The REPLAY-MOBILE Face Presentation-Attack Database

Artur Costa-Pazo,[1] Sushil Bhattacharjee,[2] Esteban Va zquez-Fernandez[1] and Sebastien Marcel[2]

Abstract: Existing databases for evaluating face-PAD methods do not capture the specific characteristics of mobile devices. We introduce a new database, REPLAY-MOBILE, for this purpose. This publicly available database includes 1,200 videos corresponding to 40 clients. The database contains genuine-presentations as well as a variety of presentation-attacks. The database also provides three non-overlapping sets for training, validating and testing classifiers for face-PAD. These sets constitute a standardized protocol for comparing new approaches to existing algorithms fairly. We also provide baseline results with state-of-the-art approaches based on image quality analysis and face texture analysis.[3]

Keywords: Face recognition, face presentation attack detection, anti-spoofing, video database

1 Introduction

Presentation attacks (PA) present one of the most significant road-blocks to wide acceptance of facial authentication technology on mobile devices[VFGJ16]. State of the art face-*presentation attack detection* (PAD) methods achieve low error performance on current datasets [GMF14, Tea15]. However, these methods usually do not generalize well, beyond the databases on which they are trained [Fr13]. This lack of generalization becomes critical in the space of mobile devices. The quality of *presentation attack instruments* (PAI) (*i.e.*, mobile devices, printers, monitors, 3D scanners, *etc.*) is also improving rapidly. This implies not only that new methods for PAD need to be developed, but also that new datasets should be generated for realistic testing scenarios. Well known databases such as REPLAY-ATTACK or CASIA (see Table 1), still extensively used for evaluating new face-PAD methods, are no longer representative of the technology in current mobile devices. A modern database should consist of high resolution genuine videos and attacks, *presented as well as recorded* using modern mobile devices.

We present here the REPLAY-MOBILE database for face-PAD experiments. The database consists of 1,200 video clips of genuine- as well as attack-presentations, by 40 clients, under various lighting conditions. The database has been collected based on three guiding principles. (1) Sequences are captured on representative mobiles devices using the frontal camera. Both, tablets (*iOS*) and smartphones (*Android*) are used to represent the current spectrum of mobile devices. (2) During recording, clients hold the device in the same way

[1] GRADIANT, Galician Research and Development Center in Advanced Telecommunications, CITEXVI, loc. 14 – CUVI, 36310, Vigo (Po.), Spain, acosta@gradiant.org, evazquez@gradiant.org

[2] Idiap Research Institute, Centre du Parc, Rue Marconi 19, PO Box 592, CH-1920, Martigny, Suisse, sushil.bhattacharjee@idiap.ch, sebastien.marcel@idiap.ch

[3] This work was partially supported by *GAIN, Axencia Galega de Innovación, Consellería de Economía, Emprego e Industria, Xunta de Galicia* (IN809A. December 30, 2014), EU H2020 project TeSLA, and by the Swiss Center for Biometrics Recognition and Test. The online version of this paper provides more detailed discussions about the database and the experiments presented here. The source-code for the experiments reported in this paper are available at the following link: https://pypi.python.org/pypi/bob.paper.BioSig2016_ReplayMobile

as they would do in a real scenario. (3) Attacks are performed using high resolution videos presented on a matte-screen (to avoid specular reflections) and high-quality prints on matte paper.

After a brief survey of related research (Section 2), the three main contributions of this paper are presented, namely: (1) the REPLAY-MOBILE database, for evaluating face-PAD algorithms specifically for mobile devices (Section 3); (2) two sets of face-PAD results, one based on image-quality measures (our baseline), and the other based on texture-analysis (Section 4); and, (3) performance results reported using newly standardized ISO metrics (see the ISO/IEC 30107-3 standard[4]). The experimental results presented in Section 5 are followed by a summary of our conclusions in Section 6.

2 Related Work

This section provides a brief overview of face-PAD approaches, and the relevant databases. For face-PAD approaches we adopt a simple taxonomy, based on two categories: liveness detection based on motion cues, and, image-quality based approaches.

For detecting printed-photo attacks, Anjos *et al.* [ACM13] use strong correlation between the estimated optical flow for the face-region and that for the background, is an indicator of a PA. Several heuristics have been developed for detecting eye-blinks [CD14]. Pinto *et al.* [Si12] treat each video as a 3D data-set (instead of a sequence of 2D frames) and compute a number of statistical descriptors over this data. A recent work [Tea15] attempts to detect involuntary movements using dynamic mode decomposition (DMD) of optical flow to characterize genuine presentations. While not attempting to capture high-level cues directly, this method can detect eye-blinks and lip movements in a face-video [Tea15]. For a face-recognition system, a PA often consists in replaying, to the camera, a video of an enrolled person whose identity is being spoofed. The process of re-capture and playback typically introduces distortions in the video-data that would not be seen in a live data-capture.

