Multimodal Anti-Spoofing in Biometric Recognition Systems

Giorgio Fumera, Gian Luca Marcialis, Battista Biggio, Fabio Roli and Stephanie Caswell Schuckers

Abstract While multimodal biometric systems were commonly believed to be intrinsically more robust to spoof attacks than unimodal systems, recent results provided clear evidence that they can be evaded by spoofing a single biometric trait. This pointed out that also multimodal systems require specific anti-spoofing measures. In this chapter we introduce the issue of multimodal anti-spoofing, and give an overview of state-of-the-art anti-spoofing measures. Such measures mainly consist of developing ad hoc score fusion rules, that are based on assumptions about the match score distribution produced by fake biometric traits. We discuss the pros and cons of existing measures, and point out the current challenges in multimodal anti-spoofing.

1 Introduction

Other chapters of this book addressed the issue of "anti-spoofing" in unimodal biometric systems. In this kind of system, anti-spoofing consists of an additional module, also called "liveness detector", that is used as a countermeasure to spoof attacks. This countermeasure is able to detect if the biometric signal acquired by some sensors belongs to a "live" person or is an artificial replica: for example, a fake finger, or a 2d photo of a face.

Anti-spoofing in multimodal biometric systems, or, for the sake of brevity, "multimodal anti-spoofing", is not a clear concept as in the unimodal case. On the basis of

Giorgio Fumera, Gian Luca Marcialis, Battista Biggio, Fabio Roli
Department of Electrical and Electronic Engineering, University of Cagliari,
Piazza d'Armi - 09123 Cagliari (Italy),
e-mail: {fumera,marcialis,battista.biggio,roli}@diee.unica.it

Stephanie Caswell Schuckers
Department of Electrical and Computer Engineering,
Clarkson University, PO Box 5720 – Potsdam, NY 13699, e-mail: sschucke@clarkson.edu

what has been discussed in other chapters, we can define multimodal anti-spoofing as follows:

1. Fusion of multiple liveness detectors for a single biometric. This can be motivated by the fact that different approaches to liveness detection can be "complementary", thus they can be combined at different levels to improve the liveness detection performance (for example, fingerprint liveness detectors which combine fake-based and live-based features according to [22]). This definition comes from the fact that liveness detection is substantially a two-class classification problem where an individual liveness detector can be seen as a classifier. Approaches aimed at fusing multiple liveness detectors appeared in [15, 14, 22, 21]. Therefore, in this case, "multimodal anti-spoofing" refers to the combination of "modalities" defined by each individual liveness detector.
2. Fusion of liveness detector and matcher. In this case, the different "modalities" are given by the two heterogeneous classifiers, namely, the uni-biometric verification system and the liveness detector. This very recent topic has been considered in [1, 19, 20].
3. Multi-biometric systems where no liveness detection is performed [27]. In this case, each "modality" consists of a different kind of biometric trait (*e.g.*, fingerprint and face), which is processed by a different biometric verification module. The proposed anti-spoofing measures for this kind of system consist of ad hoc score fusion rules, capable to exploit the information coming from a match score generated when comparing a spoofed biometric against the related genuine template(s) [25, 24, 8, 18].

Definitions 1 and 2 refer to unimodal biometric systems. Therefore, for sake of clarity and with respect to the scope of this book, we will focus on definition 3 only, where "multimodal anti-spoofing" means protecting multimodal biometric systems against spoof attacks by using robust score-level fusion rules. This topic is very recent, since the first papers were published in 2009.

The rest of this chapter is organized as follows. Section 2 states the problem. Section 3 presents the current state of the art on multimodal spoofing, with an introduction aimed at pointing out how this problem was perceived in the biometric community before 2009. Section 4 reports experimental evidences on the effectiveness on anti-spoofing measures described in Section 3. Finally, Section 5 is devoted to a discussion on current challenges in multimodal anti-spoofing, and to a summary of recent achievements in the "Tabula Rasa" research project.

2 Problem statement

In this chapter we will refer to multimodal biometric *verification* systems, whose aim is to verify a claimed identity on the basis of different biometric traits submitted by the user. In multimodal biometric verification, a person submits N different biometrics, and claims a certain identity. A feature set x_i is extracted from the i-th

biometric, and compared by a matching algorithm with the template of the claimed identity. The output is a match score s_i for the i-th biometric system. Match scores $\{s_1,...,s_N\}$ coming from such matchers are processed by a "fusion rule" module, which produces a "combined" match score that is finally used to accept of reject the claimed identity. This can be done according to two basic approaches:

1. The distribution of genuine and impostor match scores of individual matchers are used for estimating the joint likelihoods, and the Likelihood Ratio (LR) rule is used [25].
2. The distribution of genuine and impostor match scores are derived from the fused scores, and the final decision is made by setting an appropriate acceptance threshold.

Let $\{s_1,...,s_N\}$ be the set of match scores coming from the N matchers, and let $I \in \{0,1\}$ be the Boolean random variable denoting whether the verification attempt comes from an impostor ($I = 1$) or a genuine user ($I = 0$). The approach 1 above requires that the joint probability likelihood ratio is evaluated:

$$LR(s_1,...,s_N) = \frac{P(s_1,...,s_N|I=0)}{P(s_1,...,s_N|I=1)} = \frac{P(s_1|I=0)\cdot...\cdot P(s_N|I=0)}{P(s_1|I=1)\cdot...\cdot P(s_N|I=1)}, \quad (1)$$

where $P(s_i|I=0)$ and $P(s_i|I=1)$ are the match score distributions of genuine users and of impostors, respectively, for the i-th matcher. Joint distributions $P(s_1,...,s_N|I=0)$ and $P(s_1,...,s_N|I=1)$ can be derived by factorizing individual likelihoods under the hypothesis that match scores are independent given I. This hypothesis is commonly accepted in multimodal biometric verification, since it is reasonable that match scores derived from the comparison of different kinds of biometrics are independent on each other. A threshold λ is then set, and, if $LR(s_1,...,s_N) > \lambda$, the user is accepted as genuine; otherwise, it is rejected as an impostor.

In the approach 2 above, a fusion rule $f(s_1,...,s_N)$ is applied to the set of match scores, to derive a novel match score:

$$s = f(s_1,...,s_N;\Theta), \quad (2)$$

where Θ is a set of parameters which can be estimated from an additional validation set. If $\Theta = \emptyset$ no parameters are necessary and the rule is referred to as a "fixed" one, otherwise it is referred to as a "trained" rule. For example, well-known fixed rules are the maximum score, or the sum or product of the N scores. Examples of trained rules are the weighted sum and the weighted product of the N scores.[1] The final score s is compared to an acceptance threshold s^*, as usually done for uni-biometric systems. If $s > s^*$, the user is accepted as genuine; otherwise, it is rejected as an impostor. In this chapter we will consider both approaches above.

When a spoof attack is performed, one or more of the N required biometrics are counterfeited with an artificial replica. In this case, we could expect that:

[1] Also the LR rule and similar ones can be referred to as trained rules, as they require the estimation of likelihoods.

1. $M < N$ biometrics have been counterfeited, so that they should be similar to the corresponding biometrics of the claimed identity.
2. The remaining $N - M$ biometrics are non-counterfeited, so that they belong to one or more impostors.

Thus, the problem is to reject users that attempt a spoof attack, as well as users that do not attempt a spoof attack, but claim a different identity, by classifying them as impostors ($I = 1$), and to classify the other users as genuine ($I = 0$). This implies that detecting the liveness of submitted biometric is not the focus of such systems.

