Robustness Analysis of Likelihood Ratio Score Fusion Rule for Multimodal Biometric Systems under Spoof Attacks

Zahid Akhtar, Giorgio Fumera, Gian Luca Marcialis and Fabio Roli
Dept. of Electrical and Electronic Engineering
University of Cagliari
Piazza d'Armi, 09123 Cagliari, Italy
Email: {z.momin,fumera,marcialis,roli}@diee.unica.it

***Abstract*—Recent works have shown that, contrary to a common belief, multi-modal biometric systems may be "forced" by an impostor by submitting a spoofed biometric replica of a genuine user to only one of the matchers. Although those results were obtained under a worst-case scenario when the attacker is able to replicate the exact appearance of the true biometric, this raises the issue of investigating more thoroughly the robustness of multi-modal systems against spoof attacks and devising new methods to design robust systems against them. To this aim, in this paper we propose a robustness evaluation method which takes into account also scenarios more realistic than the worst-case one. Our method is based on an analytical model of the score distribution of fake traits, which is assumed to lie between the one of genuine and impostor scores, and is parametrised by a measure of the relative distance to the distribution of impostor scores, we name "fake strength". Varying the value of such parameter allows one to simulate the different factors which can affect the distribution of fake scores, like the ability of the attacker to replicate a certain biometric. Preliminary experimental results on real bi-modal biometric data sets made up of faces and fingerprints show that the widely used LLR rule can be highly vulnerable to spoof attacks against one only matcher, even when the attack has a low fake strength.**

I. Introduction

With the rapid growth in the use of biometric systems, issues about their robustness and security against external attacks are also raising. Several researchers are investigating the vulnerabilities of biometric systems, the potential attacks and the related countermeasures. Among the others, the attack which is of greatest interest in the biometric community consists in submitting to the system a counterfeit, or fake, biometric [1], which is known as "spoof attack", "direct attack", since the true biometric is replaced by a fake one. Several authors showed that biometrics such as fingerprints, iris and faces, can be stealthy procured and used to generate synthetic biometric traits to attack biometric sensors. Although several potential countermeasures have been proposed so far, no effective one exists yet.

Besides ad hoc countermeasures, it is commonly believed that multi-modal systems are intrinsically more robust against spoof attacks, since their evasion would require to spoof *all* biometric traits simultaneously [2]. However such belief is not based on theoretical or empirical evidences, but only on intuitive and qualitative arguments, which rely mainly on the higher performance of multi-modal systems with respect to mono-modal ones.

Actually, such belief has been questioned very recently in [3]–[5], where it has been shown that multi-modal systems can be cracked by faking *only one* of the biometric traits. Those results have been obtained under the stringent, worst-case scenario when the attacker is capable to produce an exact replica of the targeted client's biometric. Anyway, they raise the need of further investigations on the robustness of multi-modal systems under spoof attacks, and of developing effective countermeasures.

In this work, we address this issue by proposing a method to evaluate the robustness of a multi-modal system against spoof attacks. Our goal is to avoid the straightforward but cumbersome solution of constructing spoofed biometric traits to test the system. Since no multi-modal data set containing spoof attacks has been made available so far, our method is based on simulating the effects of a spoof attack on the distribution of the corresponding matching scores, as in [3]–[5]. However, differently from these works, our aim is to take into account also more realistic, non-worst-case scenarios, in which the fake score distribution can be different than the genuine one. The distribution of fake scores may be affected by different factors, like the particular spoofed biometric, the sensor, the matching algorithm, the technique used to construct fake biometrics, the skills of the attacker, etc. However, at the state of the art their effect is unknown. We thus propose to model such distribution by assuming that by effect of the above factors they can exhibit different shapes, and in particular, that in can be identical either to the impostor or to the genuine score distributions, or lies between them. To model distributions lying between the ones of genuine and impostor scores, we introduce a single parameter that controls their relative similarity to the genuine distribution (or equivalently, the relative distance from the impostor one), which we name "fake strength": the higher the similarity, the higher the "strength" of the spoof attack. For instance, this can reflect the different ability of attackers to replicate the targeted genuine biometric, being equal all the other factors mentioned

above.

Our method can be applied to any multi-modal system, and can be used by a designer to obtain an estimation of the system performance under potential spoof attacks of different strength. It can also be used to compare the robustness of different score fusion rules applied to a given multi-modal system.

We finally present a case study related to a bi-modal system made up of a fingerprint and a face matcher, using the likelihood ratio score fusion rule (LLR). Our results show that the LLR may be highly vulnerable to spoof attacks, even of low strength. This suggests as possible, relevant follow-ups of this work, the construction of proper data sets containing spoof attacks, to verify the assumptions behind our model, and to compare the robustness of the different fusion rules proposed so far in the literature.

The paper is organized as follows. In Sect. II we summarise previous works on spoof attacks. Our robustness evaluation method is presented in Sect. III. Experimental results are reported in Sect. IV. Preliminary conclusions are drawn in Sect. V.

II. Background

A systematic analysis of the potential vulnerabilities of biometric systems has been given in [1], where eight possible different kinds of attacks have been identified (see also Fig. 5 of [1]). Spoof attacks, which are the focus of this paper, have been considered by several authors. For instance, in [6] artificial fingers were created using gum and gelatine, and a 60% impostor acceptance rate was reported, when they were submitted to a fingerprint sensor. Liveness detection at sensor level is a possible countermasure suggested by most of the researchers [7]. Hardware- and software-based algorithms for the so-called fingerprint liveness detection have been already developed. The first survey on this topic is [8]. Several other approaches have been subsequently proposed [9], [10].

