Abdenour Hadid, Nicholas Evans, Sébastien Marcel, and Julian Fierrez

Biometrics Systems Under Spoofing Attack

An evaluation methodology and lessons learned

©STOCKPHOTO.COM/GREYFEBRUARY

Biometrics Security and Privacy Protection

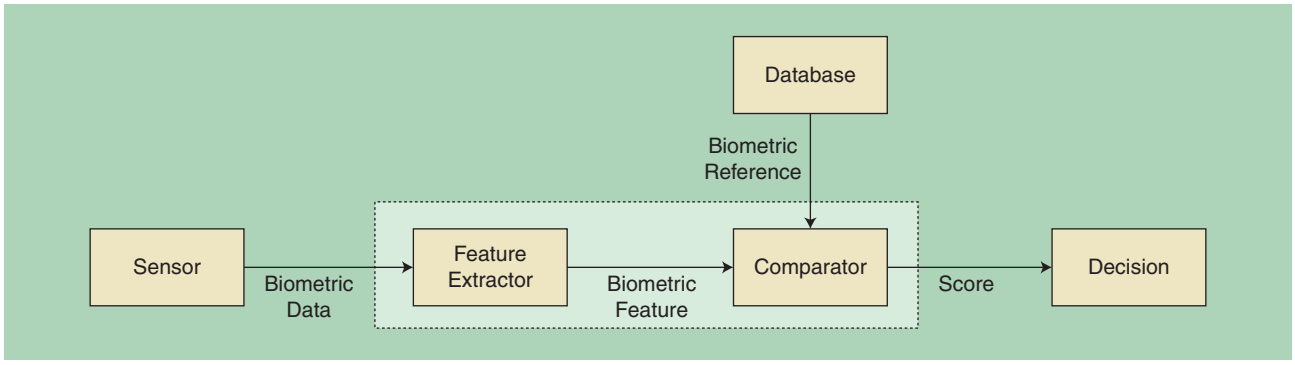
Biometrics already form a significant component of current and emerging identification technologies. Biometrics systems aim to determine or verify the identity of an individual from their behavioral and/or biological characteristics. Despite significant progress, some biometric systems fail to meet the multitude of stringent security and robustness requirements to support their deployment in some practical scenarios. Among current concerns are vulnerabilities to spoofing—persons who masquerade as others to gain illegitimate accesses to protected data, services, or facilities. While the study of spoofing, or rather antispoofing, has attracted growing interest in recent years, the problem is far from being solved and will require far greater attention in the coming years. This tutorial article presents an introduction to spoofing and antispoofing research. It describes the vulnerabilities, presents an evaluation methodology for the assessment of spoofing and countermeasures, and outlines research priorities for the future.

Digital Object Identifier 10.1109/MSP.2015.2437652
Date of publication: 13 August 2015

INTRODUCTION

The provision of security can entail the protection of sensitive data, services, or facilities by ensuring that only authorized

1053-5888/15©2015IEEE



[FIG1] **A generic biometric system.**

persons have access. Though passwords provide some protection against illegitimate access, they are often so simple that they can be guessed or easily cracked. While offering improved security, complex passwords can be difficult to remember and consequently often "stored" via less secure means. Furthermore, the same password is often used across multiple applications or platforms meaning a cracked password can enable a fraudster to access multiple resources.

An attractive alternative to passwords involves biometric recognition. Biometrics refer to a person's behavioral and biological characteristics such as their face, fingerprint, iris, voice, hand geometry, and gait. Biometric traits can be highly discriminative yet less easily lost or stolen [1]. Despite their appeal, however, biometric systems are vulnerable to malicious attacks [2]. Among them are spoofing attacks, also called *presentation attacks*, which refer to persons masquerading as others to gain illegitimate access to sensitive or protected resources. As an example, a fraudster could fool or spoof a face-recognition system using a photograph, a video, or a three-dimensional (3-D) mask bearing resemblance to a legitimate individual.

Even though the threat of spoofing is now well recognized, the problem is far from being solved, thus antispoofing research warrants far greater attention in the future. This tutorial article introduces the problem of spoofing and related research to develop antispoofing solutions. The focus is an evaluation methodology for assessing both the effect of spoofing and the performance of spoofing countermeasures. A case study in face recognition is included to illustrate the application of the evaluation methodology in practice. Finally, the article also includes a summary of the lessons learned through our own research and outlines a number of research priorities for the future. Most of the material is based upon antispoofing research performed in the scope of the European TABULA RASA research project (http://www.tabularasa-euproject.org), which was identified as a success story by the European Commission (http://europa.eu/rapid/press-release_MEMO-13-924_en.htm).

The presentation is self-contained and aimed at both the generally knowledgeable and nonspecialist. The article aims to provide an overview of the research problem, not a comprehensive survey of the plethora of antispoofing techniques in the literature; such surveys can be found elsewhere, e.g., for fingerprint recognition [3], face recognition [4], and speaker recognition [5]. The intention is also to stimulate further work, particularly the development of standard metrics, protocols, and data sets for evaluating progress.

BIOMETRICS

The term *biometrics* is derived from the Greek words *bio* (life) and *metric* (to measure). The goal of a biometric recognition system is to determine or verify the identity of an individual from his/her behavioral and/or biological characteristics. Applications include criminal identification, airport checking, computer or mobile device log-in, building and critical infrastructure access control, digital multimedia rights control, transaction authentication, voice mail, and secure teleworking. Various biometrics have been investigated, from the most conventional including fingerprint, iris, face, and voice, to more emerging modalities such as gait, hand-grip, ear, and electroencephalograms. Each modality has its own strengths and weaknesses [1]. For example, face recognition is among the most socially accepted biometric; face recognition is a natural method of identification used everyday by humans. In contrast, while fingerprint and iris recognition may be more reliable, they are also more intrusive. In practice, the choice of biometric modality depends on the application.

Biometric systems typically function in one of two distinct modes: 1) verification (or authentication) and 2) identification. An authentication system aims to confirm or deny a claimed identity (one-to-one matching), whereas an identification system aims to identify a specific individual (one-to-many matching). Although there are some differences between the two modes, their most basic operation, namely that of feature-to-reference comparison, is identical and consists of the following steps illustrated in Figure 1.