In absence of spoof attacks, biometric systems are evaluated on the basis of two kinds of errors:

- the fraction of impostors accepted as genuine users, called False Acceptance Rate (FAR);
- the fraction of genuine users rejected as impostors, called False Rejection Rate (FRR); Genuine Acceptance Rate (GAR = 1−FRR) can be used as well.

By varying the acceptance threshold one obtains different pairs of FAR and GAR values. Plotting FAR (GAR) as a function of FAR leads to the well-known DET (ROC) curve.

In presence of spoof attacks, the evaluation of a third kind of error has been proposed in [18], and is also adopted in this chapter, namely, the rate of accepted spoof attacks, called Spoof-False Acceptance Rate (SFAR). A spoof false accept is defined as the case where the fusion rule falsely accepts an attacker when one or more of the modalities in a multimodal system have been successfully spoofed. We will call SROC (Spoof-ROC) the curve obtained by plotting GAR as a function of SFAR, to distinguish it from the standard ROC curve.

It is important to note that, at a specific operating point (*i.e.*, for a given acceptance threshold), the FRR does not change under a spoof attack. Therefore, reporting the ROC and SROC curves related to the same system allows one to easily evaluate the performance when the system is under standard operation conditions or under spoof attacks, respectively. More precisely, for a given operating point whose FRR value is denoted as FRR^*, the effect of a spoof attack is represented by the SFAR that is obtained by the intersection of the SROC curve with the horizontal line given by FRR=FRR^*.

3 State of the art on multimodal biometric spoofing

In this section, we review recent works that pointed out the vulnerability of multimodal systems to spoof attacks against a single biometric trait. We then describe the anti-spoofing measures proposed so far against such a vulnerability.

3.1 Are multimodal systems more secure than unimodal ones?

About a decade ago, the use of multimodal systems to implement anti-spoofing measures was proposed [16, 17, 26]:

> "[...] multibiometric systems provide anti-spoofing measures by making it difficult for an intruder to simultaneously spoof the multiple biometric traits of a legitimate user. By asking the user to present a random subset of biometric traits, the system ensures a live user is indeed present at the point of data acquisition." [16]

The rationale is that, even if an intruder is able to fabricate a perfect replica of a given trait (*e.g.*, a fingerprint), he/she is not guaranteed to be wrongly recognized as a genuine user, since the system may not require the submission of that trait. This would force intruders to spoof more than one trait, making multimodal systems a deterrent agains such attacks.

Subsequently, a stronger belief spread in the biometric community, about the security of multimodal systems against spoof attacks, *i.e.*, evading them always requires an intruder to spoof *all* the involved traits (or at least more than one), *regardless* of the adoption of any anti-spoofing measure like the one above, and, most important, of the considered score fusion rule. This would imply that multimodal systems are *intrinsically* more secure than unimodal ones against spoof attacks, in the sense that their evasion requires a higher effort by the intruder. Even though this is far from evident, until a few years ago no work questioned or attempted to verify such a belief. This issue has been addressed first in [25, 18, 24]. These works studied the performance of multimodal biometric systems that use parallel score fusion rules, where only a subset of the modalities used in the system, or even a *single* one, are spoofed, and the attacker merely presents his/her own biometric traits for the remaining biometric(s). They even argued that this would make such systems *more vulnerable* than unimodal ones:

> "Intuitively, a multimodal system is intrinsically more secure than unimodal systems since it is more difficult to spoof two or more biometric traits than a single one. However, is it really necessary to spoof all the fused biometric traits to crack a multimodal system? This question is specially important when a very secure biometric (*e.g.*, retina scan) is combined with another that is easily spoofed (*e.g.*, face). In this scenario, the benefits of adding the face information may be negated by reduction in overall security." [25]

> "If an intruder can break the multimodal system by attacking only one mode, then the multimodal system is not more secure than the least secure mode (*i.e.*, the "weakest link"). In this case, we can even argue that the multimodal system is less secure than its unimodal systems alone since the use of several modes leads to a bigger number of vulnerability points, increasing the possibility of an intruder to take advantage of at least one of these vulnerabilities. For example, consider a multimodal system combining signature and fingerprint traits under a spoof attack. In this scenario, a forger can choose which trait to spoof according to his skills, what may increase his chances of being successful." [24]

In fact, empirical evidences provided in [25, 18, 24], and subsequently in [20] showed that parallel multimodal systems that combine two to four different modalities, and use several state-of-the-art score fusion rules, can be evaded by spoofing a single trait. Moreover, the probability of accepting spoof impostors as genuine users

(*i.e.*, the SFAR) was shown to be significantly affected by the choice of the score fusion rule. Despite this interesting result, designing score fusion rules that are robust to spoof attacks without degrading significantly the performance in the absence of spoofing is still an open issue, as well as thoroughly assessing the security of standard score fusion rules to understand whether and under what circumstances some fusion rules may be intrinsically more secure than others. Accordingly, another open issue remains that of deriving clear guidelines to help system designers to select the most appropriate score fusion rule for the task at hand (*e.g.*, depending on the level of *security* and performance in the absence of spoofing required by their system).

It is worth pointing out also that the aforementioned result was obtained by *simulating* spoof attacks, under the assumption that, in a successful spoof attack, the match score distribution of fake traits equals that of live, genuine traits. This corresponds to a scenario in which the attacker is able to produce a replica which is (statistically) indistinguishable by the matcher from the original, live trait. Accordingly, to simulate a successful spoofing attempt, a randomly drawn genuine score was used for the spoofed modality, while the impostor scores were used for the other modalities (*i.e.*, assuming the attacker supplied his/her own biometric).

In [25] a biometric system combining face and fingerprint modalities was considered, and the LR and Weighted sum were used as score fusion rules. For instance, when the operational point was set to FAR=0.1%, under a simulated face spoofing attack a SFAR of about 40% was measured. In other words, the probability that an impostor is wrongly accepted as a genuine user by submitting a replica of the face of the targeted client is 400 times higher than for standard impostors. An even more disruptive effect was observed for fingerprint spoofing, for which the SFAR became almost 100%. This means that the considered multimodal system can be evaded almost with certainty by providing a "good" fingerprint replica.

Similar results were reported in [24] for face spoofing. Contrary to [25], fingerprint spoofing was investigated using a data sets of scores coming from real spoof attacks, obtained from the Fingerprint Liveness Detection Competition 2009 [23]. In this case, setting the operational point to FAR=0.1%, the SFAR was about 8%, *i.e.*, about an order of magnitude lower than in the case of simulated attacks, but still much higher than the desired value of 0.1%.

In [18] a multimodal database consisting of face, iris, and fingerprint match scores from genuine and imposter pairs was used. This database was created by West Virginia University and is available on the CITeR website [2]. Besides two modality systems, the investigation was extended to a three modality system, where one or two modalities are spoofed. The sum fusion rule was considered. A spoof attempt was simulated using a genuine match score in place of an impostor match score, as in [25]. Even using three modalities, spoofing only one trait was found to significantly increase the probability that an attacker is accepted as a genuine user. For instance, the point at which FRR equals FAR (which is known as the equal error rate, EER) was chosen as the operating point of the system by setting the corresponding threshold level, and the corresponding SFAR was analyzed based on the same threshold. Reported results show an EER (FRR/FAR) of 0.05%. For this operating point, when one of three modalities is spoofed, the average SFAR is 4.9%,

with an associated FRR of 0.05%. When two of three modalities are spoofed the SFAR jumps up to an average of 97.4%, that is, over 97% of the time, a person will be able to spoof the system by spoofing two of three modalities.