The spoof attack does not involve only fingerprints. Recently, some papers appeared regarding iris and faces [11], [12]. In particular, it is well-known that iris may be faked by appropriate contact lens, and it has been shown that a simple photo in front of a camera may allow to deceive a face verification system.

Although several fake detection algorithms have been proposed to avoid spoof attacks, the state of the art is not yet mature. The common effect of fingerprint, or face, liveness detection systems is the increase of the false reject rate, because several "live" fingerprints are misclassified as fake ones. Recent results appeared on [13] pointed out that the average error rate of state-of-the-art fingerprint liveness detection algorithms is about 15%, using the MAP criterion. Therefore, the overall acceptability and performance of biometric verification systems decreases when they are coupled with liveness detection algorithms.

As mentioned in the introduction, multi-modal systems are commonly believed to be a possible countermeasure against spoof attacks. They have been originally proposed to improve the performance of mono-modal systems, and their effectivenss has been shown by extensive theoretical and empirical evidences [14], [15]. Moreover, they also proved to be quite robust under stress conditions, namely, deliberate attempts of the user to deceive the system by poor co-operation or some kind of forgery (e.g., wearing glasses or beard) [16]. Their claimed robustness against spoof attacks is based only on a intuitive argument instead, namely that their evasion would require to spoof *all* biometric traits simultaneously [2]. Such claim is however not supported so far by theoretical or empirical evidences, and has been questioned in [3]–[5]. In [3], the authors empirically simulated the effect of a spoof attack on the matching scores of a bi-modal system made up of a face and a fingerprint matcher, in the worst-case scenario in which the attacker is capable to exactly replicate the fingerprint of the targeted client, so that the same matching score as the client's template is obtained. In other words, the distribution of the scores of fake fingerprints is identical to the one of genuine users. Reported experiments on some real data sets, with the widely used weighted sum and likelihood ratio score fusion rules, showed that under the above scenario the false acceptance rate (FAR) of a multi-modal system can dramatically increase, allowing an attacker to crack the system by spoofing *only one* biometric trait. Such results are very discouraging, especially because LLR is known to be the optimal decision rule when the score distributions are exactly known, and lead the authors of [3] to propose two new score fusion rules to improve robustness against spoof attacks. Such rules (a fuzzy rule and a modification of the LLR rule) are based on a quality measure aimed at discriminating among fake and live traits, and (in the case of the modified LLR rule) to take into account the possibility of spoof attack at design phase, by simulating the presence of scores of spoofed traits among training data.

The results of [3] have been obtained under a worst-case scenario, which may be not realistic in many practical settings. In fact, faking some kinds of biometric traits, like fingerprints, is not trivial [6]. Moreover, although simple spoofing techniques may be effective for some biometrics (for instance, showing a photo of the targeted client in front of the camera, in the case of a face matcher), they can be easily detected in systems with human intervention.

Nevertheless, the results of [3] raise the issue of investigating more systematically and more thoroughly the robustness of multi-modal systems to spoof attacks, by focusing also on realistic scenarios where fake traits are not exact replicas of the original ones.

III. A method for robustness analysis of multi-modal biometric systems under spoof attacks

A straightforward way to evaluate the robustness of a biometric system against spoof attacks is to construct fake biometric traits and present them to the system. However this can be a difficult and impractical task [13]. An alternative solution is to *simulate* the effects of a spoof attack on the

matching score of the corresponding biometric trait, as in [3]–[5]. In this section we follow this approach, and propose a method based on a model of the fake scores distribution (Sect. III-A) to estimate the performance of a multi-modal system under spoof attacks (Sect. III-B). In Sect. III-C we show how this method can be applied to a bi-modal system using the LLR fusion rule, which is the case study of the esperiments of Sect. IV.

A. A model of the fake scores distribution

In real scenarios it is reasonable to assume that the score distribution of fake traits is different than the genuine one, and that spoof attacks carried out using different techniques, or attempted by different attackers with different forgery skills, lead to different distributions of fake scores.

To develop a model of the fake scores distribution it would be very useful to start from empirical data. However, although several data sets of live and fake biometric traits currently exist, like the LivDet09 data set (http://prag.diee.unica.it/LivDet09), they do not provide useful information on the score distribution of fake traits, since they have not been indexed by the users' identity. To our knowledge, the only exception is the data set of [10], which however is rather small and not publicly available.

The solution we propose is to make a working assumption on the possible forms that the fake scores distribution may exhibit, due to the possible effects of the different factors mentioned above. In the following we denote with s the score of a biometric matcher, and with G and I the events that the input biometric is true and comes respectively from a genuine user and an impostor, while the event that it is a fake biometric will be denoted as F. The corresponding score distributions will thus be denoted as $p(s|\mathrm{G}), p(s|\mathrm{I})$ and $p(s|\mathrm{F})$. Our working assumptions on the form of $p(s|\mathrm{F})$ are the following:

1) In the worst case, the attacker is able to fabricate exact replicas of the targeted biometric trait, and thus the distribution of fake scores is identical to the one of genuine user: $p(s|\mathrm{F}) = p(s|\mathrm{G})$. This is theonly scenario considered in [3]–[5].
2) In the best case (for the system), the fake trait is very different from the one of the targeted genuine user, such that the attacker does not get a better result than if he submitted his own original trait. Accordingly, in tis case $p(s|\mathrm{F}) = p(s|\mathrm{I})$.
3) In "intermediate" cases, we assume that $p(s|\mathrm{F})$ lies between $p(s|\mathrm{I})$ and $p(s|\mathrm{G})$, and model its possible shapes as discussed below.

