# OULU-NPU: A Mobile Face Presentation Attack Database with Real-World Variations

# OULU-NPU: A mobile face presentation attack database with real-world variations

Zinelabinde Boulkenafet[1], Jukka Komulainen[1], Lei li[2], Xiaoyi Feng[2] and Abdenour Hadid[1,2]
[1] Center for Machine Vision and Signal Analysis, University of Oulu, Finland
[2] Northwestern Polytechnical University, School of Electronics and Information, Xian, China

*Abstract*— **The vulnerabilities of face-based biometric systems to presentation attacks have been finally recognized but yet we lack generalized software-based face presentation attack detection (PAD) methods performing robustly in practical mobile authentication scenarios. This is mainly due to the fact that the existing public face PAD datasets are beginning to cover a variety of attack scenarios and acquisition conditions but their standard evaluation protocols do not encourage researchers to assess the generalization capabilities of their methods across these variations. In this present work, we introduce a new public face PAD database, OULU-NPU, aiming at evaluating the generalization of PAD methods in more realistic mobile authentication scenarios across three covariates: unknown environmental conditions (namely illumination and background scene), acquisition devices and presentation attack instruments (PAI). This publicly available database consists of 5940 videos corresponding to 55 subjects recorded in three different environments using high-resolution frontal cameras of six different smartphones. The high-quality print and video-replay attacks were created using two different printers and two different display devices. Each of the four unambiguously defined evaluation protocols introduces at least one previously unseen condition to the test set, which enables a fair comparison on the generalization capabilities between new and existing approaches. The baseline results using color texture analysis based face PAD method demonstrate the challenging nature of the database.**

## I. Introduction

The use of face modality is especially appealing in mobile biometrics because it is highly accepted among users, considering the ”selfie generation”, and can be also easily integrated in the natural interaction with the devices. Moreover, nowadays almost every mobile device is equipped with a decent front-facing camera, while fingerprint and iris sensors are just emerging. Face recognition is indeed being increasingly deployed in mobile applications. As an example, MasterCard is trialling a ”selfie verification” feature to secure its new mobile payment service.

Spoofing (or presentation attacks as defined in the current ISO/IEC 30107-3 standard [8]) poses serious security issue to face recognition or biometric systems in general. The vulnerabilities of face-based biometric systems to spoofing have been now recognized and face presentation attack detection (PAD) has finally received significant attention in the research community [1], [3], [7]. Yet we lack generalized software-based face PAD methods performing robustly in the unknown operational conditions of practical mobile authentication scenarios. For instance, in a recent study [9], six commercial face recognition systems, namely Face Unlock, Facelock Pro, Visidon, Veriface, Luxand Blink and FastAccess, were easily fooled with crude photo attacks using images of the targeted person downloaded from social networks. Even worse, also their dedicated challenge-response based liveness detection mechanisms were circumvented using simple photo manipulation to imitate the requested facial motion (liveness cues), including eye blinking and head rotation.

The existing public datasets for developing and benchmarking software-based face PAD methods are beginning to cover a variety of attack scenarios and acquisition conditions [4], [5], [13], [15]. However, the main problem is that their standard evaluation protocols do not encourage researchers to assess the generalization capabilities of their PAD methods across these variations partly due to the lack of data. Instead, the methods are evaluated using the homogeneous train and test sets, i.e. corresponding to exactly the same known operating conditions and artifacts, when many of the existing face PAD methods achieve astonishing, near 0%, error rates. The preliminary studies on generalized face spoof detection [2], [3], [6], [12], [13] have shown that these reported performances are indeed overly optimistic estimate on their actual performance in real-world authentication applications. While the existing datasets have been and continue to be useful for the research community, the remarkable results in intra-database experiments but lack of generalization capabilities among face PAD methods indicates that more challenging configurations are needed before the research on non-intrusive software-based face spoof detection can reach the next level.