Finally, in [20] four different modalities were considered, one face and three fingerprints. The LR and sum fusion rules were used. Spoof attacks were simulated again as in [25]. As an example, setting the operational point to EER=0.32%, when the sum rule was used, it was observed that spoofing a *single* fingerprint lead to a SFAR of 9 to 56% (depending on which of the three fingers is spoofed). Using the LR rule, the ERR was much lower, about 0.004%. However, the SFAR was 57 to 91%. This means that spoofing a single trait can allow an attacker to evade even a multimodal system that combines four different modalities. Moreover, we point out that the LR fusion rule was found to be more vulnerable than the simplest sum rule, although it is known to be optimal in the Neyman-Pearson sense (provided that a reliable estimate of the genuine and impostor score distributions is available). Indeed, while its EER is much lower than the one of the sum rule, its SFAR was significantly higher.

In our subsequent work we extended this investigation to *real* spoof attacks, focusing on biometric systems involving two modalities, face and fingerprint [9, 4, 6, 3, 8]. To this aim, we collected in our Lab several face and fingerprint data sets, fabricated fake traits using several techniques, and evaluated several multimodal systems using different sensors, matchers, and score fusion rules (see Sect. 4). All our results clearly confirmed that spoofing a single trait can drastically increase the probability that an impostor is wrongly accepted as a genuine user by a multimodal system. For instance, in [8] we observed that, using the sum, LR and LDA score fusion rules, and setting the operational point to FAR=0.1%, the SFAR that can be attained by face and fingerprint spoofing was respectively beyond 20% and 50% (depending on the score fusion rule and on the spoofing technique used). In [6] we found that also multimodal systems using *serial* score fusion rules exhibit the same vulnerability.

To sum up, empirical evidence collected so far clearly shows that sole use of multiple modalities does not guarantee that a biometric system can not be evaded by spoofing a single biometric trait. In the next section we overview anti-spoofing measures proposed so far for multimodal systems that use parallel score fusion rules, and always require the user to submit all the considered traits.

3.2 Countermeasures to spoof attacks in multimodal systems

In [25, 24, 18] different anti-spoofing measures have been proposed, aimed at reducing the probability (SFAR) that an intruder, by spoofing a single trait, can break multimodal systems that require users to submit *all* the considered traits, and use parallel score fusion rules. All the proposed countermeasures act on the score fusion rule. The simplest one consists of modifying the criterion for choosing the threshold on the fused score [18]; another one involves some modification of existing fusion

rules [25, 24]; another one consists of a novel, ad hoc fusion rule [25]. They are summarized in the following, and their pros and cons are discussed.

While the decision threshold on the fused score is usually set according to a desired trade-off between FRR and FAR, the anti-spoofing measure proposed in [18] consists of setting it according to a trade-off between FRR and SFAR. For instance, if the operational point is set at the ERR, where FRR=FAR, it is suggested in this paper that in order to improve security of the fusion algorithm, a new threshold is considered where SFAR equals FRR. This is called the Spoof EER (EERspoof). Other operating points could also be selected, depending on the application requirements for FRR, FAR, and SFAR. For instance, in the experiments summarized in

Operating point	FAR	FRR	SFAR (one modality)	SFAR (two modalities)
EER	0.05%	0.05%	4.9%	97.4%
EERspoof	$\ll 0.001\%$	2.89%	0.31%	2.89%

Table 1 Results obtained in [18] that show the benefits of selecting the EER operating point according to a trade-off between FRR and SFAR (EERspoof). The first row reports FAR, FRR, and SFAR when one or two modalities are spoofed attained when the EER operating point (FAR=FRR) is chosen. The second row reports the same performance measures when the EERspoof operating point is considered (FAR=SFAR when two modalities are spoofed).

Sect. 3.1, setting the threshold at the EER operating point led to a FAR (and FRR) equal to 0.05%, while the average SFAR was 4.9% for a single spoofed trait, and 97.4% for two spoofed traits. As also reported in Table 1, setting the operating point at EERspoof, a better tradeoff between FRR and SFAR can be achieved at 2.89% EERspoof for the case where two modalities are spoofed. If the operating point of 2.89% EERspoof is chosen, the corresponding FRR is 2.89% and FAR turned out to be $\ll 0.001\%$. In other words, the new operating point decreases SFAR from 97.4% to 2.89% when two modalities are spoofed. Similarly, the average SFAR attained under spoofing of one modality decreases from 4.9% to 0.31%. However, this comes at a sacrifice to FRR which increases from 0.05% to 2.89%. The tradeoff in these adjusted error rates may be preferred given the threat of a spoof attack. In summary, adjusting the operating point according to a system assessment based on SFAR can ensure for a more secure system. The main advantage of this anti-spoofing measure is that it can be applied to any score fusion rule, and is very simple to implement. One drawback is that a higher security comes at a cost of increased FRR. Another drawback is that the match score distribution produced by spoof attacks, and consequently the corresponding SFAR as a function of the decision threshold, are very difficult to estimate. This issue was addressed in [18] by assuming that such a distribution equals the one of the corresponding genuine traits, which is the approach proposed in [25] (see Sect. 3.1, and below). Note that, under this assumption, SFAR=GAR for any value of the decision threshold. This can be a pessimistic assumption. Moreover, it does not allow the proposed anti-spoofing measure to be applied, when the original criterion is the so-called zeroFAR, *i.e.*, setting the decision threshold to the lowest value at which the estimated FAR equals 0. In this

case, under the above assumption the threshold should be set so that SFAR=0. This however implies GAR=0, which is clearly not acceptable.

More complex anti-spoofing measures involve the modification of existing score fusion rules. In [25] the well-known LR rule was considered (see Eq. 1). [25] proposes to estimate $P(s_i|I=1)$ by taking int account also auxiliary information about the "security degree" of each matcher against spoof attacks, and about the probability that an attack against any subset of matchers occur. Denoting with $T_i \in \{0,1\}$ and $F_i \in \{0,1\}$ the Boolean random variables that indicate respectively whether a spoof attack has been attempted against the i-th matcher, and whether it was "successful", $P(s_i|I)$ can be estimated by marginalizing the distribution $P(s_i,T_i,F_i|I)$:[2]

$$P(s_i|I) = \sum_{T_1,\ldots,T_N,F_1,\ldots,F_N} P(T_1,\ldots,T_N|I) \times \prod_{i=1}^{M} P(F_i|T_i)P(s_i|F_i,I). \quad (3)$$

In this model, the security of the i-th matcher is defined as the probability that a spoof attack against it is successful, $P(F_i=1|T_i=1)$. As pointed out in [25], this value should be manually set according to general knowledge. Obviously, $P(F_i=1|T_i=0)=0$. The other two distributions in the right-hand side of Eq. (3) were modeled in [25] according to the following assumptions: (i) spoof attacks against any of the 2^N-1 subsets of one or more matchers are equiprobable, *i.e.*, $P(T_1,\ldots,T_N|I=1)=1-\alpha$, if $T_1=\ldots=T_N=0$, and $P(T_1,\ldots,T_N|I=1)=\frac{\alpha}{2^N-1}$ otherwise, where α is the probability that some spoof attack has been attempted; (ii) genuine users will never provide a fake trait, *i.e.*, $P(T_1,\ldots,T_N|I=0)=0$ for $T_1=\ldots=T_N=1$, and thus $P(s_i|F_i=1,I=0)$ need not to be modeled; (iii) the match score distributions in absence of successful spoof attacks, $P(s_i|F_i=0,I=0)$ and $P(s_i|F_i=0,I=1)$, equal respectively to the standard genuine and impostor distributions, and can thus be modeled from training data; (iv) the match score distribution of a successful attack equals the corresponding genuine distribution: $P(s_i|F_i=1,I=1)=P(s_i|F_i=0,I=0)$. It follows that $P(s_i|I)$ is a mixture of the score distributions of genuine users and impostors distributions.