We model $p(s|\mathrm{F})$ in "intermediate" cases using a parametric model based on a given distribution, like a Gaussian, Gamma or Beta. We assume that $p(s|\mathrm{F})$ has the same form as $p(s|\mathrm{G})$ and $p(s|\mathrm{I})$, and that the value of each of its parameters is between the values of the corresponding parameter in $p(s|\mathrm{G})$ and $p(s|\mathrm{I})$. For instance, if $p(s|\mathrm{G})$ and $p(s|\mathrm{I})$ are modeled as Gaussians with mean and variance denoted as $\mu_\mathrm{G}, \mu_\mathrm{I}, \sigma^2_\mathrm{G}$ and σ^2_I, we are assuming that $p(s|\mathrm{F})$ is Gaussian as well, and that its mean and variance satisfy the constraints:

$$\begin{aligned} \mu_\mathrm{F} &\in [\min\{\mu_\mathrm{G}, \mu_\mathrm{I}\}, \max\{\mu_\mathrm{G}, \mu_\mathrm{I}\}], \\ \sigma_\mathrm{F} &\in [\min\{\sigma_\mathrm{G}, \sigma_\mathrm{I}\}, \max\{\sigma_\mathrm{G}, \sigma_\mathrm{I}\}]. \end{aligned} \tag{1}$$

In other words, we model the possible fake score distributions as a "morphing" of the impostor distribution toward the genuine one. To simplify this model, we further constrain the parameters of $p(s|\mathrm{F})$ to satisfy a linear proportionality constraint with respect to their range, with the same value of the coefficient. For instance, in the case of Gaussian distributions this amounts to assume that the mean and variance of $p(s|\mathrm{F})$ is given by:

$$\begin{aligned} \mu_\mathrm{F} &= \alpha\mu_\mathrm{G} + (1-\alpha)\mu_\mathrm{I}, \\ \sigma_\mathrm{F} &= \alpha\sigma_\mathrm{G} + (1-\alpha)\sigma_\mathrm{I}, \end{aligned} \tag{2}$$

for some $\alpha \in (0, 1)$. From now on, we will explicitly denote the dependence of the fake scores distribution on α as $p(s|\mathrm{F};\alpha)$. Note that $\alpha = 1$ and $\alpha = 0$ lead respectively to the worst and best cases of assumptions 1 and 2. By varying α in $[0, 1]$ one obtains different distributions $p(s|\mathrm{F};\alpha)$: they approach $p(s|\mathrm{G})$ as α approaches 1 (the worst-case), and similarly they approach $p(s|\mathrm{I})$ as α approaches 0 (the best-case). Accordingly, we call the parameter α "fake strength".

Through the α parameter, our model allows one to take into account in the simplest possible way all the above mentioned different factors which can affect the fake scores distribution, in the absence of more precise information on their impact. In the next section we show how to apply this model to assess the robustness of a multi-modal biometric system under spoof attacks.

B. Robustness analysis of multi-modal systems under spoof attacks

Consider a system made up of N different matchers, whose scores are denoted as elements of a vector, $\mathbf{s} = (s_1, \ldots, s_N)$. Without loss of generality, assume that the first matcher is subject to a spoofing attack. Accordingly, the marginal distribution $p(s_1|\mathrm{F})$ follows the model described in Sect. III-A with respect to $p(s_1|\mathrm{G})$ and $p(s_1|\mathrm{I})$, while the marginal distributions of the other scores equal the ones of the corresponding impostor score, namely $p(s_i|\mathrm{F}) = p(s_i|\mathrm{I}), i > 1$, given that the corresponding traits are not subject to a spoofing attack. We further make the usual assumption that the scores are conditionally independent given the class, and extend it to the score of the spoofed trait. Accordingly, $p(\mathbf{s}|\mathrm{F})$ is modelled as $p(s_1|\mathrm{F}) \times \prod_{i=2}^{N} p(s_i|\mathrm{I})$. This model can be easily generalised to the case when more than one biometric trait is spoofed.

The distributions $p(s_i|\mathrm{G})$ and $p(s_i|\mathrm{I})$, $i > 1$, can thus be estimated from the available data set of scores using the chosen parametric model, as usual, while $p(s_1|\mathrm{F})$ is modelled as explained in Sect. III-A.

To assess the robustness of the multi-modal biometric system against a spoof attack, the threshold of the score fusion rule (and its parameters, if any) has to be estimated from training data, using the genuine and impostor score distributions, according to application requirements (for instance, setting the operational point by choosing a desired FAR value).

