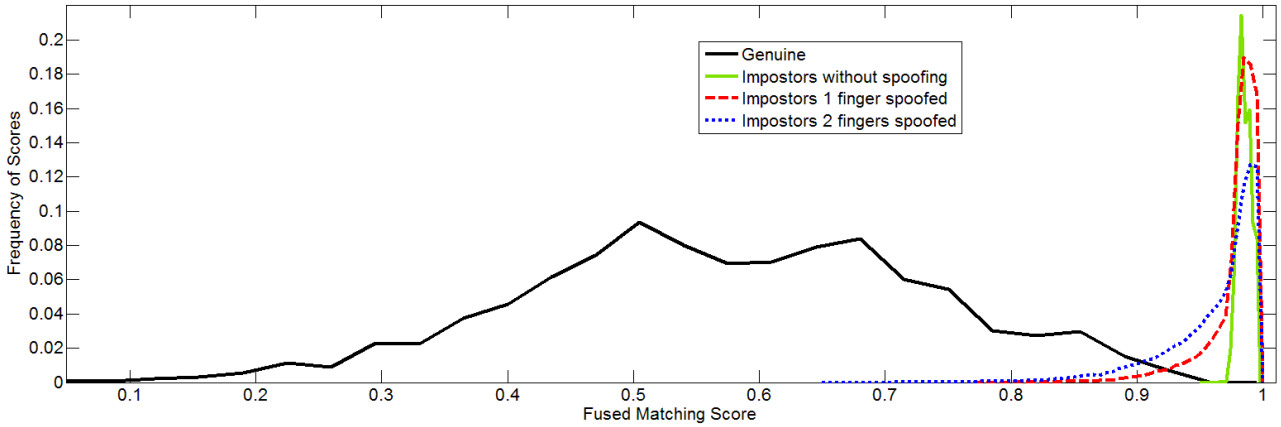

Figure 3. Genuine and impostor score distributions for multi-instance fingerprint fusion using the sum rule.

4. Experiments

In order to test the stability of proposed fusion methods under spoofing attacks, a standard fingerprint recognition system with a custom liveness detector is employed. The tested system combines $n = 4$ multi-instance fingerprints, simulating a simultaneous multi-finger acquisition device. It is evaluated using spoof attack scenarios, where an impostor has access to $m = 0, 1, \ldots, n$ latent fingerprints of different instances to create spoofed forgeries, while having to present all n fingers for authentication (m-*spoof* attack).

4.1. Test Setup

The evaluation of the multibiometric system is carried out using Receiver Operating Characteristics (ROC) by varying system threshold η (see Eq. 1) showing the relationship between Genuine Acceptance Rate (GAR, the percentage of genuine users being accepted) and False Acceptance Rate (FAR, percentage of impostors being accepted) for $m = 0$, and Spoof False Acceptance Rate (SFAR, percentage of spoof-attempts being accepted) as introduced in [6] for $m > 0$. In a similar manner (S)EER is referred to as the (Spoof) Equal Error Rate where GAR=(S)FAR and decidability index (d-Prime) as $d' = |\mu_1 - \mu_2| / \sqrt{(\sigma_1^2 + \sigma_2^2)/2}$ as measure how well distributions with mean μ_i and standard deviation σ_i are separated.

The CrossMatch sensor set of the LivDet2013 fingerprint database is employed for evaluation, using left hand finger images for training and right hand fingers for evaluation. The test dataset containing 2500 live and 2000 spoof images using BodyDouble, Latex, Playdoh and WoodGlue as spoofing materials is grouped into 186 classes providing 4 fingers each originating from the same person. Spoofing attempts are simulated by randomly replacing m out of n fingers with corresponding spoofs.

4.2. Test System

The Matlab-based system[1] employs NIST's open source Biometric Image Software [12] for fingerprint recognition and as spoofing detector a method following [11].