The above anti-spoofing strategy allows one a finer tuning of the score fusion rule, with respect to the one of [18]. On the other hand, it is tailored to the LR rule only, and is based on the same pessimistic assumption about the score distribution of successful spoof attacks (which was originally proposed in that work). Moreover, it trades a higher flexibility for the necessity of defining the probability α of a spoof attack, and the security of each matcher, $P(F_i=1|T_i=1)$, which are difficult to estimate in practice. Finally, also this countermeasure has the drawback of increasing the FRR, since the mass of the distribution $P(s_i|I=1)$ is shifted towards the genuine distribution.

A simplification of the ExtLR rule was proposed in [24], to avoid an ad hoc choice of the parameters α and $P(F_i=1|T_i=1)$. The former was set to $\alpha=0.5$, based on the rationale that no a priori information about the occurrence of spoof

[2] In [25] quality measures of the biometric samples in each modality were also considered. For the sake of simplicity we do not include them in our description.

attack is usually available. The latter was set to 1 according to the worst-case assumption that each spoof attack will be successful.

A similar approach was also proposed in [10], although not specifically tailored to biometric applications. In particular, the underlying idea of this approach was to learn *secure* classifiers in adversarial settings, including spam filtering, intrusion detection, and biometrics, by modeling the distribution of carefully crafted attacks which may be not present in the training data.

Finally, a novel, ad hoc score fusion rule was proposed in [25]. The rationale is to explicitly defining, by high-level linguistic expressions, the decision criteria to fuse the information consisting of the match scores, the quality measures of the acquired traits (if available), and the prior information about the security of each matcher. To this aim, a fuzzy score fusion rule was devised. The input data and the output score were associated respectively to the linguistic expressions "high score/quality/security", and "high output" (using the convention that the higher the output value, the higher the probability that the user is genuine), which were modeled as fuzzy variables. In particular, the output was associated to the three linguistic values "low", "medium" and "high". The proposed fuzzy rules were defined for a biometric system involving two modalities, according to the criterion that "low security biometric system cannot faithfully perform the recognition task alone; and similarity scores with low quality should have low weights in the final output" [25]. For instance, two of these rules can be phrased as: "if the two match scores are 'high', then the output is 'high' (independently on the quality measures and security levels)"; "if one of the matchers has a 'low' security and produces a 'high' score, while the other produces a 'low' match score (independently on its security level), then the output is 'high' ".

The main advantage of this anti-spoofing strategy lies in the possibility to explicitly defining high-level rules to fuse the input information. On the other hand, the drawback is that the number of fuzzy rules grows exponentially with the number of matchers, which makes it difficult to define them for biometric systems involving three or more modalities. Moreover, empirical evidence provided in [25] shows that also this anti-spoofing measure is likely to increase the FRR.

To sum up, specific anti-spoofing measures proposed so far for multimodal systems are based on more or less complex manipulations of the score fusion rule, and sometimes require the knowledge of information difficult to estimate. Moreover, an inherent feature of these measures is that they trade a higher security against spoof attacks for a higher FRR. In the next section we give an overview of the performance of some of these anti-spoofing measures, on publicly available data sets including real spoof attacks.

4 Experimental evidences and discussion

In this section we present some experimental results from our previous works, with the aim of pointing out the vulnerability of multimodal systems to spoof attacks

against a single biometric trait, and to show how such a vulnerability can be mitigated by one of the anti-spoofing measures described in Sect. 3.2. We chose the Extended LR to this aim, since it is tailored to improve the robustness of the well-known LR score fusion rule. We describe the data sets used in our experiments in Sect. 4.1, and the experimental protocol in Sect. 4.2. The results are presented and discussed in Sect. 4.3.

4.1 Data sets of spoofed samples

Fingerprint spoofing. We used the LivDet 2011 data set, created for the Second Fingerprint Liveness detection competition [28]. It includes 80 clients (distinct fingers). Different impressions of live and fake fingers were acquired in two different sessions, separated by about two weeks. All the ten fingers were considered. Fake fingerprints were created by the consensual method [13]. Gelatine, silicone, alginate, and latex, were used as the casting materials, whilst plasticine- and silicon-like materials were used for molds. Fingerprint images were acquired using the well-known Biometrika FX2000 and Italdata ET10 optical sensors, which have respectively a resolution of 569 dpi and 500 dpi, and a sensing area of 13.2×25 mm and approximately 30.5×30.5 mm. Images of latex fake fingerprints turn out to be very similar to the images of the corresponding live fingerprints, whilst fakes obtained using other materials exhibit some artifacts. The fake fingerprints LivDet 2011 data set represents the state-of-the-art in fingerprint spoofing, and thus provides a reasonable set of realistic scenarios.

Face spoofing. We used three publicly available data sets: the *Photo Attack* and Personal Photo Attack [9], and the *Print Attack* data set [7, 12]. In the Photo Attack and Personal Photo Attack data sets, two different kinds of face spoof attacks were considered. The live face images of each client were collected in two sessions, with a time interval of about two weeks between them, under different lighting conditions and facial expressions. Spoofed face images for the Photo Attack data set were constructed using the following "photo attack" method, also used in [12, 29]. It consists of displaying a photo of the targeted client on a laptop screen (or printing it on paper), and of showing it to the camera. To this end, the testing "live" face images of the clients were used. This simulates a pessimistic scenario in which the attacker can obtain photos of the targeted client under a setting similar to the one of the verification phase. The Personal Photo Attack data set has been built using personal photos voluntarily provided by 25 of the 50 clients (on average, 3 photos for each client), that were acquired in different times and under different environmental conditions than those of the live training and testing images. This simulates a more realistic scenario where the attacker is able to collect a photo of the targeted client, *e.g.*, from the Web. According to the above observations, we expect that the fake score distribution of our Photo Attack data set, provided by a given matching algorithm, will be very similar to that of the genuine users. This may not be true for the Personal Photo Attack, instead, whose effectiveness strongly

Data set	Number of clients	Number of spoofed images per client	Number of live images per client
LivDet11-Alginate	80	3	5
LivDet11-Gelatin	80	3	5
LivDet11-Silicone	80	3	5
LivDet11-Latex	80	3	5
Photo Attack	40	60	60
Personal Photo Attack	25	3 (avg.)	60
Print Attack	50	12	16

Table 2 Characteristics of the fake fingerprint and fake face data sets used in our experiments.

depends on the ability of the attacker to obtain images similar to the templates used by the system.

