Face Biometrics under Spoofing Attacks: Vulnerabilities, Countermeasures, Open Issues and Research Directions

Abdenour Hadid
Center for Machine Vision Research (CMV), University of Oulu, Finland

Abstract—Among tangible threats and vulnerabilities facing current biometric systems are spoofing attacks. A spoofing attack occurs when a person tries to masquerade as someone else by falsifying data and thereby gaining illegitimate access and advantages. Recently, an increasing attention has been given to this research problem. This can be attested by the growing number of articles and the various competitions that appear in major biometric forums. We have recently participated in a large consortium (TABULARASA) dealing with the vulnerabilities of existing biometric systems to spoofing attacks with the aim of assessing the impact of spoofing attacks, proposing new countermeasures, setting standards/protocols, and recording databases for the analysis of spoofing attacks to a wide range of biometrics including face, voice, gait, fingerprints, retina, iris, vein, electro-physiological signals (EEG and ECG). The goal of this position paper is to share the lessons learned about spoofing and anti-spoofing in face biometrics, and to highlight open issues and future directions.

I. INTRODUCTION

Identity management using biometrics has nowadays become a reality mainly because of the biometric passports (e-passports) and also because of the presence of more and more biometric enabled-applications for personal computers. However, despite the significant progress in the recent decades [11], biometric systems are, unfortunately, vulnerable to attacks. A spoofing attack occurs when a person tries to masquerade as someone else by falsifying data and thereby gaining illegitimate access and advantages. This is currently a major problem for companies willing to market information security solutions based on biometric authentication technologies. For instance, some laptops of Lenovo, Asus and Toshiba come with built-in webcams and embedded biometric systems that authenticate users by scanning their faces. However, in 2009, the Security and Vulnerability Research Team of the University of Hanoi (Vietnam) demonstrated at Black Hat 2009 conference, the world’s premier technical security conference, how to easily spoof and bypass these systems (Lenovo’s Veriface III, Asus’ SmartLogon V1.0.0005, and Toshiba’s Face Recognition 2.0.2.32 - each set to its highest security level) using fake facial images of the legitimate user, thus gaining access to the laptops. This vulnerability is now listed in the National Vulnerability Database of the National Institute of Standards and Technology (NIST) in the US. More recently (September 2013), Apple Inc. released its new device, iPhone 5s - with a Touch ID fingerprint sensor for log-in and making users’ data more secure. Less than two days later, a German hacker collective, Chaos Computer Club, claimed and then demonstrated the spoofing of the iPhone 5S with a gummy finger. These two examples, among several others, highlight tangible threats and vulnerabilities in current biometric-based information security. Thus, there is an urgent need for efficient and reliable solutions for detecting and circumventing spoofing attacks. The typical countermeasure to a spoofing attack is liveness detection that aims at detecting some physiological signs of life. It is also assumed that multi-modal systems (e.g. combining face and voice biometric modalities) are in principle more difficult to spoof than uni-modal systems. Thus, gait, face and iris verification could also be performed jointly. However, preliminary investigations indicate that spoofing only one modality can actually be enough to weaken the fusion rule and crack a biometric system protected by multiple modalities.

We have recently participated in a large consortium (TABULARASA EU project, 2010-2014), dealing with the vulnerabilities of existing biometric systems to spoofing attacks with the aim of assessing the impact of spoofing attacks, proposing new countermeasures, setting standards/protocols, and recording databases for the analysis of spoofing attacks to a wide range of biometrics including face, voice, gait, fingerprints, retina, iris, vein, electro-physiological signals (EEG and ECG). The goal of this position paper is to discuss the lessons learned about spoofing and anti-spoofing in face biometrics, and to highlight open issues and future directions.

The rest of this paper is organized as follows. In the next section, Section II, we present an experimental analysis demonstrating the vulnerability of face biometrics to spoofing attacks. Existing databases for studying face anti-spoofing are then described in Section III whereas Section IV reviews some proposed methods in the literature to cope with face spoofing attacks. Some open issues and future directions are discussed in Section V. Finally, a concluding remarks are drawn in Section VI.