Feature extractor *mindtct* automatically detects minutiae (level-2) features in the epidermis of human skin, tracking position and orientation of ridge bifurcation and termination points extractible at 350 to 500 dots per inch. It employs fingerprint enhancement, generation of local ridge orientation and frequency maps, local minutiae detection identifying pixel patterns, minutiae filtering and quality assessment.

Minutiae points are compared using the *bozorth3* algorithm in 1:1 comparison mode. This translation and rotation-invariant method uses location and orientation to

[1] www.cvg.rdg.ac.uk

estimate a pairing of minutiae in compatibility tables, traverses and links entries into clusters, and accumulates a match score. The more linked entries that exist, the larger is the returned match score. A more detailed description of fingerprint processing can be obtained from [8].

Spoofing detection employs regularized LR for classification. The method is trained using the distinct left-hand subset and achieves an error rate of 27.65% misclassified fingerprints (*ferrlive*) and 24.2% misclassified fake fingerprints (*ferrfake*) on the test-set.

4.3. Results

Table 1 lists results for (1) the reference fusion rules *sum* and - augmenting evaluations [6, 3, 1] - *median rule*, (2) the proposed *median filtering* using scores only, and (3) *1-median filtering* of score and liveness values with *logistic regression* classification. All methods refer to 4-finger fusion on LivDet2013 varying the number of spoofed fingers.

4.3.1 Impact of Spoofing on Fingerprint Fusion

Focusing on the question *"How does a spoofing of m out of n fingers impact on fusion?"* m-spoof score distributions are examined. Figure 3 illustrates, that even a single spoofed finger severely shifts the impostor score distribution (now containing also the impact of deceiving spoof fingers in the fusion result) towards the genuine scores. This is most likely due to sum rule fusion taking every score into account without any rejection of outliers. From Table 1 and Fig. 4 illustrating the ROC of standard sum rule fusion it can be seen, that every additional finger increases EER by an absolute value of ≈1.8-2.4%. Confirming results in [3] that even a spoofing of a single finger impacts on recognition accuracy, it is observed that even 4-finger spoofing does not necessarily imply success for imposter attempt - the overall reported sum rule EER in this case is 7.52% (vs. 0.14% 0-spoof). Fig. 2 illustrates spoof scores following a different distribution than genuine scores for a single finger.

4.3.2 Targeting Spoofing in the Fusion Module

For answering the question *"How to avoid a negative impact of scores originating from fake fingerprints on overall recognition accuracy?"* this paper proposed methods to suppress this information at fusion stage. Even though being ignored so far due to the reported superiority of the sum rule for zero-effort impostors [7] the experimental results indicate that the median rule is more robust in a spoofing environment, yielding a better EER for 1-spoofs (1.23%) than the sum-rule (1.91%). However, for the 0-spoof case, median rule rejects useful information. This can be targeted by median filtering in finding a better representative of non-outliers. In this case also d-Prime measures are clearly superior. ROCs illustrated in Fig. 5 for 0-spoof and 1-spoofs in the range of interest are approximately *colliding* with

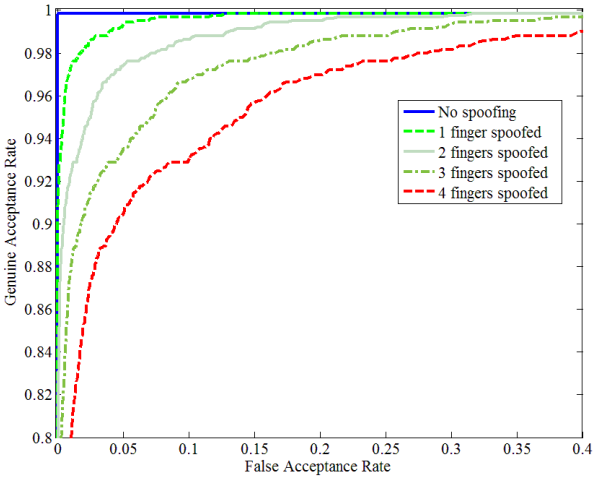

Figure 4. Fingerprint spoofing ROC using Sum rule fusion.