First, a biometric sample (e.g., a face image) is acquired from a sensor (e.g., a digital camera). Biometric features (e.g., facial intensity, color, or texture) are then extracted from the sample. These can be a set of parameters (or coefficients) that provide a compact representation of the biometric sample, which is more discriminative and amenable to pattern recognition. Biometric features should minimize variations due to acquisition or environmental factors (e.g., facial expression, pose, and illumination)

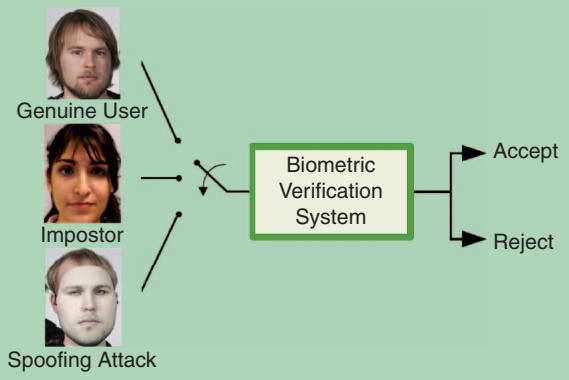
while discriminating between the biometrics collected from different individuals.

To determine or verify the identity corresponding to a given biometric sample, the features are compared to a single (verification) or set of (identification) biometric references acquired previously during an enrollment phase. These comparisons are made by a comparator that produces a score reflecting the similarity between features and references. The decision is an acceptance or rejection in the case of verification, or the identity of the closest match in the case of identification.

(a)

(b)

[FIG2] Example spoofing attacks: (a) face-recognition spoofing using a 3-D face mask; (b) fingerprint-recognition spoofing using a fake fingerprint.



[FIG3] A biometric verification system should accept genuine users but reject both zero-effort impostors and concerted-effort spoofing attacks.

SPOOFING ATTACKS

It is now widely acknowledged that biometric systems are vulnerable to manipulation. There are two broad forms of attack: direct and indirect. Direct attacks, also referred to as *spoofing* or *presentation attacks*, are performed at the sensor level, outside the digital limits of the biometric system. Indirect attacks, however, are performed within the digital limits by intruders such as cybercriminal hackers. These attacks may attempt to bypass the feature extractor or the comparator, to manipulate biometric references, or to exploit vulnerabilities in the communications channels.

Whereas traditional digital protection mechanisms such as encryption can be deployed to prevent indirect attacks, they cannot prevent direct attacks. Furthermore, indirect attacks require specialist expertise or equipment, whereas direct attacks such as those illustrated in Figure 2 can be implemented by the layman. Direct attacks are therefore of significant concern.

Vulnerabilities to spoofing are a barrier to the successful exploitation of biometrics technology. Unfortunately, some vulnerabilities have been exposed and well publicized in the international media. One of the earliest examples was demonstrated at Black Hat 2009, the world's premier technical security conference. Researchers from the Security and Vulnerability Research Team at the University of Hanoi (Vietnam) showed how the face-recognition user authentication systems introduced by three different laptop computer manufacturers could be spoofed or bypassed using photographs of legitimate users. This vulnerability is now listed in the National Vulnerability Database maintained by the National Institute of Standards and Technology (NIST) in the United States. More recently, the Chaos Computer Club, a German hacking collective, showed how an artificial finger could be used to spoof a fingerprint-user authentication system developed by one of the world's most popular smartphone manufacturers.

The typical countermeasure deployed to detect spoofing attacks involves liveness detection: systems that aim to detect signs of life. Since it is inherently more difficult to spoof multiple modalities and systems simultaneously [6], multimodal biometric systems have also been investigated as a solution to spoofing. Even so, spoofing remains a serious cause for concern, with many biometric systems remaining vulnerable even to the simplest forms of spoofing attack. Unfortunately, research to develop spoofing countermeasures is still in its infancy and warrants considerably greater attention in the future.

EVALUATION METHODOLOGY: VULNERABILITIES TO SPOOFING

This section describes the first component of the proposed methodology for the evaluation of biometric systems and its vulnerabilities to spoofing. For simplicity, the following considers only biometric verification (one-to-one matching). Traditionally, biometric systems are evaluated using large data sets of representative biometric samples with which performance is assessed according to a standard protocol of genuine and impostor trials. An evaluation essentially measures the proportion of these trials

that the system misclassifies, i.e., genuine trials classified as impostor trials and vice versa. This approach, involving only casual impostors, equates to performance in the face of zero-effort spoofing attacks, i.e., where impostors make no effort to replicate the biometric traits of another, legitimate individual.

Spoofing is accomplished using fake biometric samples expressly synthesized or manipulated to provoke artificially high comparator scores. Biometric systems should be robust to both zero-effort impostor trials and concerted-effort spoofing attacks. This ternary scenario is illustrated in Figure 3. The biometric system should produce high scores in the case of genuine trials, but low scores in the case of impostor and spoofed access attempts.

Assessment requires biometric data of three categories comprising genuine, zero-effort impostor, and spoofed trials. Biometric data are typically further divided into three nonoverlapping subsets: a training set, a development set, and a test set, the purpose of which is as follows.

- The training set is used to train the biometric system (e.g., the learning of background models) or to develop spoofing countermeasures.
- The development set is used for decision threshold optimization and the a posteriori performance at a given operating point. For each identity, the development set is partitioned into two subsets:
  - an enrollment subset used to create biometric references or models for each identity
  - a probe subset used for biometric comparisons (genuine, zero-effort impostor and spoofing attack trials).
- The test set is used to compute the a priori performance given the threshold determined from the development set. The test set is similarly partitioned into enrollment and probe subsets.

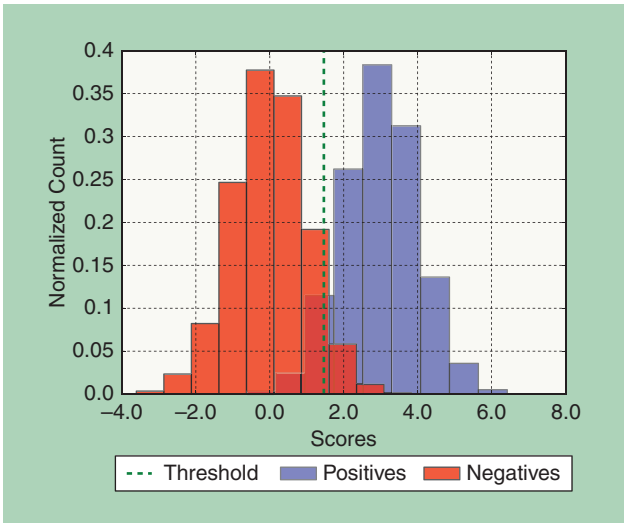
Two assessment scenarios can then be defined:

- a licit scenario for assessing baseline performance using genuine and zero-effort impostor trials
- a spoofing scenario for assessing vulnerabilities to spoofing.