II. VULNERABILITY OF FACE BIOMETRICS TO SPOOFING ATTACKS

To gain insight into the vulnerabilities of face biometric systems when confronted to spoofing attacks, we experimentally analyzed the performance of a baseline system, not trained to handle spoofing attacks, on a challenging 2D face spoofing database known as the REPLAY-ATTACK database [8]. The
chosen baseline face verification system is developed by
IDIAP research institute (Switzerland) and uses a part-based
face representation and Gaussian mixture models (GMMs)
[19]. We briefly describe below the baseline face verification
system, the spoofing attack database, the experimental setup
and importantly the obtained results which clearly assess
the significant vulnerabilities of face biometrics to spoofing
attacks.

A. The Baseline Face Biometric System

The face verification system, proposed by McCool and
Marcel in [19], is chosen as the baseline system for face
authentication. The system combines a part-based face represen-
tation and Gaussian mixture models (GMMs). 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. This cropped and normalized face is divided into
blocks (parts) and from each block (part) a feature vector is
obtained. Each feature vector is treated as a separate obser-
vation of the same underlying signal (in this case the face)
and the distribution of the feature vectors is modeled using
GMMs. The feature vectors from each block are obtained by
applying the DCT [25].

Once the feature vectors are calculated, feature distribu-
tion modelling is achieved by performing background model
adaptation of GMMs [6], [18]. Background model adaptation
first involves the training a world (background) model Ω_{world}
from a set of faces and then the derivation of client models
Ω^i_{client} for client i by adapting the world model to match the
observations of the client. The adaptation is performed using
a technique called mean only adaptation [26].

To verify an observation, x, it is scored against both the
client (Ω^i_{client}) and world (Ω_{model}) model. The two models,
Ω^i_{client} and Ω_{world}, produce a log-likelihood score which is
then combined using the log-likelihood ratio (LLR) to produce
a single score. This score is used to assign the observation to
the world class of faces (not the client) or the client class of
faces (it is the client) based on a predefined threshold τ.

B. The Face Spoofing Attack Database

To analyze the performance of the face baseline system
under spoofing attacks, we considered the REPLAY-ATTACK
face spoofing database [8] which consists of 1300 video clips
comprising of real-accesses or photo and video attack attempts
to different 50 identities, under different lighting conditions.
The data is split into 4 sub-groups comprising enrollment,
training, development and test data. Clients that appear in one
of the last three data sets do not appear in any other set, while
the enrollment set includes all clients.

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 9 seconds. In total, 20 attack videos were recorded
for each client and 6 videos were captured for real accesses
yielding in:

- **Enrollment set**: containing 100 videos (2 per client) for
  exclusively studying the baseline performance of face
  recognition systems;
- **Training set**: containing 60 real-accesses and 300 at-
tacks;
- **Development set**: containing 60 real-accesses and 300
  attacks;
- **Test set**: containing 80 real-accesses and 400 attacks.

Examples of real accesses and attacks from the REPLAY-
ATTACK database are shown in Figure 1. The full description
of the database and its associated protocol can be found in
[8].

![Fig. 1. Examples of real accesses and attacks. In the top row, samples from controlled scenario. In the bottom row, samples from adverse scenario. Columns from left to right show examples of real access, printed photograph, mobile phone and tablet attacks.](image)

C. Experimental Setup

We started by analyzing how the baseline face system
behaves on the database’s licit protocol setup (i.e. only using
real-access attempts). In this mode, the system is tested for
how well it can recognize real users authenticating against
their templates and how well it can reject real users au-
thenticating against other users’ templates (i.e. imposters).
The performance is measured objectively by observing the
rate of users rejected when authenticating against their own
template (False Rejection Rate) and by the rate of users
accepted when authenticating against someone else’s template
(False Acceptance Rate). In this way, we establish the baseline
recognition performance of the baseline face recognition for
licit access (enrollment attempts and authentication tries).

To determine the baseline face recognition performance, we
computed scores exhaustively for all videos from the develop-
ment and test set real-accesses, probing for every identity in
the set against all other models in the same set, without inter-
mixing across development and test sets. The scores generated
from matched client videos and models within the subset
are considered true client accesses and contribute to the licit
Acceptance Rate, while all others, impostors, contributing to
the licit Rejection Rate. By varying the classification threshold
on the test set, we obtain the recognition performance for the
system. In this context, the FAR (False Acceptance Rate) is
considered as the rate of impostors that are wrongly classified by the system as true-claimants. The FRR (False Rejection Rate) is the rate of true claimants that the system falsely classified as impostors.