EERs of 1.24% (0-spoof) and 1.29% (1-spoof). If 1 or 2 out of 4 fingers are spoofed, median filtering clearly outperforms the sum rule, while not using any ancillary information. This is, because the median has a breakdown point of 0.5 and is able to suppress a number of outliers (scores from spoofed fingers) up to half of samples. However, for zero-effort impostors, the sum rule is still optimal, at least for non-optimized ϕ. Experiments show, that 1-spoofing can successfully be targeted by employing suitable fusion.

4.3.3 Integrating Scores and Liveness-values

While median-filtering delivers better results than the sum rule for m-spoofing attempts with $m > 0$, as soon as the number of spoofing attempts becomes larger than half of the n instances, the performance degrades drastically. Therefore, the question *"How to integrate spoofing countermea-*

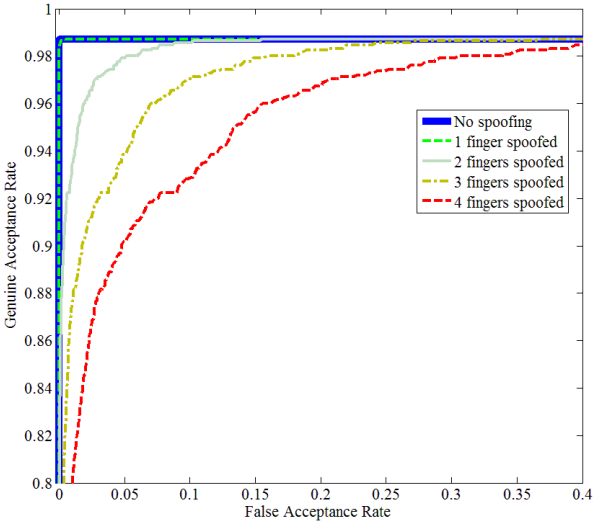

Figure 5. Fingerprint spoofing ROC using Median filtering.

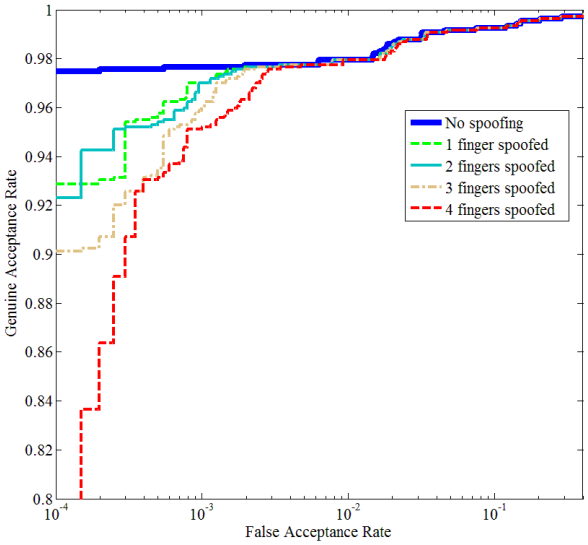

Figure 6. Fingerprint spoofing ROC using 1-Median filtering + LR.

sures in fusion rules?" is discussed in this section assessing the proposed 1-median filtering of liveness-score pairs using the trained logistic regression model. From the ROC curves for this fusion method in Fig. 6, it can be seen that this method is much more robust versus 3-spoof and 4-spoof attacks than sum or even median rule investigated before. For (S)FARs greater than 10^{-3} corresponding GARs differ minimally, with stable EERs in 1.69-1.78% (d-Prime 2.89) over all spoofing attempts. This is remarkable, given the low spoofing detection rate and illustrates the superior security performance of the proposed method compared to the other presented techniques. As the number of combined instances is low, median filtering comes at little computing overhead.

