Evaluation of Multimodal Biometric Score Fusion Rules under Spoof Attacks

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

Abstract

Recent works have shown that multimodal biometric systems can be evaded by spoofing only a single biometric trait. In this paper, we propose a method to evaluate the robustness of such systems against spoofing attacks, when score-level fusion rules are used. The aim is to rank several score-level fusion rules, to allow the designer to choose the most robust one according to the model predictions. Our method does not require to fabricate fake biometric traits, and allows one to simulate different possible spoofing attacks using the information of genuine and impostor distributions. Reported results, using data set containing realistic spoofing attacks, show that our method can rank correctly score-level fusion rules under spoofing attacks.

1. Introduction

The so-called "spoofing attacks" [1, 2, 3, 4, 5, 6, 7] have shown to be an issue for mono- and multimodal biometric systems. In fact, even multimodal systems [10], claimed in the past as "intrinsically robust", can be evaded by a spoofing attack against only one of available modalities. These results point out that the robustness of multimodal systems to spoofing attacks can not be taken for granted, and needs to be carefully evaluated. However, evaluation is difficult due to several factors: materials employed for fabricating the spoof, the particular biometric under attack (fingerprint, face, etc.), the characteristics of the targeted client, etc. Moreover, collecting realistic and representative samples of spoofing may be very inefficient. Finally, different fusion methods (e.g. each score-level fusion rules) may exhibit different levels of robustness.

To the best of our knowledge, the evaluation of the robustness of multimodal fusion rules, and especially their "ranking" according to robustness, is still an open issue. In particular, a ranking of fusion rules could help the designer to select the most robust fusion rule under a range of different, possible spoofing attacks. Previous works in the literature [3, 4] have avoided the problem of fabricating fake samples by assuming a "worst-case" scenario, where the attacker is able to exactly replicate the genuine user's trait. This scenario was simulated by sampling match scores of the spoofed trait from the available genuine scores. However, in [6, 7, 8] it has been shown that this assumption is not representative of realistic spoofing attacks, and therefore does not provide an accurate prediction of the corresponding true performance drop. Consequently, an alternative model of the distribution of fake match scores has been proposed in [6, 9], based on the idea of evaluating robustness under several potential (simulated) distribution of fake scores, different from the worst-case one, without the need of fabricating spoofs. However, the performance prediction under attack provided by the above models is useful, only if one can give a reasonable approximation of the distribution of fake scores that a system will incur, which in practice is very difficult.

Accordingly, in this paper we propose to apply the model of [6] to a different, possibly more useful aim: to predict the *relative* robustness of several score-level fusion rules, namely the ranking of their performance under attack, for a range of different, simulated distributions of fake scores. By experiments on realistic spoofing attacks, we show that the ranking of different score fusion rules under a range of spoofing attacks can be predicted more reliably than their exact performance under a single attack. This allows the designer of a multimodal system to choose the most robust score fusion rule across a range of possible attacks.

The paper is organized as follows. In Sect. 2 previous works on the robustness of multimodal systems against spoofing attacks are reviewed. In Sect. 3 we present our method to rank score fusion rules under attacks. Experimental results are reported in Sect. 4. Conclusions are drawn in Sect. 5.

2. Background

In this section we give a short illustration of multimodal biometric systems, and summarize previous works on their robustness against spoofing attacks.

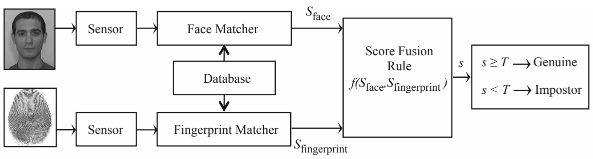

Figure 1. Outline of a multimodal biometric system made up of a face and a fingerprint matcher, with a score-level fusion rule.