To determine the robustness of the baseline system when exposed to spoofing attacks, we keep the models as trained during the licit protocol performance assessment and try attacks to the models with matching identity. A successful attack is accomplished when the system confuses a spoofing attempt with the corresponding matched user template. In this mode, the FAR corresponds to the rate of attacks that are accepted by the system when spoofed. The FRR corresponds to the rate of real-access attempts that are incorrectly dismissed by the system as attacks.

**D. Experimental Results**

The results of the experiments are presented in Figure 2 in terms of detection error trade-off (DET) profiles which illustrate the dynamic behavior of a biometric system as the decision threshold is changed, i.e. how the false acceptance rate varies according to the false rejection rate. We also show in Figure 3 the score distributions of true claimants, impostors, and spoofing attacks. By comparing these three distributions, we can observe how spoofed data is closer to information from true claimants than non-spoofed data from an average impostor trying to access the system illegally.

From these results, it is possible to notice that the face baseline system achieves perfect performance when not confronted to spoofing attacks. The performance sharply degrades in the presence of spoofing attacks. These results exemplify the vulnerability of face biometrics against spoofing attacks. These findings are used as motivations for developing countermeasures.

**III. EXISTING SPOOFING ATTACK DATABASES**

NUAA Photo Imposter Database is among the first public datasets for studying anti-spoofing in face recognition. It was released in 2010, accompanying the work of Tan et al. in [28] in which the authors explored the Lambertian reflectance model to encode the differences between the 2D images of the face presented during an attack and a real (3D) face shown.
in real-access attempts. Following the trend of similar past work [16], [4], the authors focused on the binary classification task of face spoofing detection considering pictures of real-accesses and attacks recorded with a conventional webcam. NUAA Photo Imposter Database is publicly available and is mainly useful for studying texture-based approaches to spoofing detection.

As shown by Anjos et al. [2], [1], techniques for anti-spoofing can also exploit motion artifacts present in attacks to discriminate spoofing attempts. In [2], the authors made available a public dataset composed of printed photograph attacks and real-accesses, in which the samples available for the training and evaluating spoofing classifiers are videos. The PRINT-ATTACK database can be used to devise anti-spoofing methods based on texture, motion or both [7]. An extension of this database, called the PHOTO-ATTACK database, providing photo attacks using different attack media such as mobile phones and tablets was introduced in [1]. Another extension called REPLAY-ATTACK database, also bringing video attacks using mobile phones and tablets was introduced in [8] and used in the experiments in the previous section.

Zhang et al. have also recorded and released a public dataset for face anti-spoofing containing challenging short video sequence of attacks to 50 different identities using printed photographs and videos displayed through a tablet screen [31]. The photo attacks in this database include warping. The face video attacks in this database can be used for evaluating countermeasures based on motion, texture or both.

Very recently, Erdogmus and Marcel [10] made publicly available a 3D mask database, called 3DMAD, composed of real access and mask attack videos of 17 different subjects recorded by Microsoft Kinect sensor. This database is mainly used for evaluating anti-spoofing measures on 2D face recognition.

IV. FACE ANTI-SPOOFING METHODS

We discuss in this section some existing works in the literature on face anti-spoofing. Short surveys of some schemes against photograph spoofing attacks can be found in [23], [20].

The typical countermeasure to spoofing attacks is liveness detection that aims at detecting physiological signs of life (such as eye blinking, facial expression changes and mouth movements). For instance, Pan et al. [23] proposed an eyeblink-based anti-spoofing method by integrating a structured prediction method whereas Kollreider et al. [12] presented an optical-flow based method to capture the subtle motion of face images. While such countermeasures may work in cases of attacks using photographs, they are generally ineffective when using a video (or simply shaking the photograph before the camera) as a mean of spoofing. Some researchers attempted to counter video spoofing by using structure from motion to calculate the depth information. Again, this may not work in case of spoofing attacks using 3D masks, for instance. Some current face anti-spoofing methods are based on the analysis of the skin properties such as the analysis of skin texture and skin reflectance [23], [20].

In [17], Li et al. described a method for print-attack detection by exploiting differences in the 2D Fourier spectra comparing the hard-copies of client faces and real-accesses. In that work, the authors derive the probability of attack by applying a high-pass filter to the spectra of the sample being analyzed and computing a score which is then classified according to some heuristic. The method works well for down-sampled photos of the attacked identity, but is likely to fail for higher-quality samples. The used dataset is not publicly available.