Algorithm 1 Procedure for robustness analysis of a multi-modal biometric system under spoof attacks

Inputs:

- A training set $(G_{\rm tr}, I_{\rm tr})$ and a testing set $(G_{\rm ts}, I_{\rm ts})$ made up of N-dimensional matching score vectors coming from genuine and impostor users;
- $f(\mathbf{s};\theta_f) \in \{\mathrm{G},\mathrm{I}\}$: a score fusion rule with parameters θ_f (including the threshold), where $\mathbf{s}$ is an input score vector and G and I denote the labels corresponding to the 'genuine' and 'impostor' decision;
- $\hat{P}(\cdot|\theta)$: a parametric model of the class-conditional genuine and impostor score distributions;
- $\alpha_1, \ldots, \alpha_n$: a set of fake quality values for the n matchers subject to a simulated spoof attack.

Output: The system's performance under a simulated spoof attack to matchers $1, \ldots, n$, with fake quality values $\alpha_1, \ldots, \alpha_n$.

1: Set the parameters of $f(\mathbf{s};\theta_f)$, on training data $(G_{\rm tr}, I_{\rm tr})$, according to given performance requirements.
2: Fit the model $\hat{P}(\cdot|\theta)$ to testing data $(G_{\rm ts}, I_{\rm ts})$, to approximate the genuine and impostor score distributions $\hat{P}(\mathbf{S}|\mathrm{G};\theta_{\rm G})$ and $\hat{P}(\mathbf{S}|\mathrm{I};\theta_{\rm I})$.
3: Compute the fake score distribution $\hat{P}(\mathbf{S}|\mathrm{F};\theta_{\rm F},\alpha)$ according to our model, using $\hat{P}(\mathbf{S}|\mathrm{G};\theta_{\rm G})$ and $\hat{P}(\mathbf{S}|\mathrm{I};\theta_{\rm I})$.
4: Randomly draw a set $F_{\rm ts}$ of scores from $\hat{P}(\mathbf{S}|\mathrm{F};\theta_{\rm F},\alpha)$, and label them as "impostors".
5: Evaluate the system's performance on the scores $(G_{\rm ts}, F_{\rm ts})$, using the score fusion rule $f(\mathbf{s};\theta_f)$.

The performance is then evaluated on testing data, using the genuine and fake score distributions. This procedure can be repeated for different α values in the range $[0,1]$, to get a complete picture of the system's performance as a function of the fake strength. Note that the above procedure can be carried out analitycally or numerically for some fusion rules and some parametric model of the score distributions. For instance, in the case of the LLR rule with Gaussian score distributions, the expression of the FAR as a function of the decision threshold can be obtained analitically (in integral form), and its evaluation can be done numerically, as shown in the next section. In general, the evaluation can always be carried out empirically. To this aim one must add to the genuine scores of the testing set a set of scores drawn from the distribution $p(\mathbf{s}|\mathrm{F})$, which has to be modelled based on genuine and impostor scores in the testing set. All the above procedure is summarised by Algorithm 1.

Under our model for $p(\mathbf{s}|\mathrm{F})$, the system's performance is likely to decrease from the value attained for $\alpha = 0$ (which corresponds to the absence of attacks) to the worst case corresponding to $\alpha = 1$. Therefore, the above procedure allows one to assess *how* the system's performance degrades as the attack strength increases: the more gracefully the performance degrades, the more robust a system is. This also allows one to compare the robustness of different fusion rules.

C. Case study: a bi-modal system using LLR fusion rule with Gaussian distributions

Here we show how to analitically/numerically evaluate the robustness of a bi-modal biometric system to a spoofing attack, when the score distributions are modelled as Gaussians, and the LLR rule is used. We assume that application requirements are given in terms of a desired FAR value.

The logarithm of the likelihood ratio for a system with two independent scores with class-conditional Gaussian distributions, denoted with $z(s_1, s_2)$, is given by:

$$z(s_1,s_2) = \log \frac{p(s_1|\mathrm{G})p(s_2|\mathrm{G})}{p(s_1|\mathrm{I})p(s_2|\mathrm{I})} = \log\left(\frac{\sigma_{\mathrm{I}_{s1}}\sigma_{\mathrm{I}_{s2}}}{\sigma_{G_{s1}}\sigma_{G_{s2}}}\right) + \frac{1}{2}\left[\frac{(s_1-\mu_{\mathrm{I}_{s1}})^2}{\sigma^2_{\mathrm{I}_{s1}}} + \frac{(s_2-\mu_{\mathrm{I}_{s2}})^2}{\sigma^2_{\mathrm{I}_{s2}}} - \frac{(s_1-\mu_{G_{s1}})^2}{\sigma^2_{G_{s1}}} - \frac{(s_2-\mu_{G_{s2}})^2}{\sigma^2_{G_{s2}}}\right]. \quad (3)$$

The decision function is given by $\mathrm{sign}\,(z(s_1,s_2) - \log t)$, where the value $+1$ means that the user is accepted as genuine, while a value of -1 means that he is rejected as an impostor. The threshold t has to be set so that the desired FAR is attained. The region of the score space (s_1, s_2) corresponding to genuine users, denoted as G, can be found analitically by solving the quadratic inequality $z(s_1,s_2) - \log t \geq 0$. The left-hand side of such inequality can be rewritten by re-arranging the terms of Eq. 3, as:

$$z(s_1,s_2) - \log t = As_1^2 + Bs_1s_2 + Cs_2^2 + Ds_1 + Es_2 + F, \quad (4)$$

where the threshold t is included in the term F. Depending on the value of $B^2 - 4AC$, the solution of $z(s_1,s_2) - \log t = 0$ corresponds to:

- $B^2 - 4AC < 0$: an ellipse;
- $B^2 - 4AC = 0$: a parabola;
- $B^2 - 4AC > 0$: an hyperbola.