In this paper, we address this issue and introduce a new public face PAD database, OULU-NPU, which aims at evaluating the generalization of PAD methods in more realistic mobile authentication scenarios across three covariates: unknown environmental conditions (namely illumination and background scene), acquisition devices and presentation attack instruments (PAI). Altogether, the database consists of 5940 videos corresponding to 55 subjects recorded in three different illumination conditions using high-resolution frontal cameras of six different recent smartphones. High-quality print and video-replay attacks were created using two printers and two display devices. The first three evaluation protocols assess the effect of each covariate separately, i.e. each of them introduces one previously unseen condition to the test

The financial support of the Academy of Finland, Infotech Oulu and the Northwestern Polytechnical University is acknowledged.

TABLE I: Comparison between the existing face PAD databases and the new OULU-NPU.

| Database | # subjects | Acquisition devices | # lighting scenarios | PAIs | # real/attack videos | Fixed validation set |
|---|---|---|---|---|---|---|
| Replay-Attack [4] | 50 | 1 laptop | 2 | 1 printer & 2 displays | 200/1000 | Yes |
| CASIA-FASD[13] | 50 | 3 webcams | 1 | 1 printer & 1 display | 150/450 | No |
| MSU-MFSD [15] | 35 | 1 laptop & 1 smartphone | 1 | 1 printer & 2 displays | 110/330 | No |
| Replay-Mobile [5] | 40 | 1 smartphone & 1 tablet | 5 | 1 printer & 1 display | 390/640 | Yes |
| OULU-NPU | 55 | 6 smartphones | 3 | 2 printers & 2 displays | 1980/3960 | Yes |

set which is not present in the training material. The fourth protocol is designed to simulate a real-world scenario where all the three variations were taken into consideration at the same time. In addition, the 55 subjects are divided into subject-disjoint training, development and testing because the use of unambiguous evaluation protocol with fixed validation set enables unbiased comparison between new and existing approaches. We provide baseline results of a state-of-the-art method based on color texture analysis [2] that clearly demonstrate the challenging nature of the database.

The rest of this paper is organized as follows. In Section II, we introduce the evolution of publicly available face PAD databases and discuss their advantages and shortcomings. The new OULU-NPU face presentation attack detection database is presented in Section III. Section IV describes the benchmark experiments and results. Finally, Section V concludes the paper.

## II. Related work

In the very early phase of face PAD related research, even software-based approaches were evaluated on proprietary databases. The use of private data can be seen somewhat reasonable when demonstrating proof-of-concept custom imaging solutions or (random) challenge-response based approaches introducing specific user interaction demands. The results of non-intrusive software-based methods, however, should be easily reproduced and fairly compared because they are just further processing the same images (or videos) used for the actual authentication purposes or additional data captured with conventional cameras. Furthermore, the lack of publicly available data is likely to rule out many potential researchers working on PAD. It was not a coincidence that after the release of the first public PAD dataset, NUAA Photograph Imposter Database (NUAA-PID) [11], the research on face PAD exploded.

Shortly after NUAA-PID, larger scale video-based public datasets with both print and video-replay attacks were released, namely CASIA Face Anti-Spoofing Database (CASIA-FASD) [15] and Replay-Attack Database [4], each consisting of 50 subjects. These databases introduce some variations in the acquisition conditions. The data in the CASIA-FASD was captured using three cameras with varying level of image quality and resolution, i.e. low, medium and high, while the Replay-Attack Dataset considers two authentication scenarios with two illumination conditions and backgrounds, i.e. controlled and adverse. Although the CASIA-FASD, is smaller than the Replay-Attack Database, it has shown to be more challenging benchmark dataset due to the diversity in the data, including attack types and (less-controlled) acquisition conditions in general, e.g. standoff distance and input sensor quality.