In [3], the authors proposed a method to detect spoofing attacks using printed photos by analyzing the micro-textures present in the material using a linear SVM classifier to achieve a 2.2% False-Acceptance Rate (FAR) against a 13% False-Rejection Rate (FRR). A major limitation of this method is that the input image needs to be reasonably sharp.

In contrast to the works cited above, the authors in [13], [15] presented a technique to evaluate liveness based on a short sequence of images. The work describes a binary detector that evaluates the trajectories of select parts of the face presented to the input sensor using a simplified optical flow analysis followed by an heuristic classifier. Such a classification scheme achieves an equal-error rate of 0.5% for samples of real-accesses extracted from XM2VTS and attacks produced using hard-copies of those data. The same authors also introduced in [14] a method for fusing scores from different expert systems that observe, concurrently, the 3D face motion scheme introduced on the previous work and liveness properties such as eye-blinks or mouth movements.

The works in [21] and [24] bring a real-time liveness detection specifically against photo-spoofing using (spontaneous) eye-blinks which are supposed to occur once every 2-4 seconds in humans. The system developed uses an undirected conditional random field framework to model the eye-blinking that relaxes the independence assumption of generative modelling and state dependence limitations from hidden Markov modelling. The system is tested on a dataset provided by the authors and was made publicly available. Such a dataset is composed of short video clips of eye-blinks and spoofing attempts using photographs. The attacks are not solely composed of still images but also arbitrary shaking behavior which increases the task difficulty. With this setup, the proposed detector is able to achieve 95.7% true-positive classification against a false alarm of less than 0.1% when considering a simultaneous blink of both eye lids in all test samples. A later work by the same authors [22] augment the number of countermeasures deployed to include a scene context matching that helps preventing video-spoofing in stationary face-recognition systems. To achieve this, the eye-blink detector output scores are fused with the output of a simple local-binary-pattern-\(\chi^2\) detector. The scene context detector uses some carefully chosen fiducial points coming from near regions outside the face boundaries that characterize the expected scene context. To test this new setup, the authors constructed a new private dataset with which they obtained an almost perfect scoring - 99.5% true-rejection against 100% true-acceptance.
In [5], Bao et al. proposed a method to detect attacks produced with planar media (such as paper or screens) using motion estimation by optical flow. Movement of planar objects is categorized as translation, rotation, normal or swing and 8 quantities that express the amount of these movements extracted from the analyzed (already) cropped face. The probability of an attack is then computed taking the 8 values and applying them to an ad-hoc equation that outputs a single score indicating the probability of a 3D face given the input data. Experiments on a private dataset showed a 6% false-alarm against and 14% false-acceptance in best case.

V. OPEN ISSUES AND FUTURE DIRECTIONS

A. Generalization to Unknown Attacks

Many visual cues for non-intrusive spoofing detection have been already explored and impressive results have been reported on individual databases. However, the varying nature of spoofing attacks and acquisition conditions makes it impossible to predict how single anti-spoofing techniques, e.g. facial texture analysis, can generalize the problem in real-world applications. Moreover, we cannot foresee all possible attack scenarios and cover them in databases because the imagination of the human mind always finds out new tricks to fool existing systems. As one obviously cannot foresee all possible types of fake faces, one-class approach modeling only the genuine facial texture distribution could be a promising direction. This has been successfully applied in voice anti-spoofing [7], for instance.

B. Fusion of Countermeasures

It is reasonable to assume that no single superior technique is able to detect all known, let alone unseen, spoofing attacks. Therefore, the problem of spoofing attacks should be broken down into attack-specific subproblems that are solvable if a proper combination of complementary countermeasures is used. In this manner, a network of attack-specific spoofing detectors could be used to construct a flexible anti-spoofing framework in which new techniques can be easily integrated to patch the existing vulnerabilities in no time when new countermeasures appear. This obviously raises the problem of fusing different spoofing countermeasures which has not been studied much besides the algorithms [27], [29], [30] proposed within the context of the IJCB 2011 competition on counter measures to 2D facial spoofing attacks [7].