2.1. Multimodal Biometric Systems

Fig. 1 illustrates the architecture of multimodal biometric systems, in reference to a bi-modal system composed of a face and a fingerprint matcher, that will be considered in this paper. The information coming from the two sensors can be integrated at the sensor, feature, match score, and decision level. Fusion at match score level is the most commonly adopted approach, which is also depicted in Fig. 1. The user provides his face and fingerprint to the respective sensors, and claims his identity. Then the system separately compares the two traits with the templates of the claimed identity provided at enrollment phase, and produces a face and a fingerprint match score, S_{face} and $S_{\text{fingerprint}}$, respectively. These scores are fused according to a given fusion rule $f(S_{\text{face}}, S_{\text{fingerprint}})$. If the fused score s is below a given acceptance threshold T, the user is classified as an impostor, otherwise it is classified as a genuine user.

2.2. Robustness of Multimodal Biometric Systems against Spoofing Attacks

Spoofing attacks are a major threat to biometric systems. For instance, in [1], 60% of fake fingerprints replicated using gum and gelatine were accepted as genuine. Similarly, it has been shown that face recognition systems can be evaded by showing simple photographs of genuine users [2]. Several liveness detection methods have been proposed as a potential countermeasure [11], but their common drawback is to increase the false rejection rate.

Although multimodal biometric systems were also commonly believed to be a natural countermeasure against spoofing attacks, based on a intuitive argument that evading them requires spoofing *all* traits simultaneously, some recent works provided experimental evidence that multimodal systems made up of two or three matchers can be evaded by spoofing *only one* biometric trait [3, 4, 5]. Most of the results were obtained by simulating the match scores coming from spoofing attacks, under the worst-case assumption that the attacker is able to replicate genuine traits perfectly, and thus that the match score distribution of fake traits is identical to the genuine one. However experiments carried out in [5] using real spoofed fingerprints, showed that the resulting score distribution of fake fingerprints was considerably different from the genuine one. Similar results were reported by the authors in [7, 8], on a data set of real fake fingerprints and fake faces. This means that the worst-case assumption is not always representative of realistic spoofing attacks.

The above results show that the robustness of multimodal systems to spoofing attacks should not be taken for granted, and needs to be carefully investigated.

3. Modelling the fake score distribution for robustness evaluation

A straightforward solution to evaluate the robustness of multimodal systems in realistic scenarios is to fabricate spoofing attacks. However, this is a time-consuming and difficult task [12]. The alternative solution proposed by the authors in [6] is to devise a simple model of the fake score distribution, that takes into account the fact that it is likely to lie between the genuine and the standard impostor score distributions, and that it can be affected by many factors such as the particular biometric trait spoofed, the forgery technique, the skill of the attacker, etc. However, at the state-of-the-art there is no clear information on how such factors affect the fake score distribution. For this reason, we proposed a very simple model of the fake score distribution, characterized by a single parameter which controls the relative distance to the impostor and genuine score distributions. This allows one to make a simple comparison of the robustness of different score fusion rules for different possible fake score distributions. We modelled fake scores by replacing each impostor score s_{I} with a fictitious score s_{F} obtained as:

$$s_{\text{F}} = (1 - \alpha)s_{\text{I}} + \alpha s_{\text{G}} \ , \tag{1}$$

where s_{G} is a randomly chosen genuine score, and $\alpha \in [0, 1]$ is the parameter that controls the relative distance of the resulting distribution to the ones of impostor and genuine users. Intuitively, the higher the α value, the closer the fake score distribution to the genuine one, and thus the more effective the spoofing attack; $\alpha = 0$ and $\alpha = 1$ lead respectively to the impostor and genuine score distribution. The latter corresponds to the worst case considered in [3, 4, 5].

A preliminary validation of the above model on realistic spoofing attacks was given in [6], on mono-modal systems based on a data set of faces and fingerprints. The main result was that, contrary to the worst-case assumption of [3, 4, 5], the model of [6] is capable to approximate reasonably well a realistic fake score distribution, and thus the corresponding FAR, provided that a suitable α value is chosen. However, different spoofing attacks may need a very different α value, which can not be known when a multi-modal system is being designed.