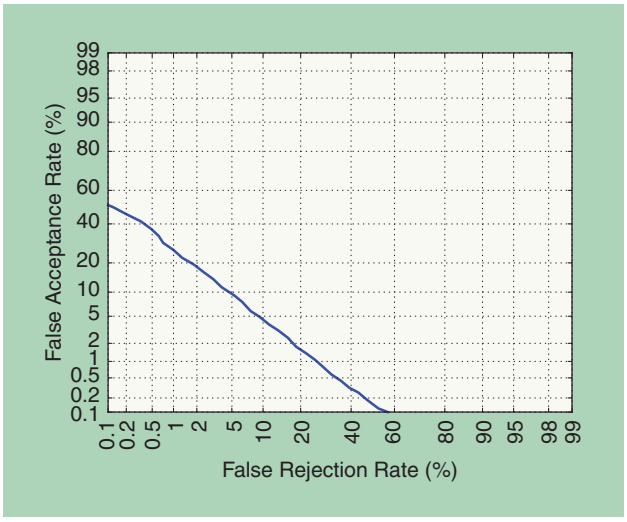
Figure 4 illustrates example score distributions for a licit scenario (i.e., genuine and impostor trials with no spoofing attacks). With only a little overlap between the two distributions, there is high potential to distinguish between genuine and impostor trials. Nonetheless, the system will still make mistakes, leading to two different error rates:

- the false rejection rate (FRR), which reflects the proportion of genuine trials misclassified as zero-effort impostor trials
- the false acceptance rate (FAR), which reflects the proportion of zero-effort impostors trials misclassified as genuine trials.

Both the FAR and FRR are dependent upon a decision threshold τ, an example of which is illustrated by the vertical, dashed line in Figure 4. Impostor trails corresponding to a score higher than this threshold will be misclassified as genuine trials, whereas genuine trials with a score lower than this threshold will be misclassified as impostor trials.



[FIG4] Score distribution of genuine users (positives) and zero-effort impostors (negatives) under a licit scenario. A decision threshold τ is illustrated by the vertical, dashed line.



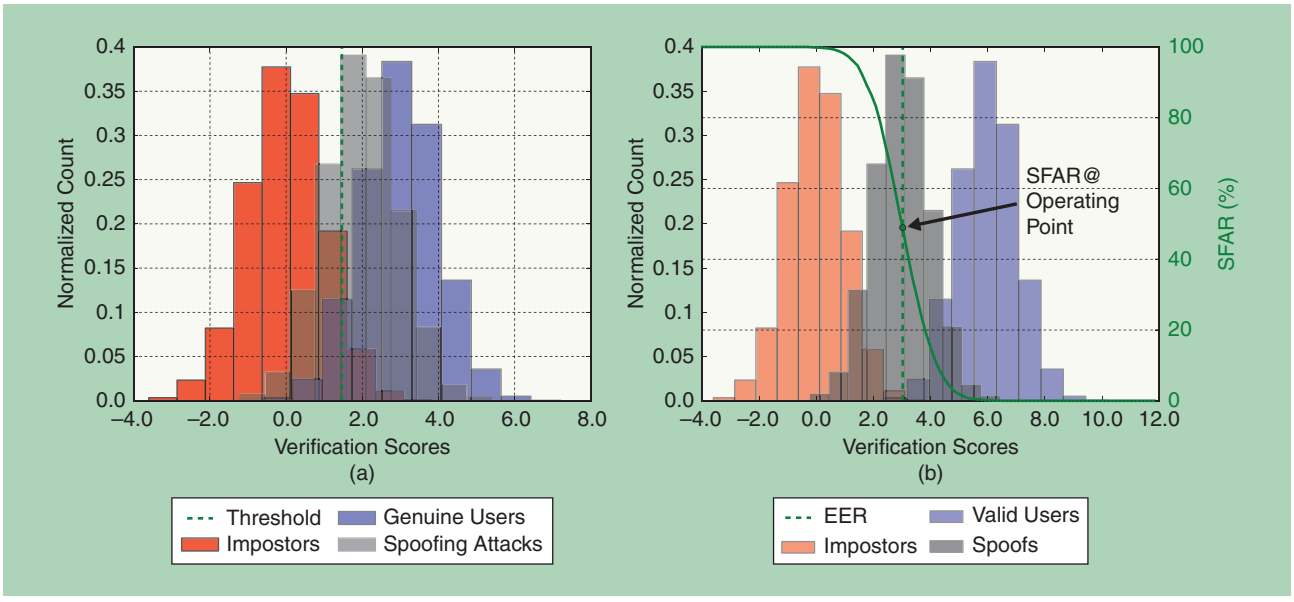
[FIG5] An example of a detection error tradeoff (DET) plot showing an EER in the order of 6%.

Since the two errors are inversely related, it is often desirable to illustrate performance as a function of the threshold τ. For a specific data set $\mathcal{D}$, one such measure is the half total error rate (HTER):

$$\mathrm{HTER}(\tau, \mathcal{D}) = \frac{\mathrm{FAR}(\tau, \mathcal{D}) + \mathrm{FRR}(\tau, \mathcal{D})}{2}. \qquad (1)$$

Performance can also be illustrated with detection-error tradeoff (DET) profiles, an example of which is illustrated in Figure 5. DET profiles illustrate the behavior of a biometric system for varying decision thresholds, τ, and show the tradeoff between the FAR and the FRR.

In practice, the decision threshold is chosen to minimize an a posteriori performance criteria, which is normally application



[FIG6] Example score distributions: (a) genuine users, zero-effort impostors, and (b) spoofing attacks with the SFAR as a function of the decision threshold τ.

dependent. Without focusing on a specific application, the equal error rate (EER) is also a common reference metric. For a development data set $\mathcal{D}_{\text{dev}}$, the EER is given by

$$\tau^*_{\text{EER}} = \arg\min_{\tau} |\,\text{FAR}(\tau, \mathcal{D}_{\text{dev}}) - \text{FRR}(\tau, \mathcal{D}_{\text{dev}})\,|, \qquad (2)$$

where the threshold τ^*_{EER} is set to equalize the FAR and FRR. The most reliable estimate of a priori performance is then determined using a test set $\mathcal{D}_{\text{test}}$ using the predetermined threshold. For example, the HTER is determined using the decision threshold τ^*_{EER} according to

$$\text{HTER}(\tau^*_{\text{EER}}, \mathcal{D}_{\text{test}}) = \frac{\text{FAR}(\tau^*_{\text{EER}}, \mathcal{D}_{\text{test}}) + \text{FRR}(\tau^*_{\text{EER}}, \mathcal{D}_{\text{test}})}{2}. \qquad (3)$$

Performance in the face of spoofing is assessed with the spoofing scenario, i.e., by replacing the subset of zero-effort impostor trials with spoofed trials. Figure 6(a) illustrates example score distributions for a population of genuine, zero-effort impostor and spoofed trials. The shift between the impostor and spoofed trial distributions illustrates the likely effect of spoofing on biometric recognition performance; the overlap between the score distributions for genuine and spoofed trials is significantly greater than that between genuine and impostor distributions.