The Replay-Attack Database and CASIA-FASD are still the main datasets used for developing and benchmarking face PAD methods. However, these datasets are not representative of the current mobile authentication scenarios. First, the data acquisition was conducted with generic web cameras or conventional digital cameras whose image quality and resolution is either too low or too high considering the latest generations of mobile devices. Furthermore, the use of stationary cameras does not correspond to the mobile applications where the user holding the device poses additional variations, thus new challenges, in the acquired face videos, including global motion, sudden illumination changes, extreme head poses and various background scenes. Face PAD in mobile scenarios does not have to be more difficult by default but the nature of the development and benchmark data must be replicate realistic of mobile authentication scenario [13].

Recently, the MSU Mobile Face Spoof Database (MSU-MFSD) [13] and the Replay-Mobile database [5] introduced mobile authentication scenarios to public face PAD benchmark datasets. In both datasets, two different acquisition devices were used for recording the real accesses and attack attempts. While the MSU-MFSD considers only small illumination variations as the real subjects were recorded in the same laboratory environment, the Replay-Mobile Database includes five different mobile scenarios and paying special attention to the lighting conditions. Therefore, it is very unfortunate that the dataset suffers from a severe flaw as the background scenes differ between the real accesses and the attack attempts. Thus, the dataset can be probably easily broken with algorithms utilising the whole video frame (context) for PAD, like [10].

The current publicly available databases have been a very important kick-off for finding out best practices for face PAD and have provided valuable insight on the different aspects in solving the problem. Many potentially useful approaches for face PAD, including from liveness cues, like eyeblink detection [10], to static image propeties, like texture [2], [3], [10] and distortions in image quality [13], have been explored. However, the databases have been partially misleading the research into wrong direction as well as a relatively large part of the research has been concentrating on breaking the datasets instead of really trying to bring new theoretical insight into the problem of face PAD. As an outcome, we still lack low-cost generalized methods that could be transferred to practical applications like mobile authentication scenarios.

Fig. 1: Samples of the subjects recorded in the database.

While existing publicly available databases still continue to be valuable tools for the community, more challenging datasets are needed to reach the next level and solve some fundamental generalization related problems in face PAD.

As seen above and in Table I, the existing public datasets are beginning to cover the different variations in e.g. illumination, acquisition devices and the attacks themselves, that will be definitely faced in real operational conditions. However, the main issue is that they do not provide default evaluation protocols for evaluating the actual generalization capabilities of the new PAD methods across these covariates. One reason for this is that the databases are rather small, when also the variations in some factors are still limited. For instance, the MSU-MFSD considers only one illumination condition and only two different cameras were employed in collecting both the MSU-MFSD and the Replay-Mobile Database. The variation in PAIs is another important factor that cannot be extensively studied using the existing benchmarks because they include at most one high-quality print and video-replay attack.

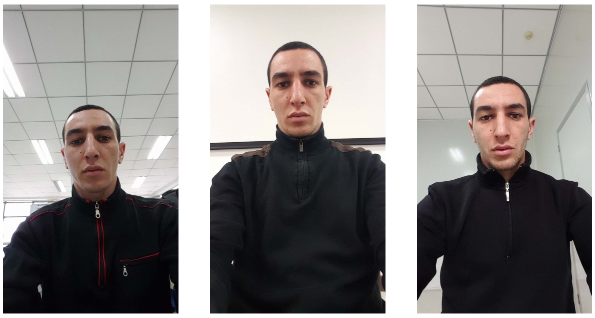
It is also worth highlighting that some of the benchmark datasets, like the CASIA-FASD and MSU-MFSD, contain separate folds only for training and testing, which may cause bias due to ”data peeking”. While independent (third-party) testing [14] is practically impossible to arrange without collective evaluations, the use of pre-defined training, development and test sets would mitigate the effect of tuning the methods on the test data, thus allowing a fairer direct comparison between new and existing approaches.

## III. The OULU-NPU face PAD database

In this work, we address many of the issues mentioned in the previous section and introduce the new OULU-NPU face PAD database. The aim of the dataset is particularly at evaluating the generalization of new PAD methods in more realistic mobile authentication scenarios by considering three covariates: unknown environmental conditions (namely illumination and background scene), acquisition devices and presentation attack instruments (PAI), separately and at once. In the following, we describe the new OULU-NPU face PAD database and its evaluation protocols in detail.