The Print Attack data set was constructed during the Competition on Countermeasures to 2D Facial spoof attacks, in 2011. It consists of 200 video clips of printed-photo attack attempts to 50 clients, under different lighting conditions, and of 200 genuine access attempts from the same clients. We extracted the "live" and spoofed face images from the corresponding videos. In particular, for each client, we extracted 12 "live" face images and 16 spoofed face images from each video clip.

Table 2 reports the size of all the data sets described above.

4.2 Experimental protocol

We used a similar experimental protocol as in [25, 18]:

- We built $5 \times 3 = 15$ *chimerical* data sets, by randomly associating face and fingerprint images of pairs of clients of the available five fingerprint and three face data sets. Building chimerical data sets is a widely used approach in experimental investigations on multimodal biometrics [27].
- Each chimerical data set was randomly subdivided into five pairs of training and testing sets. 40% of the "virtual" clients was included into the training set,[3] and the remaining 60% for the testing set. All the above procedure was repeated five times, for different random associations of face and fingerprint images of pairs of clients (namely, creating different "virtual" clients). In each run, the parameters of the trained fusion rules have been estimated on the training set. The results reported below refer to the average testing set performance, over the resulting twenty-five runs.
- The fake match scores were computed by comparing each fake image of a given client with the corresponding template image.

[3] The clients of a chimerical data set are usually referred to as "virtual" clients, since they do not correspond to a real person or identity. They are indeed created by randomly associating the biometric traits of different "real" clients.

- We normalized all match scores in $[0, 1]$ using the min-max technique [27]. The normalization parameters were estimated on the training set.
- The performance was assessed by computing the DET curves (FRR vs FAR). Note that, in the evaluation of spoof attacks, the DET curve reports FRR vs SFAR [18]. In both cases, the performance increases as the curve gets closer to the origin.

We used the NIST Bozorth3[4] and the VeryFinger[5] matching algorithms, for fingerprint verification. They are both based on matching the fingerprint minute details, called "minutiae". As they exhibited very similar behaviors, we will only report the results for Bozorth3. The Elastic Bunch Graph Matching (EBGM) algorithm was used for face verification.[6] It is based on representing a face with a graph, whose nodes are the so-called face "landmarks" (centered on the nose, eyes, and other points detected on the face). These nodes are labeled by a feature vector, and are connected by edges representing geometrical relationships among them. We also carried out some preliminary experiments using the Principal Component Analysis (PCA) and the Linear Discriminant Analysis (LDA), which yield again very similar results to that of the EBGM algorithm, and are thus omitted for sake of clarity.

We investigated three attack scenarios using the fake traits of our data sets: (a) only fingerprints are spoofed, (b) only faces are spoofed, (c) both fingerprints and faces are spoofed (bi-modal or double spoofing). For the scenarios (a) and (b), we also evaluated simulated spoof attacks under the pessimistic scenario defined in [25, 24, 18]: fictitious fake scores were generated by randomly drawing a set of genuine match scores from the testing set.

The fusion rules we considered are: sum, product, weighted sum (LDA), LR, and Extended LR. Since the bi-modal system considered in these experiments is the same as in [25], we used for the Extended LR rule the same values of the parameters as in [25] (*i.e.*, the probability that a spoofing attack against either matcher is successful, and the prior probability of a spoof attack: see Sect. 3.2).

4.3 Results and discussion

We report in the following a representative set of results, related to the following pairs of spoof attacks:

1. LivDet11-Latex and Photo Attack;
2. LivDet11-Gelatin and Print Attack;
3. LivDet11-Alginate and Personal Photo Attack.

We do not report here any result related to fake fingerprints fabricated with silicone (LivDet11-Silicone), since they attained very similar results to that fabricated with

[4] `http://www.nist.gov/itl/iad/ig/nbis.cfm`

[5] `http://www.neurotechnology.com/verifinger.html`

[6] `http://www.cs.colostate.edu/evalfacerec/algorithms5.php`

Rule	*no spoof*	*face*	*s-face*	*fing.*	*s-fing.*	*both*
	EER %	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	9.98 ± 2.1	33.25 ± 3.9	37.82 ± 3.8	44.07 ± 4.8	79.85 ± 3.8	60.89 ± 2.9
Product	3.49 ± 1.4	5.72 ± 2.1	6.43 ± 2.2	70.06 ± 5.4	96.11 ± 1.8	73.10 ± 4.9
LDA	3.32 ± 1.5	8.39 ± 4.3	9.87 ± 4.8	70.79 ± 5.6	96.36 ± 2.2	74.09 ± 5.4
LR	3.60 ± 1.4	5.58 ± 2.8	6.36 ± 3.2	71.41 ± 5.1	96.46 ± 2.2	73.47 ± 5.1
Ext. LR	3.61 ± 1.4	5.64 ± 2.7	6.40 ± 3.1	71.49 ± 5.0	96.38 ± 2.2	73.57 ± 5.1
	FRR % at FAR=1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	17.41 ± 3.2	15.58 ± 1.6	20.46 ± 1.9	28.38 ± 5.2	69.00 ± 6.1	46.00 ± 3.8
Product	5.15 ± 2.7	1.93 ± 0.4	2.28 ± 0.5	63.22 ± 4.7	94.37 ± 3.1	66.57 ± 4.4
LDA	5.05 ± 2.6	2.17 ± 0.5	2.73 ± 0.7	64.91 ± 4.7	95.12 ± 3.1	67.83 ± 4.6
LR	5.46 ± 2.6	1.22 ± 0.4	1.43 ± 0.5	64.94 ± 4.7	95.22 ± 3.1	66.38 ± 4.7
Ext. LR	5.63 ± 2.8	1.17 ± 0.4	1.38 ± 0.5	64.68 ± 4.8	94.94 ± 3.3	66.03 ± 4.8
	FRR % at FAR=0.1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	22.76 ± 3.5	8.84 ± 1.1	12.68 ± 1.5	21.65 ± 4.9	62.68 ± 6.7	37.30 ± 3.9
Product	8.59 ± 4.1	0.30 ± 0.1	0.36 ± 0.1	53.41 ± 5.5	90.62 ± 4.1	56.79 ± 4.9
LDA	7.99 ± 3.7	0.26 ± 0.1	0.32 ± 0.2	56.32 ± 5.3	92.45 ± 3.8	58.66 ± 5.2
LR	8.91 ± 4.1	0.14 ± 0.1	0.17 ± 0.1	56.23 ± 6.0	92.39 ± 3.9	57.48 ± 6.0
Ext. LR	9.46 ± 5.2	0.16 ± 0.1	0.19 ± 0.1	56.13 ± 5.7	90.97 ± 5.5	57.27 ± 6.0

Table 3 EER, FRR at FAR=1%, and FRR at FAR=0.1% for the considered fusion rules on the Livdet11-Latex and Photo Attack data sets (*no spoof*). The SFAR corresponding to the same operating points is reported for real spoof attacks against fingerprint (*fing.*), face (*face*), and both traits (*both*), and for the simulated spoof attacks against fingerprints (*s-fing.*), and face (*s-face*). Results are averaged over 25 runs, and are reported as mean and standard deviation.

latex (LivDet11-Latex), as also reported in [8]. Tables 3–5 report the average performance and the standard deviation attained on each data set by all fusion rules, for three operating points: EER (when FAR=FRR), FAR=1%, and FAR=0.1%. This allows us to directly compare the performance of the different fusion rules under standard operating conditions (in terms of FAR and FRR), and their robustness to spoof attacks (in terms of SFAR). Each operating point was fixed on the DET curve obtained without spoof attacks, namely, by considering genuine users and impostors (non-spoof attacks). The FRR at each selected operating point is reported in the first column (labelled as *no spoof*) of Tables 3-5. The SFAR attained by the different spoof attacks at the same operating point is reported in the remaining columns. This allows us to understand to which extent a fusion rule is robust: once the operating point is fixed, the effect of spoofing is only to increase the FAR (actually, the SFAR) as it only affects impostor match scores, while the FRR remains unchanged.