A quantitative measure of system vulnerability can be expressed in terms of the spoof FAR (SFAR) which reflects the percentage of spoofed trials misclassified as genuine trials given a decision threshold τ. An example SFAR profile is illustrated as a function of the threshold τ in Figure 6(b).

The vulnerability to spoofing is thus reflected in the difference between the SFAR (spoof scenario) and the FAR (licit scenario), again as a function of τ. These differences are illustrated in the example DET plots for both licit (FAR versus FRR) and spoof (SFAR versus FRR) scenarios in Figure 7. Finally, it is often of interest to express system vulnerability for a specific operating point defined by a given FRR (e.g., the EER). System vulnerability is then given by the difference between the FAR and the SFAR for the same FRR (vertical, dashed line in Figure 7). In this illustrative example for an FRR in the order of 6%, an FAR of under 10% increases to an SFAR of over 60%. This system would misclassify approximately two in three spoofed trials as genuine accesses.

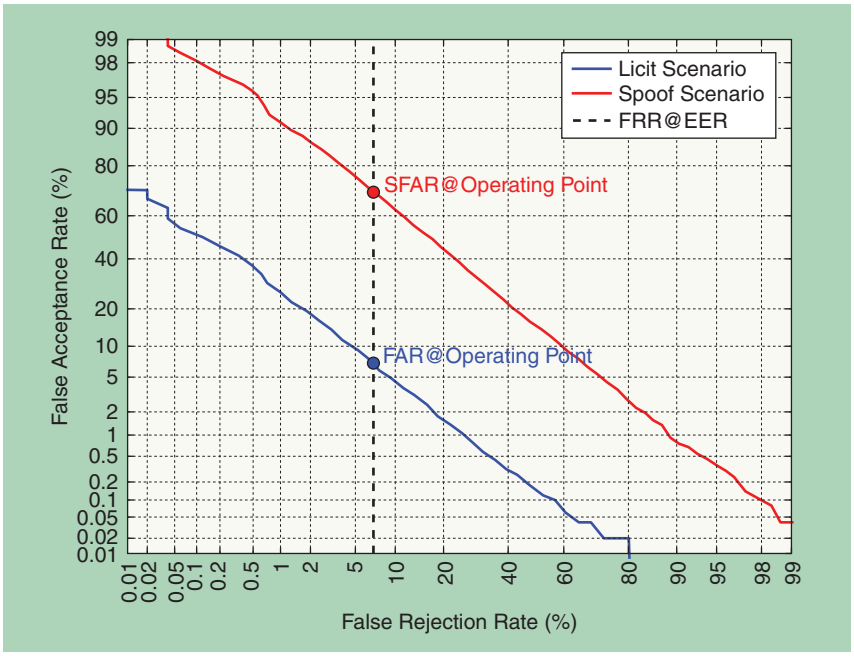
EVALUATION METHODOLOGY: SPOOFING COUNTERMEASURES

This section describes the extension of the evaluation methodology to assess spoofing countermeasures. Figure 8(a) illustrates the deployment of spoofing countermeasures as independent subsystems. Just like the biometric system, the countermeasure is a two-class classifier, its role being to distinguish between genuine and spoofed trials. Again, just like the biometric system, it will also make mistakes, leading to two new error rates referred to as the false living rate (FLR) and the false fake rate (FFR). The FLR reflects the percentage of spoofed trials misclassified as genuine trials, whereas the FFR reflects the percentage of genuine trials misclassified as spoofed trials.

Even as independent subsystems, spoofing countermeasures impact directly on the performance of the biometric system as a whole; while aiming to reduce vulnerabilities to spoofing, they also have potential to reject genuine trials [7]. Thus, while countermeasure subsystems can be assessed independently, it is often of greater interest to assess the performance of the system as a whole, for example, using a score fusion strategy illustrated in Figure 8(b).

Countermeasure subsystems may alternatively be integrated, not in parallel but in series, with the biometric system as illustrated in Figure 8(c). Assessment involves four different system configurations, for each of which a different DET profile is produced. Examples are illustrated in Figure 9. The first

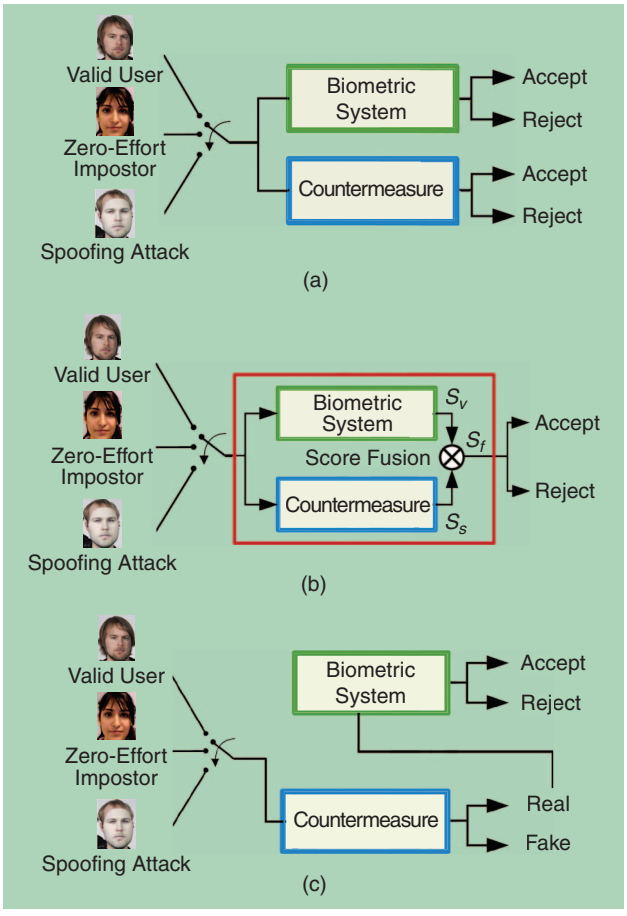
configuration (blue profile) illustrates the performance of the baseline biometric system (no spoofing, no countermeasures). The second configuration (black profile) illustrates the performance of the same system when subjected to spoofing attacks. The third configuration (green profile) illustrates the improvement in performance with active countermeasures. While seemingly complete, a fourth configuration (red profile) is still needed to determine the performance of the integrated biometric and countermeasure systems in the absence of spoofing. This final configuration is essential to gauge the effect of the countermeasure on genuine trials that may be misclassified as spoofing attacks. To reiterate, all four configurations are needed to properly observe the complex impact of spoofing and countermeasures on integrated systems.