This allows to find analitycally the region G.

The FAR for a given t value is defined as:

$$FAR(t) = \int\int_{\mathrm{G}} p(s_1|\mathrm{I})p(s_2|\mathrm{I})\mathrm{d}s_1\mathrm{d}s_2. \quad (5)$$

The above integral can be computed numerically.

Now the threshold t can be set to the value t^* which gives the desired FAR on training data. Assuming that s_1 corresponds to the matcher subject to a spoofing attack, the corresponding FAR on testing data can be found as:

$$FAR(t^*) = \int\int_{\mathrm{G}} p(s_1|\mathrm{F})p(s_2|\mathrm{I})\mathrm{d}s_1\mathrm{d}s_2, \quad (6)$$

where $p(s_1|\mathrm{F})$ and $p(s_2|\mathrm{I})$ are now obtained from testing data as described in the previous section. The above integral can be computed numerically as well.

IV. Experimental Results

In this section we report a case study involving the evaluation of the robustness of a bi-modal biometric system with the LLR fusion rule. The interest on the LLR rule is motivated by three main reasons: it is widely used in multi-modal systems; it is the optimal rule when the score distributions are exactly

known (in the sense that it gives the minimum false rejection rate, FRR, for any given FAR value, and vice-versa); its robustness to spoof attacks has already been questioned in previous works mentioned above.

A. Data set and experimental set up

We used the well known NIST biometric score set Release1 (BSSR1).[1] It contains raw similarity scores obtained from two different face matchers (denoted as C and G) and from one fingerprint matcher, both using left and right index (denoted as RI and LI), on a set of 517 people. For each individual, one genuine score and 516 impostor scores are available for each matcher and each modality.

We considered four different multi-modal systems by pairing in all possible ways the scores of the face and fingerprint matchers of the same individual. The resulting systems are therefore (Face G, Fingerprint LI), (Face G, Fingerprint RI), (Face C, Fingerprint LI), and (Face C, Fingerprint RI). In the following they will be denoted for short with the corresponding symbols: G-LI, G-RI, C-LI and C-RI. The scores were normalized using the hyperbolic tangent method [2].

In these experiments we used the whole set of scores both as training and testing data, which corresponds to the ideal situation in which the score distributions are exactly known. This allows us to evaluate the performance degradation due to spoofing attacks only, decoupling it from the effect of a mismatch between training and testing score distributions. To this aim, we applied Algorithm 1 with $G_{\mathrm{ts}} = G_{\mathrm{tr}}$, and $I_{\mathrm{ts}} = I_{\mathrm{tr}}$. In Algorithm 1 we used a Gaussian distribution to model the genuine and impostor score distributions, and computed the FAR values under a simulated spoof attack as described in Sect. III-C.

Three different operational points were considered: 0.01%, 0.1% and 1% FAR. The values we used for the fake strength α range form 0 to 0.1 with steps of 0.01, plus the values from 0.1 to 1 with steps of 0.1. Note that in this setting the simulated spoofing attacks affect only the FAR, while the FRR remains unchanged. Accordingly, the robustness of the considered multi-modal systems can be evaluated in terms of the behaviour of their FAR as a function of the fake strength.

B. Results

In Figs. 1–4 the FAR attained by the four multi-modal systems under a simulated spoofing attack on either the face or fingerprint matcher is shown, for each operational point, as a function of the fake strength α. Note that the FAR attained for $\alpha = 0$ is the one corresponding to the absence of attacks, while the one for $\alpha = 1$ corresponds to the worst-case considered in [3]. Note also that in each plot we reported the FAR vs α both for fingeprint and face spoofing only for the sake of space. However, using our method it is not possible to compare the FAR attained under spoof attacks against *different* biometrics, being equal the α value, as there is no relationship among the α values related to fake score distributions of different biometrics.

[1] http://www.itl.nist.gov/iad/894.03/biometricscores/index.html

As expected, the FAR under attack increases as the fake strength increases, namely as the simulated distribution of fake scores approaches to genuine score distribution.

The most interesting result is that in all the considered systems the increase in FAR is very steep when fingerprints are spoofed, to the extent that the FAR becomes unacceptably high even for low α values. For instance, in the G–RI system with 1% FAR operational point (Fig. 1, bottom), the FAR under attack exceeds 50% as the fake strength is above 0.15. This means that 1% of the impostors are erroneously recognised as genuines, when they provide their real fingerprint and face. Instead, when impostors provide a spoofed fingerprint of a genuine user together with their real face, 50% of them would be recognised as genuine users, as long as the mean and variance of the score distribution of the spoofed fingerprints is shifted from the one of the real impostors' fingerprints of just 15% towards the corresponding parameters of the genuine score distribution. Face spoofing causes instead a relatively more graceful increase of FAR as a function of the fake strength, in all the considered systems. Nevertheless, it always leads to FAR values exceeding 10%, for a sufficently high fake strength. The reason of this different behaviour is that the genuine and impostor score distributions of the face matchers in the considered data sets turn out to be more overlapping than the ones produced by the fingerprint matchers. Consequently, for a same value of the fake strength α, when the face is spoofed, the fingerprint matcher allows to detect a higher fraction of impostors than vice versa.