When no spoofing attack is performed, all fusion rules exhibited almost the same performance, except for the Sum rule (see the *no spoof* column in Tables 3-5). The Sum rule performed much worse, because of the strong performance imbalance between the fingerprint and the face matcher: the genuine users and impostors score distributions of the face matcher exhibited indeed a much higher overlapping.

Spoofing a single biometric trait always led to a SFAR higher than the corresponding FAR in the considered multimodal systems, for all the adopted fusion rules (see the *fing.*, *s-fing.*, *face* and *s-face* columns in Tables 3-5). This provides evidence

Rule	*no spoof*	*face*	*s-face*	*fing.*	*s-fing.*	*both*
	EER %	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	14.35 ± 2.2	46.31 ± 3.0	47.26 ± 2.6	29.98 ± 3.0	76.38 ± 5.1	58.97 ± 3.9
Product	5.25 ± 1.6	16.31 ± 4.4	21.20 ± 5.5	53.54 ± 8.6	92.94 ± 3.0	68.01 ± 7.9
LDA	4.32 ± 1.8	29.54 ± 9.9	38.27 ± 8.5	53.28 ± 10.3	94.02 ± 3.4	69.51 ± 10.1
LR	4.16 ± 1.6	17.68 ± 8.7	28.65 ± 12.3	56.31 ± 9.5	94.88 ± 2.7	66.64 ± 10.8
Ext. LR	4.18 ± 1.6	16.52 ± 7.9	27.68 ± 12.3	56.02 ± 9.6	94.84 ± 2.8	66.16 ± 10.9
	FRR % at FAR=1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	28.04 ± 5.1	32.84 ± 2.3	39.50 ± 1.0	8.42 ± 2.8	56.24 ± 8.9	40.90 ± 3.5
Product	8.87 ± 3.2	4.87 ± 0.9	7.50 ± 1.5	38.40 ± 7.7	88.61 ± 4.9	52.89 ± 6.7
LDA	6.43 ± 2.9	10.42 ± 4.1	23.48 ± 7.8	41.88 ± 7.7	91.61 ± 4.7	56.57 ± 6.5
LR	6.58 ± 3.0	3.34 ± 1.4	8.18 ± 5.9	45.46 ± 7.9	92.98 ± 4.1	52.52 ± 8.2
Ext. LR	6.64 ± 3.0	3.15 ± 1.3	7.03 ± 4.8	45.39 ± 7.8	92.95 ± 4.2	52.22 ± 8.0
	FRR % at FAR=0.1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	34.52 ± 5.7	22.08 ± 4.5	36.08 ± 1.4	3.86 ± 1.9	46.04 ± 9.8	31.93 ± 4.6
Product	13.82 ± 4.2	1.05 ± 0.3	1.90 ± 0.4	26.08 ± 7.2	82.45 ± 6.5	38.21 ± 7.2
LDA	9.98 ± 3.7	1.39 ± 0.6	4.61 ± 2.9	31.64 ± 7.8	88.62 ± 5.6	41.27 ± 7.0
LR	10.73 ± 4.3	0.37 ± 0.2	0.81 ± 0.5	33.78 ± 8.2	89.53 ± 5.3	38.04 ± 8.4
Ext. LR	10.61 ± 4.4	0.40 ± 0.2	0.81 ± 0.4	34.15 ± 7.9	89.67 ± 5.4	38.59 ± 7.9

Table 4 Results attained on the Livdet11-Gelatin and Print Attack data set. See the caption of Table 3 for the details.

that multimodal systems are not intrinsically robust to spoof attacks against a single biometric trait. In particular, fingerprint spoofing was almost always much more harmful than face spoofing. Also this behavior is due to the performance imbalance between the fingerprint and the face matcher, which caused all fusion rules, except for Sum, to give a higher weight to the former. Therefore, the Sum rule exhibited the worse performance in the absence of attacks, and the highest vulnerability to face spoof attacks, while it turned out to be the least vulnerable to fingerprint spoofing. Spoofing both traits led to an even higher SFAR (see the *both* column in Tables 3-5). No that the SFAR values reported in Tables 3–5 under the *fing.* columns show that latex-based fingerprint spoofing appears more effective than using gelatin-based fake fingerprints; in turn, the latter appears more effective than using alginate-based fake fingerprints.

Let us now focus on the Extended LR rule, which was specifically designed to counteract spoof attacks. At the considered operating points, it performed similarly to the LR rule, independently on the kind of attack, although it was expected to improve the robustness of the LR rule. Moreover, it exhibited a SFAR higher than that of the LR rule, at operating points characterized by FAR values lower than the ones considered in Tables 3-5 (these results are not reported here, due to lack of space). This suggests that the assumption about the score distribution of fake traits on which the Extended LR rule is based is too pessimistic. This can also be argued by the fact that the SFAR of the simulated spoof attacks always overestimates the corresponding SFAR under a real spoof attack.

Rule	*no spoof*	*face*	*s-face*	*fing.*	*s-fing.*	*both*
	EER %	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	10.57 ± 1.5	10.75 ± 3.5	37.97 ± 4.1	14.80 ± 1.6	78.32 ± 3.2	16.63 ± 4.0
Product	4.08 ± 1.1	6.12 ± 1.8	8.05 ± 2.0	25.09 ± 6.0	95.62 ± 1.6	30.36 ± 9.0
LDA	3.89 ± 1.3	6.48 ± 2.4	11.63 ± 3.7	25.16 ± 5.6	95.81 ± 1.8	28.75 ± 7.5
LR	4.14 ± 1.1	5.03 ± 1.9	7.89 ± 2.4	25.43 ± 6.0	95.97 ± 1.7	28.52 ± 6.9
Ext. LR	4.14 ± 1.1	5.17 ± 1.7	8.31 ± 2.8	25.88 ± 5.5	96.03 ± 1.7	28.78 ± 7.1
	FRR % at FAR=1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	18.80 ± 3.0	1.40 ± 1.0	20.37 ± 2.3	2.83 ± 0.8	66.84 ± 4.8	2.94 ± 1.6
Product	6.38 ± 2.2	1.47 ± 0.6	2.33 ± 0.3	13.61 ± 4.4	93.18 ± 2.8	15.36 ± 6.1
LDA	6.21 ± 2.4	1.47 ± 0.7	2.84 ± 0.8	14.53 ± 4.4	94.09 ± 2.8	15.48 ± 6.2
LR	6.64 ± 2.5	1.15 ± 0.5	1.71 ± 0.4	15.01 ± 4.7	94.41 ± 2.8	14.81 ± 5.9
Ext. LR	6.63 ± 2.5	1.13 ± 0.5	1.64 ± 0.3	14.86 ± 4.6	94.29 ± 2.9	14.64 ± 5.8
	FRR % at FAR=0.1%	SFAR %	SFAR %	SFAR %	SFAR %	SFAR %
Sum	24.59 ± 3.3	0.17 ± 0.2	12.53 ± 1.4	0.82 ± 0.5	60.14 ± 5.0	0.66 ± 0.7
Product	9.81 ± 3.1	0.17 ± 0.1	0.38 ± 0.1	6.27 ± 3.1	89.14 ± 3.6	6.05 ± 3.9
LDA	9.21 ± 3.1	0.17 ± 0.1	0.36 ± 0.2	7.23 ± 3.2	91.26 ± 3.5	6.40 ± 3.9
LR	10.55 ± 4.6	0.15 ± 0.1	0.15 ± 0.1	6.82 ± 3.8	90.86 ± 3.9	5.52 ± 3.8
Ext. LR	10.85 ± 6.5	0.15 ± 0.1	0.18 ± 0.1	7.33 ± 3.8	88.69 ± 10.8	6.24 ± 3.7

Table 5 Results attained on the Livdet11-Alginate and Personal Photo Attack data set. See the caption of Table 3 for the details.