[FIG7] **A DET profile illustrating the SFAR. FRR@EER refers to the FRR on the test set for a decision threshold τ for which the FAR and FRR are equal (determined from the development set).**

Even if it is the effect on overall recognition performance that is of greatest interest, there will always be an interest to first assess countermeasure performance independently. This configuration might be limited to system development. The second, fused approach is simple and straightforward. However, unless separate decisions are applied to the countermeasure and recognition subsystem scores, it does not support the explicit detection of spoofed trials. The third approach is therefore the most appealing in practice, allowing for explicit spoof detection and an evaluation of countermeasure impacts on overall system performance.

Finally, whatever the approach to integration, performance is dependent on a separate countermeasure decision threshold τ^{CM}. Since it impacts upon overall performance, the FAR/FRR/SFAR can be determined for a range of different operating points specified, for example, by an application-dependent FFR. So as to reflect the likely performance for a variety of different application scenarios, the TABULA RASA project considered values of FFR $= \{1, 5, 10\}\%$. Even so, countermeasure integration and assessment is not a solved problem and very much a topic of ongoing research. This is discussed further in the section "Lessons Learned."



[FIG8] **Integrated biometric systems and spoofing countermeasures. (a) Two independent components. (b) Fused subsystems. (c) A two-step process.**

EVALUATION METHODOLOGY: A CASE STUDY IN 2-D FACE VERIFICATION

As an illustration of the evaluation methodology in practice, the following presents a case study in two-dimensional (2-D) face verification.

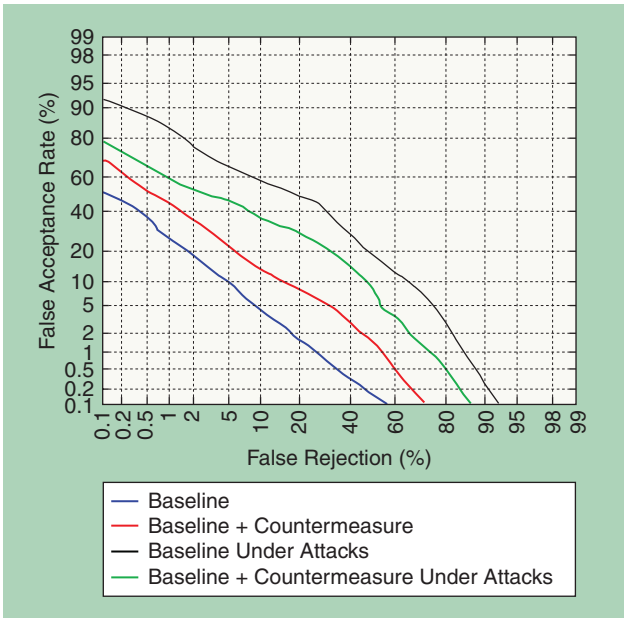
PREVIOUS WORK

The vast majority of past research in 2-D face recognition has focused on maximizing the discrimination between the faces of different persons [8]. Liveness detection has received considerably less attention, even though 2-D face-recognition systems have been known to be vulnerable to spoofing for some time.

Many face-recognition systems can be spoofed by presenting to the camera photographs, videos, or 3-D masks of enrolled persons [2]. While makeup or plastic surgery are also viable spoofing attacks, photographs are among the most easily implemented, most effective, and therefore the most likely in practice. Video attacks can be especially effective since they are dynamic signals that more closely resemble a genuine trial. Furthermore, it is natural to assume that systems that are vulnerable to photo attacks will also be vulnerable to video attacks.

SPOOFING IS ACCOMPLISHED USING FAKE BIOMETRIC SAMPLES EXPRESSLY SYNTHESIZED OR MANIPULATED TO PROVOKE ARTIFICIALLY HIGH COMPARATOR SCORES.

As is the case for the spectrum of other modalities, the typical countermeasure to prevent the spoofing of 2-D face-recognition systems involves liveness detection. Here, liveness indicators include eye blinking, changes in facial expression, mouth movements, estimates of skin texture, structure and motion analysis, and depth information, etc. Multispectral and reflectance analysis has also been used successfully to differentiate between living faces and lifeless fakes [2]. Alternatively, face recognition can be combined with other biometric modalities such voice or gait recognition. A survey of face-spoofing detection approaches can be found in [4].



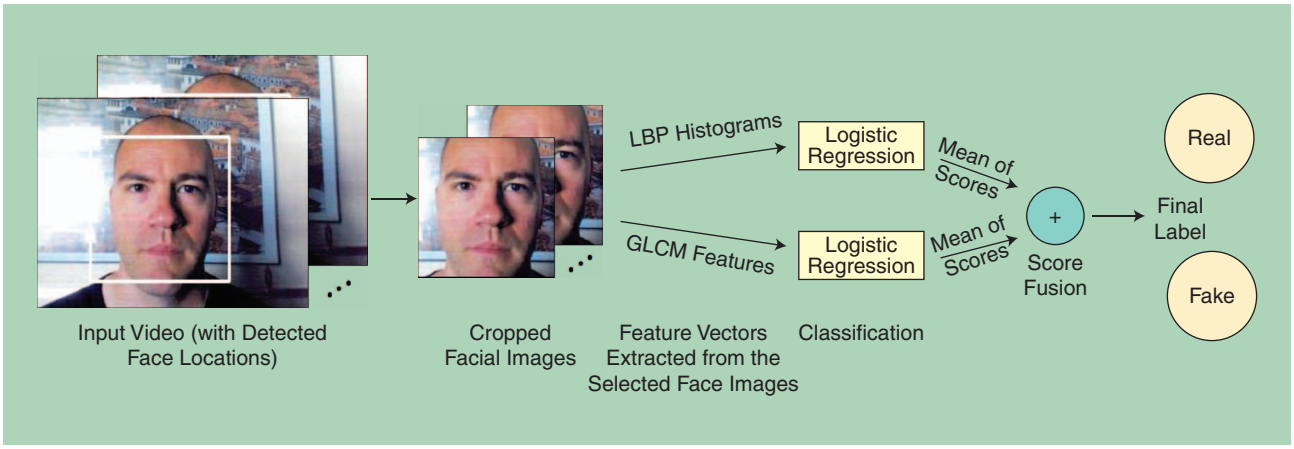
[FIG9] The FAR and FRR of a biometric system with and without spoofing attacks, and with and without countermeasures.

BASELINE FACE BIOMETRIC SYSTEM

The system combines a part-based face representation with Gaussian mixture models (GMMs) [9]. The system divides the face into blocks and treats each block as a separate observation of the same underlying signal (the face). A feature vector is thus obtained from each block by applying the discrete cosine transform (DCT). The distribution of the feature vectors is then modeled using GMMs.