Based on the above results, in this work we propose to use our previous model not to predict the FAR of different score fusion rules under a specific spoofing attack, but to predict their *ranking* with respect to a range of potential

Algorithm 1 Prediction of the ranking of score fusion rules, based on their robustness to spoofing attacks.

Inputs: A multimodal system made up of N matchers;
A training set $(G_{\mathrm{tr}}, I_{\mathrm{tr}})$ and a testing set $(G_{\mathrm{ts}}, I_{\mathrm{ts}})$ made up of N-dimensional vectors of match scores coming from genuine (G) and impostor (I) users;
A set of score fusion rules;
A set of $\alpha = \{0, \alpha_1, ..., \alpha_n, 1\}$ values is defined.
Output: The ranking of score fusion rules according to their predicted robustness to spoofing attacks, as a function of the parameter α.

```
Set the parameters (if any) and the decision threshold $t$ of each fusion rule on training data $(G_{\mathrm{tr}}, I_{\mathrm{tr}})$, according to application requirements.
for each $\alpha$ value do
    Replace each score of the matcher under attack in $I_{\mathrm{ts}}$ with a fictitious fake score generated by Eq. (1). Let $F_{\mathrm{ts}}$ denote the set of fake scores.
    Evaluate the FAR of each score fusion rule on the scores $(G_{\mathrm{ts}}, F_{\mathrm{ts}})$.
    Rank the score fusion rules according to their FAR.
end for
```

attacks, namely different α values in the range $[0, 1]$, that lead to different (simulated) potential distributions of fake scores. This can give the designer of a multimodal system useful information about the relative robustness of different score fusion rules to spoofing attacks characterised by a different "effectiveness", namely by a fake score distribution more or less close to the one of genuine scores. The procedure is summarised in Algorithm 1.

In detail, given a multimodal system, a training set of genuine and impostor scores for each modality, and a set of score fusion rules, first the parameters of each rule (including the decision threshold T) are computed on training data, according to application requirements (for instance, they can be set to the zeroFAR operational point). The performance of the system in a standard operational setting, namely without any attack, can be evaluated on testing data. To evaluate robustness according to our method, a set of α values is chosen, and the impostor scores are replaced with a set of fictitious fake scores obtained by Eq. 1. The corresponding FAR of each score fusion rule can be then computed for each α value, and the different rules can be ranked in terms of their predicted FAR. Finally, the ranking of score fusion rules can be analyzed as a function of α.

4. Experiments

In this section we provide an experimental evaluation of the proposed robustness evaluation method of score fusion rules under spoofing attacks. The goal of these experiments is to assess whether our method can accurately predict the ranking of score fusion rules in terms of their FAR, when the α value that best fits a real fake score distribution is known, and how a choice between different rules can be made, based on their predicted robustness across the whole range of α values.

4.1. Data sets

We used two distinct data sets of faces and fingerprints previously collected by our research group, and fabricated spoofing attacks [6, 12]. Live face images were collected under various lighting and expression conditions, in two sessions separated by about two weeks interval. We fabricated the spoofed faces using the "photo attack" method as in [2]: we put in front of the camera the photo of each user displayed on a laptop screen. Fingerprint images were collected, using the Biometrika FX2000 optical sensor, from 40 volunteers aged between 20 and 70. Fake fingers were replicated by "consensual methods" as in [11, 12]: we used two-compound mixture of liquid silicon with a catalyst as cast while plasticine-like material as mold.

Since the two data sets have no users in common, we randomly combined their users to created a "chimerical" multimodal data set, which is made up of fictitious users with a face and a fingerprint coming from different real users. This data set is composed of 40 users, with 40 genuine samples and 40 spoofing attack samples per user.