5. Conclusion

Recent studies indicated a high impact of spoofing on fusion recognition accuracy. In this paper we contributed a fusion scheme using outlier detection employing the 1-median with two implemented methods as a means to detect score anomalies for increased security in multibiometric systems. Results illustrated how scores in a multi-finger spoofing scenario degrade, if m out of n fingers are spoofed. While median filtering comes at the cost of slightly reduced 0-spoof performance, it is much more robust than sum rule fusion if a minority of features is spoofed. Experiments on LivDet2013 data showed that the proposed integrated scores and spoofing-countermeasures fusion using trained LR classification with median filtering is able to almost eliminate the impact of spoofing, retaining stable 1.69-1.78% EER over all m-spoof tests. As future research we envision the incorporation of quality into the scheme, optimized selection of the median filter parameter ϕ, and an evaluation on other multimodal databases.

Acknowledgements

This work was supported by the EU FASTPASS project under grant agreement 312583.

References

[1] Z. Akhtar, G. Fumera, G. Marcialis, and F. Roli. Evaluation of multimodal biometric score fusion rules under spoof attacks. In *Proc. Int'l Conf. on Biometrics (ICB)*, pages 402–407, 2012.

[2] Z. Akhtar, G. Fumera, G. Marcialis, and F. Roli. Evaluation of serial and parallel multibiometric systems under spoofing attacks. In *Prof. Int'l Conf. Biometrics: Theory, Applications and Systems (BTAS)*, pages 283–288, Sept 2012.

[3] Z. Akhtar, S. Kale, and N. Alfarid. Spoof attacks on multimodal biometric systems. In *Proc. Int'l Conf. on Information and Network Technology*, volume 4, pages 46–51, 2011.

[4] P. Coli, G. Marcialis, and F. Roli. Vitality detection from fingerprint images: A critical survey. In S.-W. Lee and S. Li, editors, *Advances in Biometrics*, volume 4642 of *LNCS*, pages 722–731, Berlin Heidelberg, 2007. Springer.

[5] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G. Marcialis, F. Roli, and S. Schuckcrs. Livdet 2013 fingerprint liveness detection competition 2013. In *Proc. Int'l Conf. on Biometrics (ICB)*, pages 1–6, 2013.

[6] P. Johnson, B. Tan, and S. Schuckers. Multimodal fusion vulnerability to non-zero effort (spoof) imposters. In *Proc. Int'l Wksp. on Inf. For. and Sec. (WIFS)*, pages 1–5, 2010.

[7] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[8] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2005.

[9] E. Marasco, Y. Ding, and A. Ross. Combining match scores with liveness values in a fingerprint verification system. In *Proc. Int'l Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pages 418–425, 2012.

[10] E. Marasco, P. Johnson, C. Sansone, and S. Schuckers. Increase the security of multibiometric systems by incorporating a spoofing detection algorithm in the fusion mechanism. In C. Sansone et al., editors, *Multiple Classifier Systems*, volume 6713 of *LNCS*, pages 309–318. Springer, Berlin, 2011.

[11] S. B. Nikam and S. Agarwal. Gabor filter-based fingerprint anti-spoofing. In J. Blanc-Talon et al., editors, *ACIVS*, volume 5259 of *LNCS*, pages 1103–1114. Springer, 2008.

[12] NIST. Biometric Image Software V4.2. 2013.

[13] A. Rattani, N. Poh, and A. Ross. A bayesian approach for modeling sensor influence on quality, liveness and match score values in fingerprint verification. In *Proc. Int't Wksp. on Inf. Forensics and Sec. (WIFS)*, pages 37–42, 2013.

[14] R. Rodrigues, N. Kamat, and V. Govindaraju. Evaluation of biometric spoofing in a multimodal system. In *Proc. Int'l Conf. Biometrics: Th. App. & Syst. (BTAS)*, pages 1–5, 2010.

[15] R. N. Rodrigues, L. L. Ling, and V. Govindaraju. Robustness of multimodal biometric fusion methods against spoof attacks. *J. Vis. Lang. Comput.*, 20(3):169–179, 2009.

[16] A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer, 2006.