For feature extraction, the face is normalized, registered, and cropped. The cropped and normalized face is divided into blocks (parts) from each of which a feature vector is extracted. Each feature vector is treated as a separate observation of the same underlying signal and the distribution of the feature vectors is modeled using GMMs. The feature vectors from each block are obtained by applying the DCT. Once the feature vectors are calculated, feature distribution modeling is achieved by performing background model adaptation of GMMs. Background model adaptation first involves the training of a world (background) model Ω_{world} from a large data set of faces. Client models $\Omega^{i}_{\text{client}}$ for client i are learned through the adaptation of the world model toward the observations of the client.

Verification is achieved by scoring an observation x against both the client ($\Omega^{i}_{\text{client}}$) and world ($\Omega_{\text{model}}$) models. They produce log-likelihood scores, which are combined to give the single log-likelihood ratio (LLR). The LLR is used to assign the observation



[FIG10] A block diagram of the texture-based countermeasure.

[FIG11] The setup and sample images from the REPLAY-ATTACK database. Spoofed data collection using (a) printed photographs and (b) examples of genuine accesses and spoofing attacks. In (b), the columns from left to right show examples of real accesses, printed photographs, mobile phone, and tablet attacks.

to the world class (not the client) or the client class based on a predefined threshold τ.

SPOOFING COUNTERMEASURE

Genuine and spoofed face images exhibit differences at the texture level. Spoofed face images are captured or acquired twice, first from the live client and second from the spoofing medium at recognition time. Accordingly, spoofed face images tend to exhibit degraded facial texture due to reproduction or printing on the spoofing medium (a photograph) and the two imaging systems [10]. As a result, estimates of facial texture quality, here obtained using logistic regression [11], serve to identify spoofing attacks. The method is computationally efficient and does not require any additional user cooperation; it is nonintrusive.

AS IS THE CASE FOR THE SPECTRUM OF OTHER MODALITIES, THE TYPICAL COUNTERMEASURE TO PREVENT THE SPOOFING OF 2-D FACE-RECOGNITION SYSTEMS INVOLVES LIVENESS DETECTION.

Figure 10 illustrates a block diagram of the texture-based spoofing countermeasure. The approach analyzes the texture of single facial images using two different texture descriptors: local binary patterns (LBPs), which encodes the microtexture patterns into a feature histogram and features computed from the gray-level co-occurrence matrices (GLCMs), which describe the distribution of different combinations of gray-level pairs in an image block. The combination of these two measurements provides an effective representation of the overall facial texture quality. The facial texture descriptions are fed into logistic regression classifiers. Score-level fusion of the individual classifier outputs determine whether the facial image corresponds to a living person or a spoofed reproduction. As the database consists of video sequences, several cropped face images from each video are used for feature extraction at intervals of 0.6 seconds. The final score for the two individual texture representations is determined by averaging the scores of each face image. The outputs of each classifier are then fused using the sum rule with min-max score normalization [6] to obtain the final label for each video sequence.

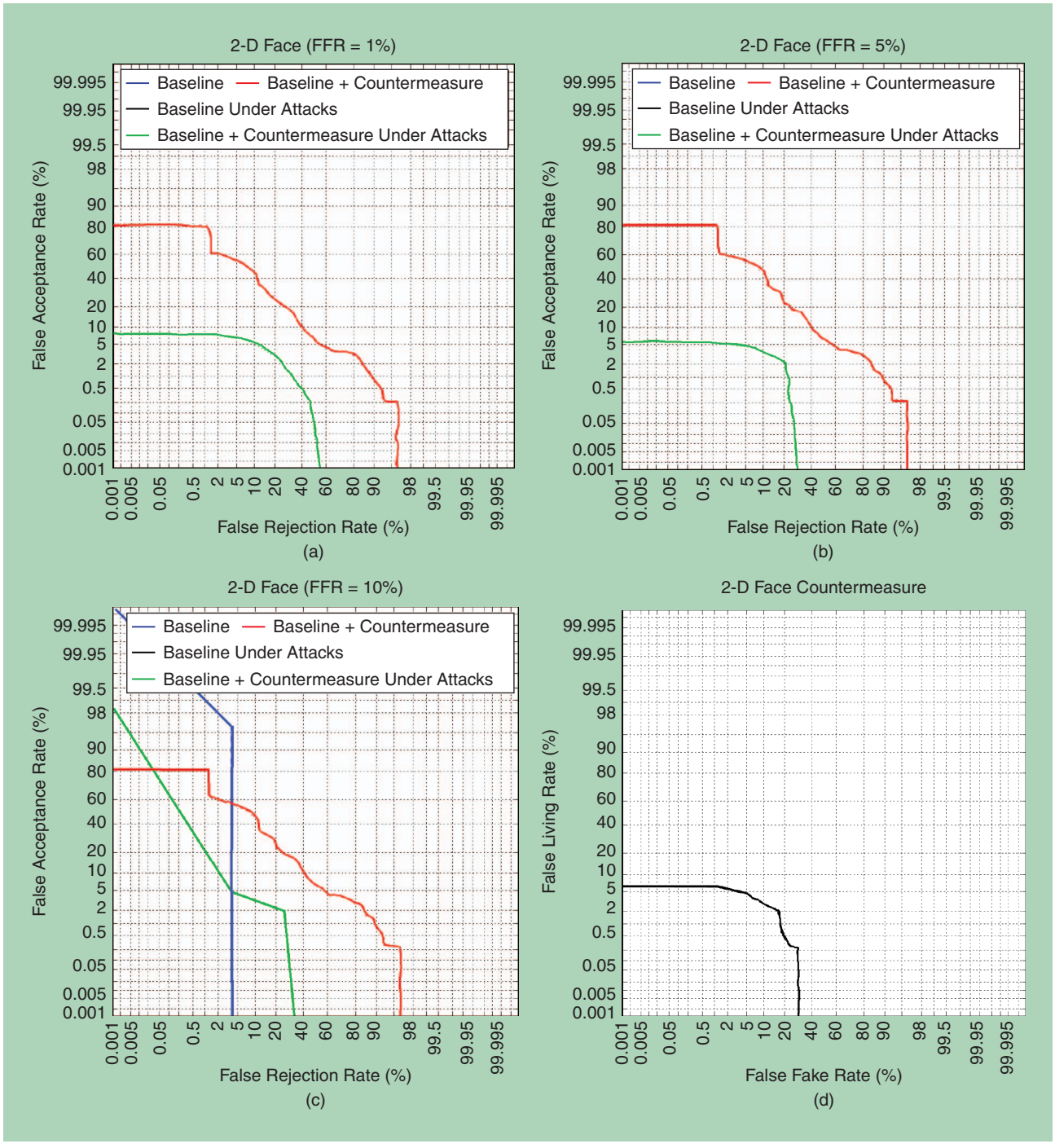
THE FACE SPOOFING ATTACK DATABASE

Experiments were performed with the REPLAY-ATTACK face spoofing database [12]. It consists of 1,300 samples comprising genuine trials, photo, and video spoofing attacks. The database contains samples collected from 50 persons under varying lighting conditions. The data is split into four subgroups comprising enrollment, training, development, and test data. While the enrollment set includes samples collected from all clients, there is no client overlap in the training, development, and test data sets.