C. Biometric System + Countermeasures

A spoofing counter-measure is usually not designated to operate as a stand-alone procedure but in a joint operation with a recognition system. However, most works on anti-spoofing tend to focus only on the spoofing detection part hence omitting to integrate the counter-measure into a recognition system. In practice, integrating the counter-measure will affect the performance of the recognition system. While it will reduce its vulnerability to spoofing attacks, it may also decrease the recognition performance. The open issue is how to combine the spoofing counter-measure and the biometric recognition so that the combined biometric recognition system is robust to spoofing and does not suffer from reduced recognition accuracy [9].

D. Contextual Information

Face images captured from face spoofs may visually look very similar to the images captured from live faces. Thus, face spoofing detection may be difficult to perform based on only single face image or a relatively short video sequence. Depending on the imaging and face quality, it is nearly impossible, even for humans, to tell the difference between a genuine face and a fake one without any scene information or unnatural motion or facial texture patterns. However, we can immediately notice if there is something suspicious in the view, e.g. if someone is holding a video display or a photograph in front of the camera. Therefore, scenic cues can be exploited for determining whether display medium is present in the observed scene.

E. Challenge-Response Approach

Liveness and motion analysis based spoofing detection is rather difficult to perform by observing only spontaneous facial motion during short video sequences. This problem can be simplified by prompting the user to do some specific random action or challenge (such as a smiling and moving the head to the right). The user’s response (if any) will provide liveness evidences. This is called challenge-response approach for spoofing detection. The drawback of such an approach is that it requires user cooperation, thus making the authentication process a time-consuming. Another advantage of non-intrusive techniques is that from challenge-response based countermeasures it is rather easy to deduce which liveness cues need to be fooled. For instance, the request for uttering words suggests that analysis of synchronized lip movement and lip reading is utilized, whereas rotating head in a certain direction reveals that the 3D geometry of the head is measured. For non-intrusive approaches, it is usually not known which countermeasures are used, thus the system might be harder to deceive [22].

VI. CONCLUSION

To evaluate the vulnerabilities of face biometric systems when confronted to spoofing attacks, we discussed the performance of a baseline system on a challenging 2D face spoofing database consisting of 1300 video clips of real-accesses and attack attempts to different 50 identities. The 2D face spoofing attack database allows measuring the effectiveness of spoofing attacks or counter-measures to 2D face recognition systems. It is composed of two sets of data: real-accesses and attacks. Real-accesses are used to establish reference performance figures for recognition systems whereas attacks can be used to train spoofing classifiers or measure the impact of spoofing to existing baseline systems. The chosen baseline face verification system uses parts-based Gaussian Mixture Models and provides state-of-the-art performance. The experimental results showed that the face baseline system
achieves perfect performance when not confronted to spoof attacks. The performance sharply degrades in the presence of spoofing attacks. For instance, for a False Rejection Rate (FRR) of 0.1%, the FAR on real impostors goes from 0% to more than 80% when attacks are introduced. These results exemplify the vulnerability of face biometric systems against spoofing attacks.

Without spoofing counter-measures, most of the state-of-the-art facial biometric systems are indeed vulnerable to attacks, since they try to maximize the discriminability between identities without regards to whether the presented trait originates from a living legitimate client or not. The proposed anti-spoofing methods in the literature have shown very encouraging results on individual databases but may lack generalization to varying nature of spoofing attacks that can be encountered in real-world applications. This suggests that a network of attack-specific spoofing detectors maybe needed to tackle different spoofing attacks. The existing databases for spoofing and anti-spoofing analysis have been and are still useful for studying the spoofing problems but one cannot foresee all possible attack scenarios and cover them in databases. As the field evolves, new and more challenging databases can be expected. The imagination of the human mind always finds out new tricks to fool existing biometric systems. As one obviously cannot foresee all possible types of fake faces, one-class approach modeling only the genuine facial texture distribution could be a promising direction.

The open issues and the research directions that have been discussed in this paper are not specific to face biometrics but also hold for other biometric modalities. Due to lack of space, we focused this paper on face biometrics and did not report the results on voice, gait, fingerprints, retina, iris, vein, electro-physiological signals (EEG and ECG). In summary, the investigations in the TABULA RASA EU project showed that most of the biometrics modalities, including the multi-modal combinations, are vulnerable to spoofing attacks with different degrees.

Acknowledgments: The financial support of the Academy of Finland and NOKIA Foundation is acknowledged. This work is done within the FP7 EU European Project TABULARASA.

REFERENCES