With regard to this behavior, Tables 3–5 show that the difference between the SFAR of simulated face spoofing and of real face spoofing (*face* and *s-face* columns) is much lower than in the case of fingerprint spoofing (*fing.* and *s-fing.* columns), for all the considered fusion rules. In particular, for face spoofing this difference is often very small. This means that the assumption underlying the simulation of spoof attacks made in [25] is too pessimistic, especially for fingerprint spoofing. Note that results of Tables 3 and 4 refer to face spoof attacks obtained by using face images taken from the testing set, that are thus similar to the template of the targeted users. Thus, the corresponding fake score distribution is likely to be similar to that of the genuine users, but this is a pessimistic attack scenario.

The above results suggest that the assumption that the fake score distribution equals the one of genuine users can be often violated by real spoof attacks, and that a more realistic modeling of the fake score distribution should be adopted for designing robust score fusion rules.

5 Current challenges in multimodal anti-spoofing

In this chapter, we reviewed the main achievements in the field of multimodal anti-spoofing. We summarized empirical evidence showing that multimodal biometric systems are not intrinsically robust to spoof attacks against one biometric trait, and that the probability of accepting spoof impostors as genuine users especially de-

pends on the chosen score fusion rule. In particular, multimodal systems that use well-known *trained* score fusion rules, like the LR, turned out to be potentially more vulnerable than systems that use simpler, *fixed*, rules (*e.g.*, the product rule), even though the LR is optimal according to the Neyman-Pearson criterion, provided that the genuine and impostor distributions are reliably estimated. Despite the reported results, no specific guidelines on how to determine the most suitable fusion rule for the task at hand have been given yet. It is still an open issue to understand whether and to what extent score-level fusion rules are vulnerable to spoof attacks, and under what circumstances some rules may be intrinsically more secure than others. Nevertheless, the reported empirical evidences suggest us that ad hoc anti-spoofing measures should be adopted also in multimodal systems (like the one originally proposed in [16]).

We then considered recently proposed anti-spoofing measures that consist of modifying existing score fusion rules, and of developing novel ones, by exploiting additional information about spoof attacks, in terms of specific assumptions on the corresponding match score distribution. We focused in particular on the approach proposed in [25], and subsequently used in [24, 18, 20], for evaluating the vulnerability of multimodal systems against spoof attacks, and for designing robust score fusion rules. It is based on the assumption that the match score distribution produced by successful spoof attacks equals that of genuine users. While this approach does not require one to fabricate spoof attacks in order to estimate the corresponding match score distribution, the above assumption is often violated in practice, and turns out to be too pessimistic. This can lead one to overestimate the vulnerability of multimodal systems. Moreover, when the above assumption is used in the design of robust score fusion rules like the Extended LR [25], it can even make the resulting multimodal system *less* robust to real spoof attacks.

Accordingly, a suitable assumption for the match score distribution produced by spoof attacks is crucial in the design of robust score fusion rules, if no information from liveness detection modules is exploited. In principle, this requires one to fabricate a large variety of spoof attacks to analyze the corresponding match score distribution. However, this would affect the scalability of a multimodal biometric system, since the fake score distribution should be re-estimated as novel genuine users are added. It would also affect the acceptability of the system, since genuine users would have to provide replicas of their own biometrics.

A possible solution is to develop realistic models of the fake score distribution that are *representative* of different kinds of spoof attacks, and can be used by the designers of multimodal systems without the need of actually fabricating any spoof attack. A preliminary attempt towards this direction was made by the authors in [5, 3]. The model proposed in that work consists of simulating a family of fake score distributions that exhibit an *intermediate* behavior between the impostor and the genuine distribution, parametrized by a measure of the relative distance to the latter. However, such a model was based on a limited empirical evidence, and, thus, it may not properly account for the wide variability of spoof attack distributions induced, *e.g.*, by different spoof fabrication techniques and materials. Extending this preliminary model and collecting larger empirical evidence on spoof attack distri-

butions to overcome such limitations is indeed part of the authors' ongoing work. Further, we advocate that security of multimodal biometric systems to spoof attacks should be evaluated and improved based on a *proactive* what-if analysis, *i.e.*, by anticipating potential (and novel) spoof attacks that may be incurred during system operation, as also suggested in our recent work on the security of pattern recognition systems [11]. To this end, relying on a well-suited *simulation* model of the spoof distribution has the main advantage of allowing system designers to thoroughly assess the security of multimodal systems against a large number of spoof attack scenarios, instead of considering a limited number of cases corresponding to spoof attacks fabricated in a laboratory setting. This approach may also shed some light on the open issue mentioned at the beginning of this section, *i.e.*, it may help understanding the security properties of standard score-level fusion rules in a more systematic manner.

Another approach for improving the robustness of multimodal systems is to integrate the match score with the score provided by liveness detectors, through suitable fusion rules. Although building a liveness detector requires one to collect samples of spoof attacks, it is not required that such samples are taken from the genuine users of the multimodal biometric system under design. This approach is under investigation in the "Tabula Rasa" research project,[7] and preliminary results are reported in [20]. Note that this approach does not fall within the definition of "multimodal anti-spoofing" considered in Sect. 1, since state-of-the-art liveness detection is related to the integration of liveness and match scores in *unimodal* biometric systems [19]. We also argue that the aforementioned spoof simulation model may be also exploited to the same end, *i.e.*, to train fusion rules that combine matching algorithms and liveness detectors without submitting any fake trait to the matching algorithm to estimate the corresponding score distribution.

To sum up, the issue of multi-biometric anti-spoofing has been raised only recently in the biometric community, and it has quickly become one of the most relevant open problems in this field. Further and more systematic theoretical and experimental investigations of this issue are therefore needed, taking into account the large variety of biometrics and of possible score-level fusion rules.

Acknowledgements This work has been partly supported by the TABULA RASA project, 7th Framework Research Programme of the European Union (EU), grant agreement number: 257289; by the project CRP-18293 funded by Regione Autonoma della Sardegna (RAS), L.R. 7/2007, Bando 2009; and by a grant awarded to B. Biggio by RAS, PO Sardegna FSE 2007-2013, L.R. 7/2007 "Promotion of the scientific research and technological innovation in Sardinia".