Galbally *et al.* have proposed a set of 25 image-quality measures (IQM) [GMF14], well known in the image-compression community, to detect PAs. Wen *et al.* [WHJ15] have proposed a different set of image-quality features, that attempt to characterize color-diversity, image-sharpness, and the amount of specularity present in the image. Whereas the IQMs used in [GMF14] are computed on gray-images (the Y component of a color-frame in YCbCr representation), the features proposed in [WHJ15] are evaluated on color-images (except for the image-sharpness features).

Videos re-captured from a digital display device often exhibit Moiré patterns. Several researchers have used Moiré pattern detectors [Zh12, GQ15, Pa15] for face-PAD. Zhang *et al.* [Zh12] have proposed a Moiré pattern detector based a two-class classifier to detect the presence of high-frequency components in the image. Garcia *et al.* [GQ15] use a set of Mexican-Hat filters to decompose the image. They then assume that a Moiré pattern is present if the energy in any of the filter-responses is stronger than a threshold. Patel *et al.*

[4]https://www.iso.org/obp/ui/#iso:std:iso-iec:30107:-1:ed-1:v1:en

use multi-scale LBP, to detect Moiré patterns in the spatial domain. Overall, however, these methods have limited success in face-PAD, since Moiré patterns are not guaranteed to be present in all PAs. One efficient way to use a Moiré pattern detector is as a pre-filtering step.

Face-PAD methods relying on texture information present in the face-region have also been proposed [CAM12, Yea13]. For our purposes, such methods may be included in the category of image-quality based approaches.

Database Name	Max. Res.	Notes
REPLAY-ATTACK (2012) [CAM12]	320×240	1200 videos for 50 subjects
CASIA (2012) [Zh12]	1920×1080	videos captured on SLR camera, not mobile devices
MSU-MFSD (2015) [WHJ15]	720×480	350 videos for 35 subjects. Videos captured on a mix of devices including smartphones, tablets, SLR

Tab. 1: Existing public-domain databases for face-PAD experiments.

Some databases commonly used for benchmarking face-PAD methods are listed in Table 1. Some of these databases (REPLAY-ATTACK, and MSU-MFSD) have videos that do not have a resolution high enough to be representative of current mobile-devices. Although CASIA database includes some high-resolution videos, these were captured on a conventional camera (Sony NEX-5), and not on mobile devices such as smartphones.

The robustness of a PAD method depends on the training and evaluation dataset used, as well as on the technology used for face presentation and acquisition. This leads to the question: can we fairly evaluate the performance face-PAD method designed for use on mobile devices, without mobile-specific databases? This question motivates the REPLAY-MOBILE database presented in this work.

3 The REPLAY-MOBILE Database

The REPLAY-MOBILE database[5] consists of short video recordings of both real-access and attack attempts to 40 different identities. This section presents the details of the data-collection process, as well as an explanation of the evaluation protocols that are provided.

The videos comprising this database have been collected in two sessions separated by an interval of two weeks. In the first session both enrollment videos and media for manufacturing the attacks were collected under two different illumination conditions, namely *lighton* (electric lights in the room are switched on) and *lightoff* (electric lights are turned off). In both scenarios the background of the scene is homogeneous and a tripod is used for the capturing device. (More details are provided in Section 3.)

In the second session each client recorded 10 videos, under the following 5 different scenarios and paying special attention to the lightning conditions:

[5] The database may be downloaded using the following URL: `https://www.idiap.ch/dataset/replay-mobile`

1. *controlled*: uniform background, office-light on, window blinds down.
2. *adverse*: uniform background, office-light off, window blinds halfway up.
3. *direct*: complex background, user facing a window with direct sunlight.
4. *lateral*: complex background, user near a window under lateral sunlight.
5. *diffuse*: complex background, user in an open indoor-space, with diffuse light.

When recording the video, the user was asked to stand, to hold the mobile device at the eye level and to center the face on the screen of the video capture application. Each video is approximately 10 seconds long ($\sim$ 300 frames @ 30fps) and HD resolution (720×1280). (Note that the videos in the MSU-MFSD database have relatively lower resolution.) In each lighting condition the user captured two videos, one using an *iPad Mini 2* tablet and another using a *LG-G4* smartphone. Figure 1a shows examples of genuine presentations in the database.