### A. Collection of real access attempts

The OULU-NPU presentation attack detection database includes short video sequences of real access and attack attempts corresponding to 55 subjects (15 female and 40 male). Figure 1 shows samples of these subjects. The real access attempts were recorded in three different sessions separated by a time interval of one week. During each session, a different illumination condition and background scene were considered (see Figure 2):



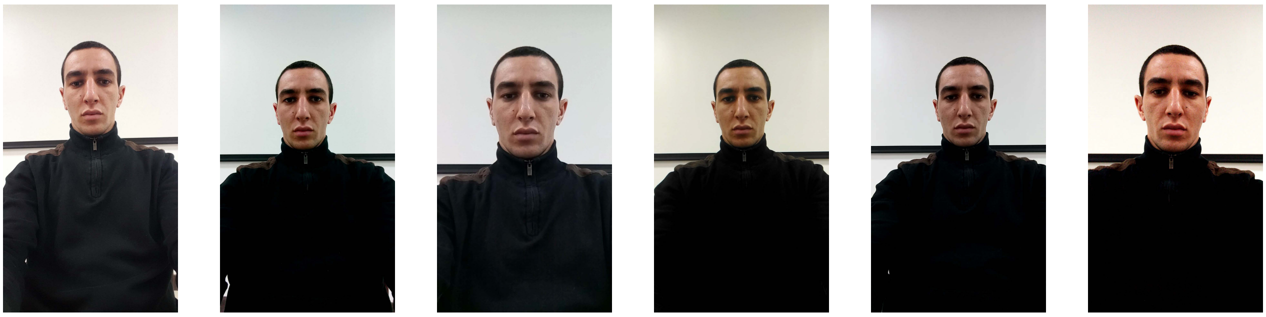
(a) Session 1 (b) Session 2 (c) Session 3

Fig. 2: Sample images of a real subject highlighting the illumination conditions across the three different scenarios.

- *Session 1*: The recordings were taken in an open-plan office where the electronic light was switched on and the windows blinds were up and the windows were located behind the users.
- *Session 2*: The recordings were taken in a meeting room where the electronic light was the only source of illumination.
- *Session 3*: The recordings were taken in a small office where the electronic light was switched on and the windows blinds were up and the windows were located in front of the users.

During each session, the subjects recorded two videos of themselves (one for the enrollment and one for the actual access attempt) using the frontal cameras of the mobile devices. In order to simulate realistic mobile authentication scenarios, the video length was limited to five seconds and the clients were asked to hold the mobile device like they were being authenticated but without deviating too much from their natural posture while normal device usage.

The recent advances in sensor technology have introduced high-resolution cameras also to the mid range models of



(a) Samsung (b) HTC (c) MEIZU (d) ASUS (e) Sony (f) OPPO

Fig. 3: Sample images showing the image quality of the different camera devices.

the last generation mobile devices capable of capturing good quality images (and videos) in daylight and indoor conditions. Considering that the acquisition quality of the embedded (both front and rear) cameras can be expected to be growing generation by generation, we selected six smartphones with high-quality front-facing cameras in the price range from €250 to €600 for the data collection:

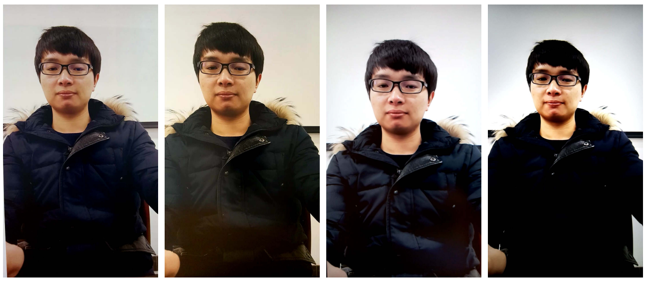
- Samsung Galaxy S6 edge (Phone 1) with 5 MP frontal camera.
- HTC Desire EYE (Phone 2) with 13 MP frontal camera.
- MEIZU X5 (Phone 3) with 5 MP frontal camera.
- ASUS Zenfone Selfie (Phone 4) with 13 MP frontal camera.
- Sony XPERIA C5 Ultra Dual (Phone 5) with 13 MP frontal camera.
- OPPO N3 (Phone 6) with 16 MP rotating camera.