All videos are generated by either having a (real) client trying to access a laptop through a built-in webcam or by displaying a photo or a video recording of the same client for at least nine seconds. In total, 20 attack videos were recorded for each client, whereas six videos were captured for real accesses. The enrollment set contains 100 videos (two per client) and is used exclusively to evaluate baseline performance. The training set contains 60 real accesses and 300 attacks. The development set contains 60 real accesses and 300 attacks, whereas the test set contains 80 real accesses and 400 attacks. Examples of real accesses and spoofing attacks from the REPLAY-ATTACK database are shown in Figure 11. A complete description of the database and associated protocol can be found in [12].

EXPERIMENTAL RESULTS

Vulnerabilities to spoofing and countermeasure impacts were assessed according to the evaluation methodology introduced in this article. Results in the form of DET profiles are illustrated in Figure 12: (a)–(c) represent overall system performance under a spoofing attack for three different operating points where countermeasures are tuned to produce FFRs of 1%, 5%, and 10%, respectively. For completeness, (d) is included to illustrate countermeasure performance in independence from biometric recognition. The EER of the countermeasures is in the order of 5%.



[FIG12] DET profiles for the 2-D face verification systems with/without spoofing and with/without countermeasures at different countermeasures operating points [FRRs of (a) 1%, (b) 5%, and (c) 10%]. Missing profiles indicate EER = FRR = FAR = 0%. Also illustrated in (d) is a DET profile for the independent countermeasure.

The texture-based countermeasure delivers significant improvements in robustness to spoofing at almost every operating point. For example, when the countermeasure is tuned to an FFR of between 1 and 5% as illustrated in Figure 12(a) and (b), and for a fixed FRR of 1%, then the FAR under a spoofing attack drops from 80% to just over 5%. However, if the countermeasure is tuned to a higher FFR such as 10% as illustrated in Figure 12(c), then the FAR increases to approximately 20%. This is high-security configuration. Even if almost no spoofing trials are misclassified as genuine trials, the countermeasure misclassifies 10% of genuine trials as spoofed trials.

In this case, the countermeasure degrades usability; this effect is illustrated by the blue profile in Figure 12(c). There is an inherent tradeoff between the security and usability, which is heavily

dependent on the modality, recognition system, countermeasure, and database. For this particular configuration, a sensible compromise is achieved when the countermeasure FFR is tuned to 5%. Here, 5% of genuine trials will be misclassified as spoofed trials, whereas for all but the highest FRRs, only 5% of spoofing attacks will succeed. Between 40 and 80% of attacks would otherwise be successful without the texture-based countermeasure.

WHILE POTENTIALLY MORE INTRUSIVE, CHALLENGE-RESPONSE COUNTERMEASURES CAN BE COMPLEMENTARY TO THE MORE TRADITIONAL APPROACHES TO LIVENESS DETECTION.

LESSONS LEARNED

The evaluation methodology presented in this article is based upon work funded through the European TABULA RASA project. Through a collaboration involving a large team of researchers, the study included a wide range of biometric modalities including face (2-D, 3-D, and multispectral), voice, gait, fingerprint, iris, vein, and electrophysiological signals. The study established state-of-the-art authentication technologies for each modality, investigated the vulnerabilities of each to spoofing, and proposed novel countermeasures integrated and evaluated according to the methodology outlined in the sections "Evaluation Methodology: Vulnerabilities to Spoofing" and "Evaluation Methodology: Spoofing Countermeasures." The findings described in the face-recognition case study are illustrative of the general trend across all modalities considered in TABULA RASA. The following discusses some of the lessons learned and the most pressing directions for future research.

VULNERABILITIES

Unless they are equipped with suitable countermeasures, all biometric systems were shown to be vulnerable to spoofing. Even so, some modalities (e.g., gait) are more robust than others (e.g., fingerprint), however, this should not be interpreted as meaning they are more reliable; in the absence of spoofing, fingerprint recognition generally outperforms gait recognition. Multimodal biometric systems are also vulnerable and can be overcome by the spoofing of only a single modality [13].

METRICS AND PROTOCOLS

Spoofing and countermeasure assessment is considerably more complex than might first appear. Countermeasures tend to be developed and evaluated independently from recognition systems and, although this methodology supports the comparison of different antispoofing approaches, it is the influence of countermeasures on overall system performance that is of greatest interest. New, standard metrics and protocols reflecting the robustness of integrated biometric and countermeasure systems should be adopted in the future.

USABILITY

While countermeasures are successful in reducing the vulnerability of biometric systems to spoofing, increased robustness comes at the expense of increased FRR. This impact on usability is generally manageable; increases in the FRR are usually negligible in comparison to increases in the FAR caused by spoofing. It is stressed, however, that the tradeoff is dependent on the modality and application.

GENERALIZATION

A large number of antispoofing studies have been reported in the literature, and encouraging results have been detailed for a small number of standard databases, e.g., [14]–[16]. Spoofing attacks are, however, varying and unpredictable in nature and it is difficult to predict how countermeasures will generalize to spoofing attacks "in the wild," where their true nature can never be known with certainty. As the field evolves, and as new forms of spoofing emerge, it will be necessary to collect databases with increasingly challenging and diverse spoofing trials, calling for new and increasingly generalized countermeasures. An alternative or complementary strategy involves one-class classification approaches [17]. As a form of anomaly detection that relies only on models of genuine data, they offer greater potential for generalization.

COUNTERMEASURE FUSION

The growing interest in spoofing and countermeasures and the number of diverse countermeasure strategies and algorithms in the literature lends support to research in fused countermeasures. While not increasing robustness to specific attacks, fused countermeasures offer a flexible antispoofing framework with which newly emerging vulnerabilities can be quickly patched with adaptable fusion strategies or the addition of new countermeasures.

CHALLENGE-RESPONSE COUNTERMEASURES

Interactive, challenge-response countermeasures require a user to perform a randomized, specific action such as smiling or blinking. The correct response to the given challenge infers liveness. While potentially more intrusive, challenge-response countermeasures can be complementary to the more traditional approaches to liveness detection. Further work is required to validate their potential.

COMBINED HARDWARE- AND SOFTWARE-BASED COUNTERMEASURES

The majority of previous work, including all that within TABULA RASA, has investigated software-based countermeasures. Hardware-based countermeasures and new sensors have great potential to detect more elaborate spoofing attacks, and a new generation of countermeasures that combine hardware- and software-based approaches may be needed.