(a) Genuine presentation examples

(b) Attack presentation examples

Fig. 1: Examples of genuine- and attack-presentations in different scenarios. (a) Geunine presentation samples. Top row: samples acquired on a smartphone. Bottom row: samples captured on a tablet. Columns from left to right show video frames in *controlled*, *adverse*, *direct*, *lateral*, and *diffuse* scenarios, respectively. (b) Attack presentations. Top row: samples captured on a smartphone. Bottom row: samples captured on a tablet. Columns, from left to right, show examples of *mattescreen-lighton*, *mattescreen-lightoff*, *print-lighton*, and *print-lightoff* scenarios, respectively.

To create the attacks, a separate set of high resolution photos and videos was first collected, under the same illumination conditions as in the video collection sessions. Each user was asked to sit down in front of two devices while the acquisition operator captured the data under the conditions previously defined (*lighton* and *lightoff*). For photo-based attacks, a *Nikon Coolpix P520* camera was used to capture high resolution images (18 Mpixel). Video-based attacks were recorded by using the back camera of the *LG-G4* smartphone, which records *1080p FullHD* video clips.

The attacks have been created using two different PAIs: *mattescreen*: photos and videos for each client are displayed on a *Philips 227ELH* monitor with a resolution of 1920×1080 pixels; and *print*: hard-copies of high-resolution digital photographs are printed on plain A4 matte paper (using a *Konica Minolta ineo+ 224e* color laser printer).

Each attack was recorded on each mobile device (tablet and smartphone) for 10 seconds. For recording *mattescreen* attacks the capturing mobile device was supported on a fixed

support. Each *print* video, however, was captured in two different attack modes: *hand-held attack*, where the operator holds the capture device; and *fixed-support attack*, where the capture device is fixed on a support. Thus, four different PAIs are represented in REPLAY-MOBILE. Figure 1b shows examples of attacks available in the database.

Videos in the REPLAY-MOBILE database are grouped into 3 disjoint subsets: *train*, *development* and *test*. Identities for each subset have been selected via demographic analysis: each subset has equable distribution for identities based on gender, age and eye-wear.

Table 2 summarizes the organization of videos in the various protocols for face-PAD experiments. Each row in the table (a specific *Scenario-Type* pair) corresponds to one PAI. The column-labels *Mobile* and *Tablet* indicate the capture-device used. Besides the two protocols (*mattescreen* and *print*), a *Grandtest* protocol is also provided, for global performance evaluation[6].

		Mobile			Tablet			
Scenario	Type	Train	Devel	Test	Train	Devel	Test	Total
	real-access	60	80	60	60	80	60	400
mattescreen-attack	photo-fixed	24	32	24	24	32	24	160
	video-fixed	24	32	24	24	32	24	160
print-attack	print-fixed	24	32	24	24	32	24	160
	print-hand	24	32	24	24	32	24	160
Grandtest-attack		156	208	156	156	208	156	1040

Tab. 2: Number of videos in each subset of the REPLAY-MOBILE database.

4 The Studied Face-PAD Approaches

In this section we describe the two face-PAD methods that we have applied to the REPLAY-MOBILE database. The first method is based on image-quality measures, and serves as our baseline. We also propose a new method for face-PAD, based on Gabor-jets. Experimental results for these methods are reported in Section 5.

Our baseline, against which to compare the results of the proposed method, is derived from a set of image-quality measures (IQM), first used for face-PAD by Galbally *et al.* [GMF14]. Some of the IQMs used by Galbally *et al.* [GMF14], have been computed using third-party executables and are not easily reproducible. Our experiments are based a subset of reproducible features. Specifically, from the set of 25 IQMs proposed by Galbally *et al.* [GMF14], we have used a subset of 18 IQMs. The features used in our experiments are listed in Table 3.

Here we propose a new texture-based approach for face-PAD, using Gabor-jets [Wea97], previously used by GRADIANT [Gu13] for face-recognition. To our knowledge this texture-descriptor has not previously been applied to the problem of face-PAD. The face-images, cropped to 85×100 pixels, are preprocessed using an adaptation of the retina layer model

[6] In Table 2, each element in the *Grandtest* row is the sum of the remaining elements in the corresponding column.