The face and fingerprint recognition systems were implemented using the elastic bunch graph matching (EBGM) method [13] and the minutiae-based Neurotechnology's Verifinger 6.0, respectively. They produced match scores in the range $[0, 1]$ and $[0, 990]$, respectively. Since score normalization is necessary before using fusion rules at the score-level, the fingerprint scores were normalized into the range $[0, 1]$ using the hyperbolic tangent method [10].

4.2. Fusion Rules

We used three fixed score fusion rules (sum, product and Bayesian) and five trained ones (weighted sum, weighted product, perceptron, likelihood ratio, and the extended likelihood ratio of [3]). They are described below with reference to the bi-modal system used in the experiments, denoting the scores of the two matchers as S_1 and S_2.

Sum: the fused score is obtained as $s = S_1 + S_2$.

Product: $s = S_1 \times S_2$.

Bayesian: $s = \frac{S_1 \times S_2}{(1-S_1)(1-S_2)+(S_1 \times S_2)}$.

Weighted sum: $s = w \times S_1 + (1 - w) \times S_2$. The weight w was computed by maximising the system performance on a given operational point.

Weighted product: $s = S_1^w \times S_2^{1-w}$. The weight w was computed as described above.

Perceptron: $s = \frac{1}{1+exp[-(w_0+w_1 S_1+w_2 S_2)]}$, where weights w_0, w_1, and w_2 were computed by maximizing the

Fisher distance between the score distributions of genuine and impostor users.

Likelihood ratio (LLR): the threshold T is set on the ratio $\frac{p(S_1|\mathrm{G})p(S_2|\mathrm{G})}{p(S_1|\mathrm{I})p(S_2|\mathrm{I})}$, where $p(\cdot|\mathrm{G})$ and $p(\cdot|\mathrm{I})$ denote the match score distributions of genuine and impostor users, respectively, and the usual assumption that the scores of a given user are conditionally independent is considered. We modeled $p(\cdot|\mathrm{G})$ and $p(\cdot|\mathrm{I})$ using a Gamma distribution as in [3], which provided a rather good approximation of our data. This can be seen from the dissimilarity values between the normalized score histograms and the corresponding approximation, which was computed using the L1-norm Hellinger distance [14]. Given two distributions $f(x)$ and $g(x), x \in \mathcal{X}$, it is defined as $\int_{\mathcal{X}} |f(x)-g(x)|\mathrm{d}x \in [0, 2]$, where the value of 0 and 2 corresponds respectively to identical and non-overlapping distributions. Denoting with S_1 and S_2 respectively the face and fingerprint score, the Hellinger distance for $p(S_1|\mathrm{G})$, $p(S_2|\mathrm{G})$, $p(S_1|\mathrm{I})$ and $p(S_2|\mathrm{I})$ was respectively 0.30, 0.25, 0.17 and 0.26.

Extended LLR (ExtLLR): it is an extension of LLR, which was proposed in [3] to make it robust against spoofing attacks. The basic idea is to model the distribution of impostor scores of each matcher $p(\cdot|\mathrm{I})$ in the original LLR rule as a mixture of the original impostor distribution (which does not include spoofing attacks) and of a fictitious component corresponding to spoofing attacks against either of the matchers, or both of them. For each matcher, the component corresponding to a spoofing attack was assumed to be identical to the distribution of genuine scores, according to the worst-case assumption. We set the mixture parameters to the same values used in [3], and modelled $p(\cdot|\mathrm{G})$ and $p(\cdot|\mathrm{I})$ using a Gamma distribution as explained above. We refer the reader to [3] for further details.

Since no parameter tuning is required by our method, we used all the available data both as training and testing sets.

4.3. Experimental Results

In Table 1, we report the ranking of the eight score fusion rules considered, according to the FAR attained on testing data under the spoofing attacks in our data set, at two high security operational points that were chosen from training data (in absence of spoofing attacks): zeroFAR and 1% FAR on training data. Since under a spoofing attack the FAR changes, while the FRR does not (as spoofing attacks do not affect genuine scores), only the FAR values are reported.