OUTLOOK

Until relatively recently, the main focus of biometrics research has centered on the discrimination between genuine and zero-effort impostor trials. In the face of well-acknowledged vulnerabilities, greater effort to develop countermeasures will be required in the future to defend against concerted-effort spoofing attacks. The

growing body of related literature and number of competitive evaluations is testament to the increasing interest and importance of this research.

The past work is characterized by the lack of a standard evaluation methodology, which is needed to assess the influence of countermeasures on biometric system performance. This article presents what is, to the best of our knowledge, the first proposal for a formal evaluation methodology. Needless to say, however, further work is required to extend and adapt the methodology in view of the lessons learned through recent work.

While the performance of spoofing countermeasures proposed so far gives cause for optimism, their generalization to previously unseen spoofing attacks remains unclear. This leads to a number of research directions for the future, including networks of independent, fused countermeasures and one-class classification strategies. As the field evolves, new and more challenging databases will be essential for biometric system developers to stay one step ahead of the fraudsters. While extremely difficult, it will be critical to estimate the reliability of countermeasures in practical application scenarios including, not just their ability to detect spoofing, but also their impact on system usability.

ACKNOWLEDGMENTS

The following support is gratefully acknowledged: the European Commission (TABULA RASA: grant agreement number 257289, BEAT: grant agreement number 284989); the Academy of Finland; and the Spanish MICINN (BIO-SHIELD: grant agreement number TEC2012–34881).

AUTHORS

Abdenour Hadid (hadid@ee.oulu.fi) received the doctor of science in technology degree in electrical and information engineering from the University of Oulu, Finland, in 2005. He has been an adjunct professor and academy research fellow in the Center for Machine Vision Research of the University of Oulu since 2010. His research focuses on computer vision and pattern recognition with a particular interest in face analysis and biometrics. He made significant contributions to the state of the art, and his work is gaining increasing interest in the scientific community. According to Google Scholar (as of May 2015), his h-index is 24 and his work has been cited more than 5,600 times.

Nicholas Evans (evans@eurecom.fr) is an assistant professor at EURECOM, in Sophia Antipolis, France. He received the M.Eng. (1999) and Ph.D. (2003) degrees from the University of Wales, Swansea, United Kingdom. He was the lead guest editor of *IEEE Transactions on Information Forensics and Security* (special issue on biometric spoofing and countermeasures) as well as this special issue of *IEEE Signal Processing Magazine* on biometric security and privacy. He is an associate editor of *EURASIP Journal on Audio, Speech, and Music Processing* and is a member of the IEEE Speech and Language Technical Committee.

Sébastien Marcel (marcel@idiap.ch) is a researcher at the Idiap Research Institute, Switzerland, where he heads the biometrics group. He is also a lecturer at the Ecole Polytechnique Fédérale de Lausanne. He is an associate editor of *IEEE Transactions on Information Forensics and Security* (*T-IFS*), coeditor of the *Handbook of Biometric Anti-Spoofing*, and was a guest editor of *T-IFS* (special issue on biometric spoofing and countermeasures) and of this special issue of *IEEE Signal Processing Magazine* on biometric security and privacy.

Julian Fierrez (julian.fierrez@uam.es) received the M.Sc. and Ph.D. degrees from Universidad Politécnica de Madrid, Spain, in 2001 and 2006, respectively. Since 2004 he has been affiliated with Universidad Autonoma de Madrid, where he is currently an associate professor. From 2007 to 2009, he was a Marie Curie postdoctoral researcher at Michigan State University. His research interests include image processing, pattern recognition, authentication using biometrics such as fingerprints and handwritten signatures, and security of person authentication systems. He is actively involved in multiple EU projects and has received multiple research distinctions including the EURASIP Best Ph.D. Award in 2012.

REFERENCES

[1] A. Jain, A. Ross, and S. Pankati, "Biometrics: A tool for information security," *IEEE Trans. Inform. Forensics Security*, vol. 1, no. 2, pp. 125–143, June 2006.

[2] S. Marcel, M. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing*. New York: Springer, 2014.

[3] E. Marasco and A. Ross, "A survey on antispoofing schemes for fingerprint recognition systems," *ACM Comput. Surv.*, vol. 47, no. 2, pp. 28:1–28:36, Nov. 2014.

[4] J. Galbally, S. Marcel, and J. Fierrez, "Biometric anti-spoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1–23, Dec. 2014.

[5] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and countermeasures for speaker verification: A survey," *Speech Commun.*, vol. 66, no. 0, pp. 130–153, 2015.

[6] J. Fierrez, "Adapted fusion schemes for multimodal biometric authentication," Ph.D. dissertation, Universidad Politecnica de Madrid, May 2006.

[7] I. Chingovska, A. Anjos, and S. Marcel, "Biometrics evaluation under spoofing attacks," *IEEE Trans. Inform. Forensics Security*, vol. 9, no. 12, pp. 2264–2276, Dec. 2014.

[8] S. Z. Li and A. K. Jain, Eds., *Handbook of Face Recognition*, 2nd ed. New York: Springer, 2011.

[9] L. El-Shafey, C. McCool, R. Wallace, and S. Marcel, "A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 35, no. 7, pp. 1788–1894, July 2013.

[10] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint and face recognition," *IEEE Trans. Image Processing*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[11] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using texture and local shape analysis," *IET Biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[12] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. D. Martino, A. Hadid, M. Pietikäinen, and S. Marcel, "Face liveness detection using dynamic texture," *EURASIP J. Image Video Process.*, vol. 2014, no. 2, Jan. 2014.

[13] G. L. Marcialis, G. Fumera, and B. Biggio, "Anti-spoofing: Multimodal," *Encyclopedia of Biometrics*. New York: Springer, 2014.

[14] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G. L. Marcialis, F. Roli, and S. Schuckers, "Livdet 2013—Fingerprint liveness detection competition," in *Proc. IEEE/IAPR Int. Conf. Biometrics*. IEEE Press, June 2013, pp. 1–6.

[15] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kahm, C. Glaser, N. Damer, et al., "The 2nd competition on counter measures to 2D face spoofing attacks," in *Proc. IEEE/IAPR Int. Conf. Biometrics*. IEEE Press, June 2013, pp. 1–6.

[16] Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilci, M. Sahidullah, and A. Sizov, "ASVspoof 2015: The first automatic speaker verification spoofing and countermeasures challenge," in *Proc. INTERSPEECH*, 2015.

[17] F. Alegre, A. Amehraye, and N. Evans, "A one-class classification approach to generalised speaker verification spoofing countermeasures using local binary patterns," in *Proc. IEEE 6th Int. Conf. Biometrics: Theory, Applications and Systems (BTAS 2013)*, 2013, pp. 1–8.

[SP]