F#	Name	Abbrev.	F#	Name	Abbrev.
1	Mean Squared Error	MSE	10	R-averaged Max. difference (r=10)	RAMDv
2	Peak Signal to Noise Ratio	PSNR	11	Mean angle similarity	MAS
3	Average difference	AD	12	Mean angle magnitude similarity	MAMS
4	Structural content	SC	13	Spectral magnitude error	SME
5	Normalized cross-correlation	NK	14	Gradient magnitude error	GME
6	Max. difference	MD	15	Gradient phase error	GPE
7	Laplacian MSE	LMSE	16	Structural similarity index	SSIM
8	Normalized Abs. error	NAE	17	Visual information fidelity	VIF
9	Signal to noise ratio	SNRv	18	High-low frequency index	HLFI

Tab. 3: List of image-quality measures (IQM) used in the baseline experiments. See [GMF14] for details of these features.

[VC09]. Gabor-jets are then computed over a regular 10×10 grid using 40 Gabor wavelets with default parametrization [Wea97]. For each face-image a 4000-element feature-vector is recorded.

5 Experimental Results

To evaluate the PAD performance, we have used standard ISO/IEC 30107-3 metrics, namely, APCER: Attack Presentation Classification Error Rate; and BPCER: *Bona fide* Presentation Classification Error Rate. We also provide the ACER (Average Classification Error Rate), defined as $(APCER + BPCER)/2$. To aid comparison with previously published works, we also report the half-total error rates (HTER) for our experiments. Galbally *et al.* [GMF14] have used linear discriminant analysis (LDA) in their experiments, to achieve a HTER of 15.2% on the REPLAY-ATTACK database. Our experiments show that a support-vector machine (SVM) with a radial-basis function (RBF) kernel yields better face-PAD results (HTER = 5.3%) on REPLAY-ATTACK database, than LDA (using the same features). Therefore, baseline results on the REPLAY-MOBILE database are reported in Table 4 using only the SVM-RBF classifier. The table reports the EER for the development set, and the HTER achieved on the test set, for different combinations of *scenario* and *type* defined in the REPLAY-MOBILE database. The overall performance, as shown for the experiment [Grandtest] in Table 4, is 7.8%. We use this method to set our face-PAD baseline on this database.

	Mattescreen		Print		Grandtest
	photo	video	fixed	hand	
Dev. EER (%)	7.93	11.70	5.31	4.98	7.50
Test. HTER (%)	7.70	13.64	4.22	5.43	7.80

Tab. 4: Baseline results using SVM-RBF classifier ($\gamma = 1.5$) on IQM features (see Table 3) computed for the face-PAD protocol of the REPLAY-MOBILE database.

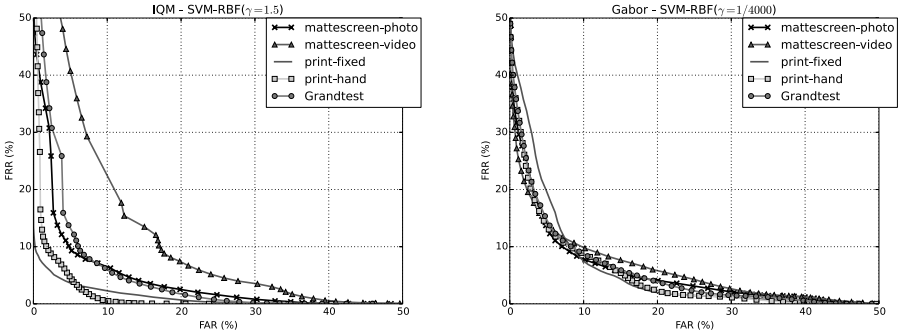
A two-class classifier is constructed for the 4000-D Gabor-jet feature-vectors using SVM-RBF ($\gamma = \frac{1}{4000} = 0.00025$). The results achieved on REPLAY-MOBILE and the comparison with the IQM baseline are shown in Table 5. The table also highlights the advantage

of using APCER and BPCER, over HTER. Both methods (IQM-based and Gabor-based face-PAD) show similar HTER. Using APCER and BPCER, however, we note that the Gabor-based approach seems to be more consistent among different presentation attack instruments (PAI), which indicates that it is the more robust of the two approaches.

	Test. HTER (%)					Test. ACER (%)	Test. APCER (%)	Test. BPCER (%)
	MP	MV	PF	PH	GT			
IQM	7.70	13.64	4.22	5.43	7.80	13.64	19.87	7.40
Gabor	8.64	9.53	9.40	8.99	9.13	**9.53**	7.91	11.15

Tab. 5: The proposed Gabor-jet based face-PAD method compared with the baseline in REPLAY-MOBILE database using the HTER, ACER, APCER, BPCER (%) measures. The HTER results are reported for each protocol, indicated by the following column-headings: MP – *mattescreen-photo*, MV – *mattescreen-video*, PF – *print-fixed*, PH – *print-hand* and GT – *Grandtest*.