To investigate whether realistic fake score distributions, and the corresponding FAR, can be reasonably approximated by the model of Sect. 3, for some α value, we computed the α value that minimizes their Hellinger distance. The corresponding α values are reported in Table 2.

We first evaluated the accuracy of our method in approximating the FAR of a multimodal system under attack, as a function of α. Table 3 shows the FAR attained by the LLR rule for α values $0.1, 0.2, \ldots, 1.0$, and the ones of Table 2. Qualitatively similar results were obtained with the other fusion rules, and are not reported for lack of space. Our model provides a good approximation of the real FAR under attack, provided that the optimal α value is used, in the case of face spoofing. The approximation is not as accurate in case of fingerprint spoofing: the real FAR is overestimated (when the optimal α value is used) by an amount of about 16-18%. The FAR predicted by the worst-case assumption of [3, 4, 5] (corresponding to $\alpha = 1$) is as accurate as the one provided by the considered model, for face spoofing, but is much more inaccurate for fingerprint spoofing, where it overestimates the FAR of about 40 to 50%. This is an evidence that our model is more appropriate than the one based

Face Spoofing			
zeroFAR		1% FAR	
FAR(%)	Rules	FAR(%)	Rules
0.04	ExtLLR	2.26	ExtLLR
0.05	LLR	2.29	LLR
0.27	W. Product	10.72	W. Product
0.48	W. Sum	18.37	W. Sum
1.30	Perceptron	20.95	Perceptron
6.75	Bayesian	23.47	Bayesian
6.80	Sum	23.49	Sum
6.82	Product	23.57	Product
Fingerprint Spoofing			
zeroFAR		1% FAR	
FAR(%)	Rules	FAR(%)	Rules
0.00	Bayesian	1.05	Bayesian
0.00	Sum	1.15	Sum
0.00	Product	1.33	Product
24.56	W. Sum	42.59	W. Sum
27.73	Perceptron	44.11	Perceptron
34.87	W. Product	51.10	W. Product
50.42	ExtLLR	60.31	ExtLLR
50.43	LLR	60.32	LLR

Table 1. Ranking of fusion rules according to their FAR under realistic spoofing attacks, when either the face (top) or the fingerprint is spoofed (bottom), at two operational points.

Data set	Hellinger distance	α
Face	0.0939	0.9144
Fingerprint	0.4397	0.0522

Table 2. Minimum values of the Hellinger distance between the real distributions of fake scores and the ones obtained by the considered model, as a function of α, for the face and fingerprint data sets. The corresponding α value is also shown.

Face Spoofing		
$\alpha = 0$	zeroFAR	1% FAR
0.1	0.00	1.09
0.2	0.00	1.15
0.3	0.01	1.26
0.4	0.01	1.35
0.5	0.01	1.45
0.6	0.01	1.59
0.7	0.01	1.76
0.8	0.02	2.01
0.9	0.04	2.24
0.9144	**0.04**	**2.33**
1	0.06	2.68
Realistic attack	0.05	2.29
Fingerprint Spoofing		
$\alpha = 0$	zeroFAR	1% FAR
0.0522	**66.04**	**78.04**
0.1	86.94	91.28
0.2	95.06	97.28
0.3	97.90	99.08
0.4	99.12	99.43
0.5	99.34	99.70
0.6	99.63	99.79
0.7	99.73	99.84
0.8	99.75	99.86
0.9	99.76	99.87
1	99.85	99.89
Realistic attack	50.43	60.32

Table 3. FAR (%) attained by the multimodal system under a simulated spoofing attack against the face (top) and the fingerprint matcher (bottom), as a function of α, using LLR rule, at two operational points. The FAR under the α value that best fitted the real fake score distributions (see Table 2) is shown in boldface.

on the worst-case assumption.