These preliminary results provide further evidence with respect to [3]–[5] that multi-modal systems can be vulnerable to spoof attacks against only one matcher. In particular, they can be very vulnerable also in more realistic, non-worst-case scenarios than the one considered in [3]–[5], when the spoofed traits are not perfect replicas of the real genuine traits. In particular, their performance can become unacceptable even when the fake score distribution is much closer to the impostor distribution than to the genuine one.

V. Conclusions

The main contribution of this work is a model of the matching score distribution produced by fake biometric traits under different possible realistic scenarios characterised by different spoofing techniques and different degrees of fakes' "strength" due to attackers' capability. Such factors are summarised in our model in a single parameter associated to the degree of similarity of the fake score distribution to the genuine one, which is named accordingly "fake strength". Our model allowed us to develop a method to empirically or analitically/numerically evaluate the robustness of multi-modal biometric systems to spoofing attacks, by simulating their effect on the matching scores.

We applied our robustness evaluation method to a case study involving bi-modal systems made up by a face and a fingeprint matcher, whose scores are fused using the well known LLR rule. Under the assumption of our model, experiments on

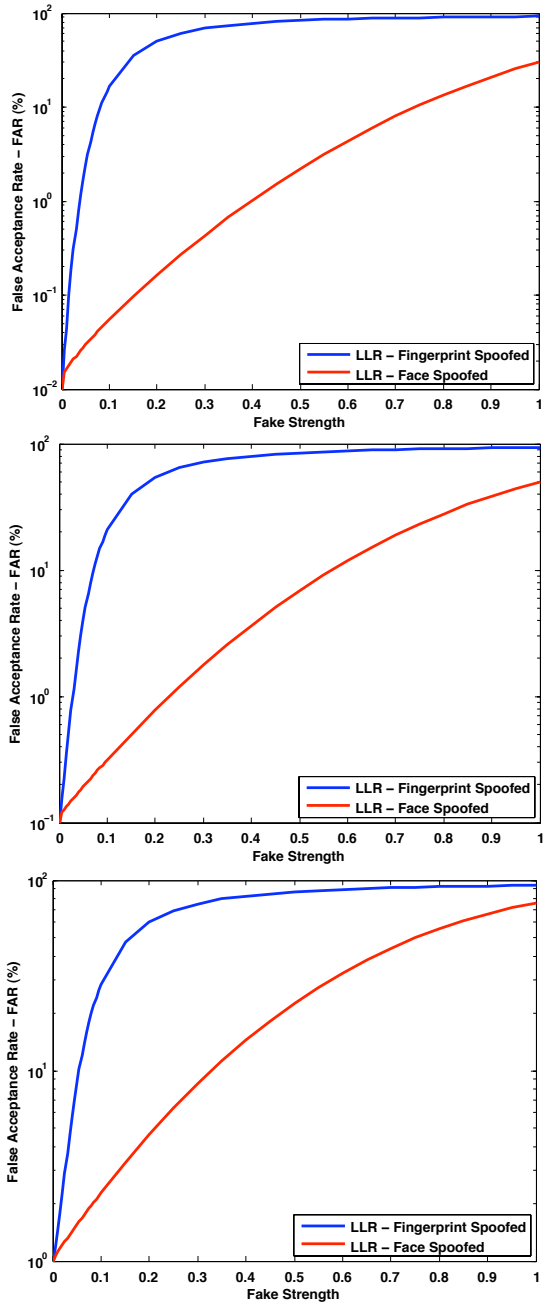

Fig. 1. FAR (%) of the G–RI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength α, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