The videos were recorded at Full HD resolution, i.e. 1920 × 1080 using the frontal cameras of the six mobile devices and the same camera software[1] installed on each device. Even though the nominal camera resolution of some mobile devices is the same, like Sony XPERIA C5 Ultra Dual, HTC Desire EYE and ASUS Zenfone Selfie (13 MP), significant differences can be observed in the quality of the resulting videos as demonstrated in Figure 3.

### B. Attack creation

Assuming that the legitimate users are trying to get authenticated in multiple conditions, it is important to collect the data of genuine subjects in multiple lighting conditions from the usability point of view. In contrast, the attackers try to present as high-quality artifact as they can to the input camera in order to maximize the chance of successfully fooling a face biometric system. Therefore, the attacks should be carefully designed and conducted in order to guarantee that they are indeed hard to detect.

During each of the three sessions, a high-resolution photo and video of each user was captured using the back camera of the Samsung Galaxy S6 Edge phone capable of taking 16 MP still images and Full HD videos. These high resolution photos and videos were then used to create the presentation attacks. The attack types considered in this database are print and video-replay attacks:



(a) Print 1 (b) Print 2 (c) Replay 1 (d) Replay 2

Fig. 4: Samples of print and replay attacks taken with the front camera of Sony XPERIA C5 Ultra Dual.

- *Print attacks*: The high resolution photos were printed on A3 glossy paper using two different printers: a Canon imagePRESS C6011 (Printer 1) and a Canon PIXMA iX6550 (Printer 2).
- *Video-replay attacks*: The high-resolution videos were replayed on two different display devices: a 19" Dell UltraSharp 1905FP display with 1280 × 1024 resolution (Display 1) and an early 2015 Macbook 13" laptop with Retina display of 2560 × 1600 resolution (Display 2).

The print and video-replay attacks were then recorded using the frontal cameras of the six mobile phones. While capturing the print attacks, the facial prints were held by the operator and captured with stationary capturing devices in order to maximize the image quality but still introduce some noticeable motion in the print attacks. In contrast, when recording the video-replay attacks both of the capturing devices and PAIs were stationary. Furthermore, we paid special attention that the background scene of the attacks matches the real accesses during each session and that the attack videos do not contain the bezels of the screens or edges of the prints. Figure 4 shows samples of the attacks captured using the Sony XPERIA C5 Ultra Dual.

### C. Evaluation protocols

To evaluate the performances of the face PAD methods on the OULU-NPU database, we designed four protocols.

[1] http://opencamera.sourceforge.net/

TABLE II: The detailed information about the video recordings in the train, development and test sets of each protocol.