Note that in our data set fingerprint spoofing leads to a higher increase in FAR than face spoofing. This is due to the fact that the genuine and impostor score distributions of the face matcher turned out to be more overlapping than the ones produced by the fingerprint matcher.

Lastly, note that the FAR increases very quickly as a function of α in the case of fingerprint spoofing, up to the extent that the performance drops considerably even for low α values. This suggests that the LLR rule, which in principle is the optimal fusion rule under normal operational conditions, can be very vulnerable to spoofing attacks. Similar results were observed for the other fusion rules.

Consider now how our method predicts the ranking of score fusion rules. First, we compared the real ranking of Table 1, and the one predicted by the model for the optimal α value of Table 2. We found that our method always predicted the correct ranking corresponding to the optimal α value. This is an evidence that our method is capable to provide a reliable ranking prediction for any given α value, if the fake score distribution is best approximated by our model using such value. Since in practice a multimodal system can be subject to different attacks, that are best approximated by different α values, and that such values are unknown, the above result suggests that our method can nevertheless be used to choose the score fusion rule that exhibits the highest "average" robustness across the whole range of α values. To better explain this point, consider the predicted ranking for the considered fusion rules as a function of α, which is reported in Fig. 2 for the zeroFAR and 1% FAR (the ranking was identical for both operational points). Under fingerprint spoofing, the predicted ranking of each rule remains constant, and the Bayesian rule always exhibits the best ranking. This suggests that the Bayesian rule should be a good choice in terms of robustness, even if the designer does not know the optimal α value for any given attack. For face spoofing, two different rankings are predicted instead: one for $\alpha < 0.5$, and the other $\alpha \geq 0.5$. The latter corresponds to the one observed under a real spoofing attack (Table 1, top). The Bayesian and the ExtLLR rules are the top-ranking ones in the two intervals, respectively. However, except for the weighted sum and weighted product rules, that exhibit a constant or almost constant, and rather

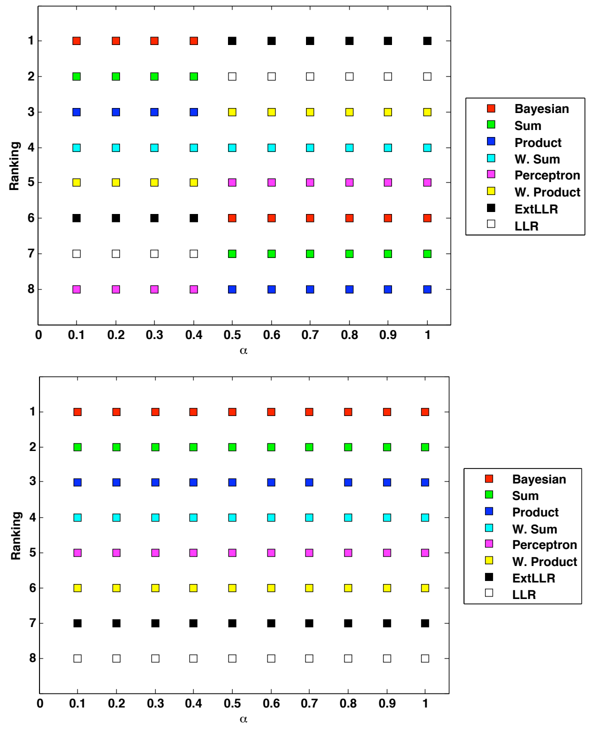

Figure 2. Ranking of the considered fusion rules as a function of α , when only face (top) or fingerprint (bottom) is spoofed, at the zeroFAR and 1% FAR (the ranking was identical for both operational points).

high ranking in both intervals, the ranking of the other rules drastically changes. This suggests that the weighted sum or the weighted product rule is a reasonable choice to avoid the risk of a very low performance under attack, unless the fake score distribution of possible face spoofing attacks is believed to be either close to the impostor one (namely, it is best approximated by a low α value), in which case the Bayesian rule is the best choice, or it is believed to be close to the genuine user distribution, and in this case the ExtLLR rule is the best choice.