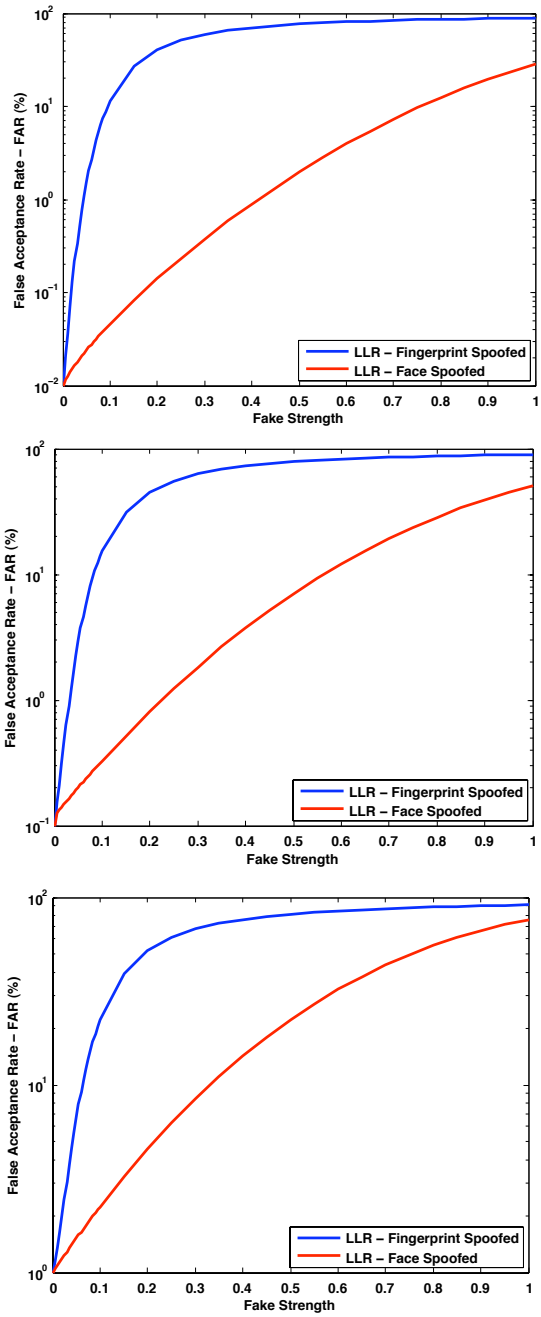

Fig. 2. FAR (%) of the G–LI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

several real data sets provided further evidence, besides [3]–[5], that multi-modal systems may be cracked by spoofing one only biometric trait, contrary to a common belief. They also showed that the LLR rule may be very vulnerable to such spoofing attacks, and that its vulnerability increases as the targeted matcher provides less overlapping score distributions.

Our results suggest several very interesting follow-ups:

- Constructing proper data sets containing spoofing attacks, to analyse the behaviour of the real distribution of fake scores under different conditions (different biometric traits, spoofing techniques, matchers, etc.). This would allow one to check whether the assumptions underlying our model provide good approximations of the fake score distributions, and to modify them if necessary. As a consequence, this would allow to make our method for robustness evaluation a practical tool for the designers of biometric systems, without requiring the actual implementation of spoofing attacks.
- Applying our method to compare the robustness of dif-