References

1. Abhyankar, A., Schuckers, S.A.C.: Integrating a wavelet based perspiration liveness check with fingerprint recognition. Pattern Recognition **42**(3), 452–464 (2009)
2. Adler, A., Schuckers, S.A.C.: Security and liveness, overview. In: S.Z. Li, A.K. Jain (eds.) Encyclopedia of Biometrics, pp. 1146–1152. Springer US (2009)

[7] `http://www.tabularasa-euproject.org/`

3. Akhtar, Z., , Fumera, G., Marcialis, G., Roli, F.: Evaluation of multimodal biometric score fusion rules under spoof attacks. In: Proc. 5th IEEE/IAPR Int'l Conf. Biometrics (ICB), pp. 402–407. New Delhi, India (2011)
4. Akhtar, Z., Biggio, B., Fumera, G., Marcialis, G.: Robustness of multi-modal biometric systems under realistic spoof attacks against all traits. In: Proc. IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS 2011), pp. 5–10. Milan, Italy (2011)
5. Akhtar, Z., Fumera, G., Marcialis, G., Roli, F.: Robustness evaluation of biometric systems under spoof attacks. In: G. Maino, G. Foresti (eds.) Int'l Conf. Image Analysis and Processing (ICIAP), *LNCS*, vol. 6978, pp. 159–168. Springer Berlin Heidelberg (2011)
6. Akhtar, Z., Fumera, G., Marcialis, G., Roli, F.: Evaluation of serial and parallel multibiometric systems under spoofing attacks. In: Proc. IEEE 5th Int'l Conf. Biometrics: Theory, Applications and Systems (BTAS), pp. 283–288. Washington DC, USA (2012)
7. Anjos, A., Marcel, S.: Counter-measures to photo attacks in face recognition: a public database and a baseline. In: Int'l Joint Conf. on Biometrics (IJCB) (2011)
8. Biggio, B., Akhtar, Z., Fumera, G., Marcialis, G., Roli, F.: Security evaluation of biometric authentication systems under real spoofing attacks. IET Biometrics **1**, 11–24 (2012)
9. Biggio, B., Akhtar, Z., Fumera, G., Marcialis, G.L., Roli, F.: Robustness of multi-modal biometric verification systems under realistic spoofing attacks. In: A.K. Jain, A. Ross, S. Prabhakar, J. Kim (eds.) Int'l Joint Conf. on Biometrics (IJCB), pp. 1–6. IEEE (2011)
10. Biggio, B., Fumera, G., Roli, F.: Design of robust classifiers for adversarial environments. In: IEEE Int'l Conf. on Systems, Man, and Cybernetics (SMC), pp. 977–982 (2011)
11. Biggio, B., Fumera, G., Roli, F.: Security evaluation of pattern classifiers under attack. IEEE Transactions on Knowledge and Data Engineering **99**(PrePrints), 1 (2013)
12. Chakka, M.M., Anjos, A., Marcel, S., Tronci, R., Muntoni, D., Fadda, G., Pili, M., Sirena, N., Murgia, G., Ristori, M., Roli, F., Yan, J., Yi, D., Lei, Z., Zhang, Z., Z.Li, S., Schwartz, W.R., Rocha, A., Pedrini, H., Lorenzo-Navarro, J., Castrillón-Santana, M., Maatta, J., Hadid, A., Pietikainen, M.: Competition on counter measures to 2-D facial spoofing attacks. In: Proc. IAPR IEEE Int'l Joint Conf. on Biometrics (IJCB). Washington DC, USA (2011)
13. Coli, P., Marcialis, G.L., Roli, F.: Vitality detection from fingerprint images: A critical survey. In: S.W. Lee, S.Z. Li (eds.) Proc. Int'l Conf. Biometrics (ICB), *LNCS*, vol. 4642, pp. 722–731. Springer, Seoul, Korea (2007)
14. Ghiani, L., Marcialis, G.L., Roli, F.: Experimental results on the feature-level fusion of multiple fingerprint liveness detection algorithms. In: Proc. 14th ACM Workshop on Multimedia and Security (MMSec), pp. 157–164. ACM, Coventry, UK (2012)
15. Ghiani, L., Marcialis, G.L., Roli, F.: Fingerprint liveness detection by local phase quantization. In: Int'l Conf. Pattern Rec. (ICPR), pp. 537–540. IEEE (2012)
16. Jain, A.K., Ross, A.: Multibiometric systems. Commun. ACM **47**(1), 34–40 (2004)
17. Jain, A.K., Ross, A., Pankanti, S., Member, S.: Biometrics: A tool for information security. IEEE Transactions on Information Forensics and Security **1**, 125–143 (2006)
18. Johnson, P., Tan, B., Schuckers, S.: Multimodal fusion vulnerability to non-zero effort (spoof) imposters. In: IEEE Int'l Workshop on Information Forensics and Security (WIFS), pp. 1–5 (2010)
19. Marasco, E., Ding, Y., Ross, A.: Combining match scores with liveness values in a fingerprint verification system. In: Proc. IEEE 5th Int'l Conf. Biometrics: Theory, Applications and Systems (BTAS). Washington DC (USA) (2012)
20. Marasco, E., Johnson, P.A., Sansone, C., Schuckers, S.A.C.: Increase the security of multibiometric systems by incorporating a spoofing detection algorithm in the fusion mechanism. In: C. Sansone, J. Kittler, F. Roli (eds.) Prof. 10th Int'l Workshop Multiple Classifier Systems (MCS), *LNCS*, vol. 6713, pp. 309–318. Springer, Naples, Italy (2011)
21. Marasco, E., Sansone, C.: Combining perspiration- and morphology-based static features for fingerprint liveness detection. Pattern Recognition Letters **33**(9), 1148–1156 (2012)
22. Marcialis, G.L., Coli, P., Roli, F.: Fingerprint liveness detection based on fake finger characteristics. Int'l J. Digital Crime and Forensics (IJDCF) **4**(3), 1–19 (2012)

23. Marcialis, G.L., Lewicke, A., Tan, B., Coli, P., Grimberg, D., Congiu, A., Tidu, A., Roli, F., Schuckers, S.A.C.: 1st Int'l Fingerprint Liveness Detection Competition - LivDet 2009. In: P. Foggia, C. Sansone, M. Vento (eds.) Proc. 15th Int'l Conf. Image Analysis and Processing (ICIAP), *LNCS*, vol. 5716, pp. 12–23. Springer, Vietri sul Mare, Italy (2009)
24. Rodrigues, R.N., Kamat, N., Govindaraju, V.: Evaluation of biometric spoofing in a multimodal system. In: Int'l Conf. Biometrics: Theory Applications and Systems (BTAS), pp. 1–5 (2010)
25. Rodrigues, R.N., Ling, L.L., Govindaraju, V.: Robustness of multimodal biometric fusion methods against spoof attacks. J. Vis. Lang. Comput. **20**(3), 169–179 (2009)
26. Ross, A.: An introduction to multibiometrics. In: Proc. 15th European Signal Processing Conference (EUSIPCO 2007), pp. 20–24. Poznan, Poland (2007)
27. Ross, A.A., Nandakumar, K., Jain, A.K.: Handbook of Multibiometrics. Springer Publishers (2006)
28. Yambay, D., Ghiani, L., Denti, P., Marcialis, G.L., Roli, F., Schuckers, S.: LivDet2011 - Fingerprint Liveness Detection Competition 2011. In: 5th Int'l Conf. Biometrics (ICB) (2012)
29. Zhang, Z., Yi, D., Lei, Z., Li, S.Z.: Face liveness detection by learning multispectral reflectance distributions. In: Int'l Conf. Automatic Face and Gesture Recognition, pp. 436–441 (2011)