| Protocol | Subset | Session | Phones | Users | Attacks created using | # real videos | # attack videos | # all videos |
|---|---|---|---|---|---|---|---|---|
| Protocol I | Train | Session 1,2 | 6 Phones | 1-20 | Printer 1,2; Display 1,2 | 240 | 960 | 1200 |
| | Dev | Session 1,2 | 6 Phones | 21-35 | Printer 1,2; Display 1,2 | 180 | 720 | 900 |
| | Test | Session 3 | 6 Phones | 36-55 | Printer 1,2; Display 1,2 | 240 | 960 | 1200 |
| Protocol II | Train | Session 1,2,3 | 6 Phones | 1-20 | Printer 1; Display 1 | 360 | 720 | 1080 |
| | Dev | Session 1,2,3 | 6 Phones | 21-35 | Printer 1; Display 1 | 270 | 540 | 810 |
| | Test | Session 1,2,3 | 6 Phones | 36-55 | Printer 2; Display 2 | 360 | 720 | 1080 |
| Protocol III | Train | Session 1,2,3 | 5 Phones | 1-20 | Printer 1,2; Display 1,2 | 300 | 1200 | 1500 |
| | Dev | Session 1,2,3 | 5 Phone | 21-35 | Printer 1,2; Display 1,2 | 225 | 900 | 1125 |
| | Test | Session 1,2,3 | 1 Phone | 36-55 | Printer 1,2; Display 1,2 | 60 | 240 | 300 |
| Protocol VI | Train | Session 1,2 | 5 Phones | 1-20 | Printer 1; Display 1 | 200 | 400 | 600 |
| | Dev | Session 1,2 | 5 Phones | 21-35 | Printer 1; Display 1 | 150 | 300 | 450 |
| | Test | Session 3 | 1 Phone | 36-55 | Printer 2; Display 2 | 20 | 40 | 60 |

*1) Protocol I:* The first protocol is designed to evaluate the generalization of the face PAD methods under different environmental conditions, namely illumination and background scene. As the data is recorded in three sessions with different illumination conditions and locations, the train, development and evaluation sets can be constructed using video recordings taken from different sessions, see Table II.

*2) Protocol II:* Since different PAI (i.e. different displays and printers) create different artifacts, it is necessary to develop face PAD methods robust to this kind of variations. The second protocol is designed to evaluate the effect of the PAI variation on the performance of the face PAD methods by introducing previously unseen PAI in the test set as shown in Table II.

*3) Protocol III:* One of the critical issues in face anti-spoofing and image classification in general is the generalization across different acquisition devices. A Leave One Camera Out (LOCO) protocol is designed to study the sensor interoperability of the face PAD methods. In each iteration, the real and the attack videos recorded with five smartphones are used to train and tune the countermeasure model. Then, the generalization of the method is assessed using the videos recorded with the remaining smartphone.

*4) Protocol IV:* In the last and most challenging scenario, the previous three protocols are combined to simulate the real-world operational conditions. To be more specific, the generalization abilities of the face PAD methods are evaluated simultaneously across previously unseen illumination conditions, background scenes, PAIs and input sensors, see Table II.

In all these protocols, the 55 subjects were divided into three subject-disjoint subsets for training, development and testing (20, 15 and 20, respectively). Tables II gives a detailed information about the video recordings used in the train, development and test sets of each protocol.

## IV. Experiments

The experimental results of the baseline method under the different protocols are presented and discussed in this section. For the performance evaluation, we selected the recently standardized ISO/IEC 30107-3 metrics [8]: Attack Presentation Classification Error Rate (APCER) and Bona Fide Presentation Classification Error Rate (BPCER). In principle, these two metrics correspond to the False acceptance Rate (FAR) and False Rejection Rate (FRR) commonly used in the PAD related literature. However, unlike the FAR and FRR, the APCER and the BPCER take the attack potential into account in terms of an attacker's expertise, resources and motivation in the ”worst case scenario”. To be more specific, the APCER is computed separately for each PAI (e.g. print or display) and the overall PAD performance corresponds to the attack with highest APCER, i.e. the most successful PAI. This indicates how easy a biometric system is to fool on average by exploiting its vulnerability (if there is any).

Since both the APCER and the BPCER depend on the decision threshold, the development set operates as a separate validation set for fine tuning the system parameters and estimating the threshold value to be used on the test set. To summarize the overall system performance in a single value, the Average Classification Error Rate (ACER) is used which is the average of the APCER and the BPCER at the decision threshold defined by the Equal Error Rate (EER) on the development set.