The detection-error tradeoff (DET) curves in Figure 2 show the influence of each attack. These plots show that the Gabor-jet based face-PAD shows consistent performance for the different kinds of attacks, whereas the performance of the image-quality based approach varies significantly among the various attack-types.



(a) IQM-based (b) Gabor-based

Fig. 2: DET curves for the various attack protocols. The performance of the IQM based PAD method (a) varies significantly among the different kinds of attacks. By contrast, the Gabor-jet based approach (b) is more consistent over the range of attack-types.

6 Conclusions

We have introduced REPLAY-MOBILE, a new, freely available database for fair evaluation of face-PAD methods on mobile devices. The key features of REPLAY-MOBILE are: (1) high-resolution videos captured under realistic conditions of device-usage, including a variety of illumination conditions; (2) a variety of presentation-attacks; (3) a pre-defined protocol for unbiased training and fair evaluation of new face-PAD methods. We have also proposed a new approach for face-PAD based on Gabor-jets, and have compared its performance a baseline approach based on image quality assessment. We demonstrate that

using the newly standardized metrics APCER and BPCER lead to a fairer comparison of anti-spoofing algorithms.

References

[ACM13] Anjos, A.; Chakka, M. M.; Marcel, S.: Motion-Based Counter-Measures to Photo Attacks in Face Recognition. Institution of Engineering and Technology Journal on Biometrics, July 2013.

[CAM12] Chingovska, I.; Anjos, A.; Marcel, S.: On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. In: Proc. IEEE Intl. Conf. of Biometrics Special Interest Group (BIOSIG). 2012.

[CD14] Chakraborty, S.; Das, D.: An Overview of Face Liveness Detection. International Journal on Information Theory (IJIT), 3(2):11–25, April 2014.

[Fr13] de Freitas Pereira, T.; Anjos, A.; de Martino, J. M.; Marcel, S.: Can Face Anti-spoofing Countermeasures Work In a Real World Scenario? In: Proc. Intl. Conf. on Biometrics (ICB). 2013.

[GMF14] Galbally, J.; Marcel, S.; Fierrez, J.: Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint and Face Recognition. IEEE Trans. on Image Processing, 23(2):710–724, February 2014.

[GQ15] Garcia, D. C.; Queiroz, R. L.: Face Spoofing 2D-Detection Based on Moiré-Pattern Analysis. IEEE Trans. on Information Forensics and Security, 10(4):778–786, April 2015.

[Gu13] Guenther, M.; Costa-Pazo, A.; Ding, C.; et al.: The 2013 face recognition evaluation in mobile environment. In: Proc. Intl. Conf. on Biometrics (ICB). 2013.

[Pa15] Patel, K.; Han, H.; Jain, A. K.; Ott, G.: Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks. In: Proc. 8th Intl. Conf. on Biometrics (ICB). 2015.

[Si12] da Silva Pinto, A.; Pedrini, H.; Schwartz, W.; Rocha, A: Video Based Face Spoofing Detection Through Visual Rhythm Analysis. In: Proc. 25th SIBGRAPI. 2012.

[Tea15] Tirunagari, S.; et al.: Detection of Face Spoofing Using Visual Dynamics. IEEE Trans. Information Forensics and Security, 10(4):762–777, April 2015.

[VC09] Vu, Ngoc-Son; Caplier, A.: Illumination-robust face recognition using retina modeling. In: Proc. of Intl. Conf. on Im. Proc. (ICIP). 2009.

[VFGJ16] Vazquez-Fernandez, E.; Gonzalez-Jimenez, D.: Face recognition for authentication on mobile devices. Image and Vision Computing, 2016.

[Wea97] Wiskott, L.; et al.: Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7):775–779, July 1997.

[WHJ15] Wen, D.; Han, H.; Jain, A. K.: Face Spoof Detection With Image Distortion Analysis. IEEE Trans. on Information Forensics and Security, 10(4):746–761, April 2015.

[Yea13] Yang, J.; et al.: Face Liveness Detection with Component Dependent Descriptor. In: Proc. IAPR IEEE Intl. Joint Conf. on Biometrics (IJCB). 2013.

[Zh12] Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S.Z.: A face antispoofing database with diverse attacks. In: IAPR Intl. Conf. on Biometrics (ICB). 2012.