To sum up, our preliminary results provide some evidence that the ranking of different fusion rules in terms of their robustness (FAR) across different potential spoofing attacks, can be reliably predicted by exploiting the model of fake score distribution of [6]. This provides the designer of a multimodal system a way to choose a score fusion rule, taking into account its robustness to spoofing attacks.

5. Conclusion

We proposed a simple method to rank biometric score fusion rules in terms of their robustness against spoofing attacks. Our method uses the information of genuine and impostor distributions only, and does not need to fabricate fake biometrics. It is based on a simple model of the fake score distribution previously proposed by the authors, that takes into account different, potential spoofing attacks, and is more realistic than a model base on a "worst-case" assumption proposed in [3]. Preliminary results provide some evidence that our method is capable to provide a reliable ranking of fusion rules. An interesting follow-up of our work is the development of more realistic models of the fake score distribution, that could further improve the effectiveness of the proposed method.

Acknowledgment

This work was partially supported by: the TABULA RASA project, 7th Framework Research Programme of the European Union (EU), grant agreement n. 257289; 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); the project "Adaptive biometric systems: models, methods and algorithms", funded by the Regione Autonoma della Sardegna, Regional Law n. 7 for Fundamental and Applied Research, grant n. CP4 442.

References

[1] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of Artificial 'gummy' Fingers on Fingerprint Systems", *Opt. Sec. Counterfeit Deterrence Tech. IV, Proc. SPIE Vol. 4677*, pp. 275-289, 2002.

[2] Z. Zhang, D. Yi, Z. Lei, and S.Z. Li, "Face Liveness Detection by Learning Multispectral Reflectance Distributions", *IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 436-441, 2011.

[3] R. N. Rodrigues, L.L. Ling, and V. Govindaraju. "Robustness of multimodal biometric methods against spoofing attacks". *J. Vis. Languages and Computing*, vol. 20, pp. 169-179, 2009.

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

[5] R. N. Rodrigues, N. Kamat, and V. Govindaraju, "Evaluation of Biometric Spoofing in a Multimodal System", *Int. Conf. Biometrics*, pp. 1-5, 2010.

[6] Z. Akhtar, G. Fumera, G.L. Marcialis, and F. Roli, "Robustness Evaluation of Biometric Systems under Spoofing Attacks", *Int. Conf. Image Analysis and Processing*, pp. 159-168, 2011.

[7] B. Biggio, Z. Akthar, G. Fumera, G.L. Marcialis, and F. Roli, "Robustness of multimodal biometric verification systems under realistic spoofing attacks", *Int. Joint Conf. Biometrics*, 2011.

[8] Z. Akhtar, B. Biggio, G. Fumera, G. L. Marcialis, "Robustness of Multimodal Biometric Systems under Realistic Spoofing Attacks against All Traits", *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, pp. 5-10, 2011.

[9] Z. Akhtar, G. Fumera, G. L. Marcialis and F. Roli, "Robustness analysis of Likelihood Ratio score fusion rule for multimodal biometric systems under spoofing attacks", *Int. Carnahan Conf. Security and Technology*, pp. 237-244, 2011.

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

[11] P. Coli, G. L. Marcialis, and F. Roli. "Vitality detection from fingerprint images: a critical survey". *Int. Conf. Biometrics*, pp. 722-731, 2007.

[12] G. L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers, "First Int. Finger. Liveness Detection Competition", *Int. Conf. Image Analysis and Proc.*, pp. 9-12, 2009.

[13] D. S. Bolme, "Elastic Bunch Graph Matching", M.Sc. Thesis, Colorado State University, 2003.

[14] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*, Springer, 1986.