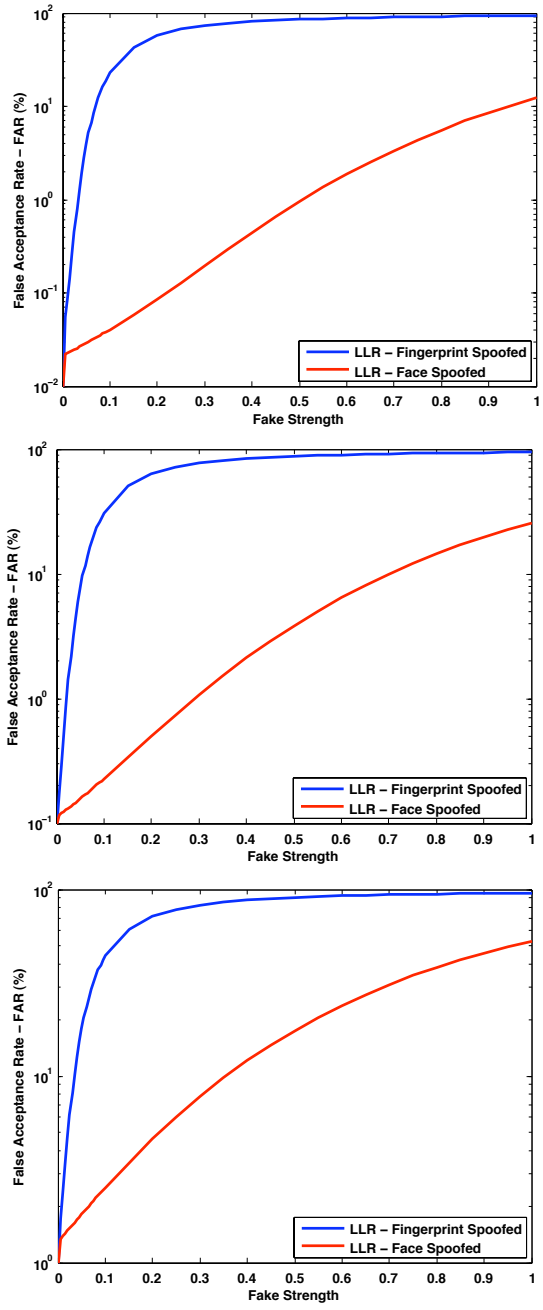

Fig. 3. FAR (%) of the C–RI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

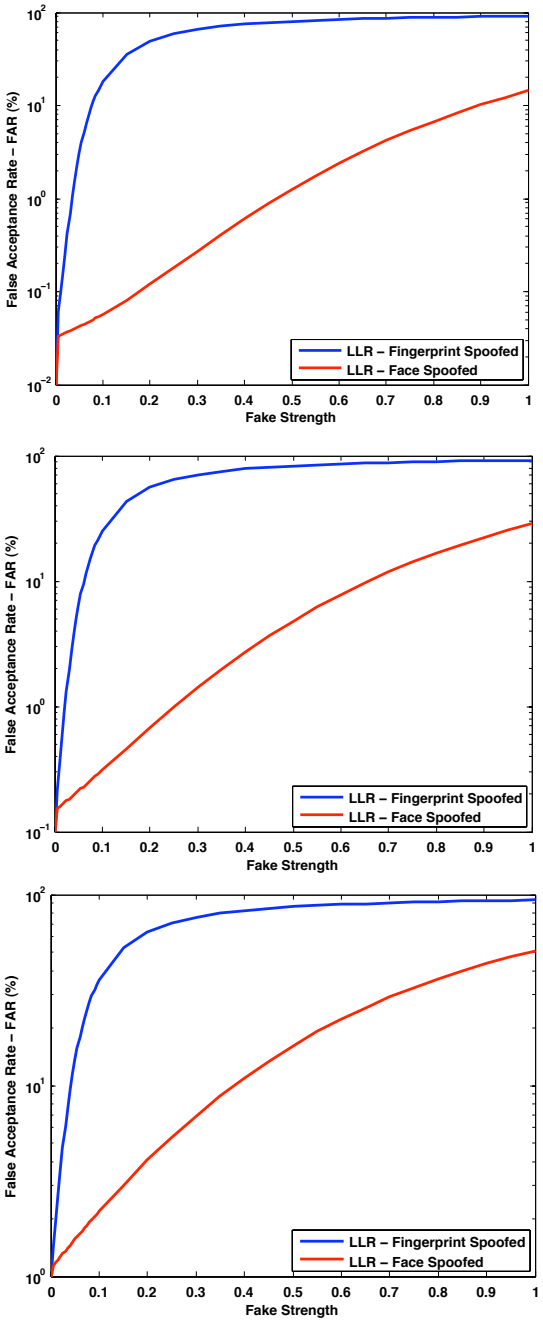

Fig. 4. FAR (%) of the C–LI system at 0.01 % (top) , 0.1 % (middle) and 1 % FAR (bottom), as function of the fake strength, when either the fingerprint (blue curve) or the face (red curve) is spoofed.

ferent score fusion rules to spoofing attacks.

- Finally, as our method allows to point out the vulnerabilities of score fusion rules to spoofing attacks, it could be exploited to develop proper countermeasure to improve their robustness.

ACKNOWLEDGMENT

This work was partially supported by the TABULA RASA project, 7th Framework Research Programme of the European Union (EU), grant agreement number: 257289; by the PRIN 2008 project “Biometric Guards - Electronic guards for protection and security of biometric systems” funded by the Italian Ministry of University and Scientific Research (MIUR); and by the RegioneAutonoma della Sardegna, through the Regional Law n. 7 for Fundamental and Applied Research, in the context of the funded project Adaptive biometric systems: models, methods and algorithms, grant n. CP4 442.

REFERENCES

[1] A.K. Jain, K. Nandakumar, A. Nagar. Biometric template security. *EURASIP J. on Adv. in Sig. Proc.*, pp. 1-17, 2008.

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

[3] R.N. Rodrigues, L.L. Ling, V. Govindaraju. Robustness of multi-modal biometric methods against spoof attacks. *J. of Visual Languages and Computing*, vol. 20, no. 3, pp. 169-179, 2009.

[4] R.N. Rodrigues, N. Kamat, V. Govindaraju. Evaluation of Biometric Spoofing in a Multimodal System. *Proc. Fourth IEEE Int. Conf. Biometrics: Theory Applications and Systems*, pp. 1-5, 2010.

[5] P.A. Johnson, B. Tan, S. Schuckers. Multimodal Fusion Vulnerability to Non-Zero Effort (Spoof) Imposters. *Proc. IEEE Workshop on Information Forensics and Security*, pp. 1-5, 2010.

[6] T. Matsumoto, H. Matsumoto, K. Yamada, S. Hoshino. Impact of artificial "gummy" fingers on fingerprint systems. *Optical Security and Counterfeit Deterrence Techniques IV*, vol. 4677 of Proc. of SPIE, pp. 275-289, 2002.

[7] P. Coli, G.L. Marcialis, F. Roli. Fingerprint silicon replicas: static and dynamic features for vitality detection using an optical capture device. *Int. J. of Image and Graphics*, vol. 8, no. 4, pp. 495-512, 2008.

[8] P. Coli, G.L. Marcialis, and F. Roli. Vitality detection from fingerprint images: a critical survey. *IEEE/IAPR 2nd International Conference on Biometrics ICB 2007*, pp. 722-731, 2007.

[9] G.L. Marcialis, F. Roli, and A. Tidu. Analysis of Fingerprint Pores for Vitality Detection. *Proc. of 20th IEEE/IAPR International Conference on Pattern Recognition (ICPR 2010)*, pp. 1289-1292, 2010.

[10] A. Abhyankar, and S. Schuckers. Integrating a wavelet based perspiration liveness check with fingerprint recognition. *Patt. Rec.*, vol. 42, no. 3, pp. 452-464, 2009.

[11] X. He, Y. Lu, and P. Shi. A Fake Iris Detection Method Based on FFT and Quality Assessment. Proc. Chinese Conf. on Pattern Recognition, pp. 1-4, 2008.

[12] Y. Kim, J. Na, S. Yoon, and J. Yi. Masked fake face detection using radiance measurements. *J. Opt. Soc. Am.* A 26, pp. 760-766, 2009.

[13] G.L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers. First International Fingerprint Liveness Detection Competition. *Proc. of 14th International Conference on Image Analysis and Processing ICIAP 2009*, pp. 12-23, 2009.

[14] G.L. Marcialis, F. Roli, and L. Didaci. Personal identity verification by serial fusion of fingerprint and face matchers. *Pattern Recognition*, vol. 42, no.11, pp. 2807-2817, 2009.

[15] K. Nandakumar, Y. Chen, S.C. Dass, and A.K. Jain. Likelihood ratio-based biometric score fusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342-347, 2008.

[16] G.L. Marcialis, F. Roli. Score-level fusion of fingerprint and face matchers under "stress" conditions. In: Proc. Int. Conf. Image Analysis and Proc. (ICIAP), pp. 259-264, 2007.