As a baseline face PAD method, we chose the color texture based method [2] as it has shown promising generalization abilities. In this method, the texture features are extracted from the color images instead of the gray-scale representation that has been more commonly used in face PAD. The color reproduction (gamut) of different PAIs, e.g. prints, displays and masks, is limited compared to genuine faces. Gamut mapping functions are typically required in order to preserve color perception properties across different output devices, which can alter the (color) texture of the original image. In general, the gamut mapping algorithms focus on preserving the spatially local luminance variations in the original images at the cost of the chrominance information because the human eye is more sensitive to luminance than to chroma. The camera used for capturing the targeted face sample will also lead to imperfect color reproduction compared to the legitimate sample. Furthermore, other disparities in facial texture, including printing defects, video artifacts, noise signatures of display devices and moiré effects, should be more evident in the original color images compared to gray-scale images. Thus, the color texture analysis provides enhanced discrimination between the real and the attack samples.

In this paper, for each frame, the face region is detected, cropped and normalized into $64 \times 64$ pixel images. Since the studies conducted in [2], [3] depict that the color texture information extracted from both the HSV and YCbCr color spaces gives the best results compared to the RGB, or the gray-scale images, the uniform $\text{LBP}_{8,1}^{u2}$ (i.e. neighbors=8 and radius=1) features are extracted from each channel of the HSV and YCbCr image representations. Then, the resulting features are concatenated and fed into a Softmax classifier with a cross-entropy loss function.

### A. Protocol I: Effect of the illumination variation

To study the effect of the illumination variation on the robustness of the face PAD method, we train and tune the countermeasure model using the video recordings taken in Session 1 and Session 2, then evaluate its performance on the videos taken in the third session. Table III shows the effect of this variation on the color LBP based method. As we can see, using different sessions to train and evaluate the countermeasure model results in a degraded performance compared to the results of the countermeasure model trained with video recordings from the same session as the evaluation set (Session 3). The performance degradation from 2.7% to 13.5% indicates that the illumination variation can indeed pose a big issue for the face PAD methods, especially for the texture based methods in terms of BPCER.

### B. Protocol II: Effect of the PAI variation

The effect of the PAI variation on the generalization performance is investigated by selecting the spoofing attacks created with different PAIs in the train and test conditions. In the train set, we used the print and the video-replay attacks created with Printer 1 and Display 1. Then, for the evaluation, we used the attacks created with Printer 2 and Display 2. To show how much this variation can affect the generalization performance, we have also reported the results without any PAI variation (i.e. the attacks in the training, development and test sets are created using Printer 2 and Display 2). The results reported in Table IV show that the variation in the PAI decreases the performance of the baseline method from 7.2% to 14.2% in terms of ACER. It is worth highlighting that the baseline method is able to deal with the PAI variation much better in the case of video-replay attacks than print attacks as the ACER increases from 7.2% to 9.2% and from 6.1% to 14.2%, respectively. It is not surprising to notice that illumination variation increases specifically the BPCER, while PAI variation has more significant effect on the APCER.

### C. Protocol III: Effect of the camera device variation

To study the effect of the camera device variation, we compared the results obtained with the LOCO protocol to the results obtained without any camera device variation (i.e. the videos in the training, development and test sets are recorded using the same mobile device). The results are presented in Table V and Table VI. In addition to reporting the performance on each mobile phone separately, the average and the standard deviation over all folds are also computed in order to summarize the results. From Table V and Table VI, we can clearly see that sensor interoperability is another major issue in face PAD that needs further attention.

### D. Protocol IV: Effect of illumination, PAI and camera device variations

This part demonstrates the combined effect of the illumination, PAI and camera device variations on the generalization performance, which gives us a better idea about the robustness of the developed face PAD methods in more realistic mobile authentication scenarios. The results reported in Table VII show that combining the three variations causes significant degradation in performance. Although the color texture based method shows relatively good generalization abilities in the previous experiments (in which only one variation was taken into account), it fails completely to deal with the different covariates at the same time, especially in the case of some mobile phones.

## V. Conclusions

In this paper, we introduced a new mobile face presentation attack detection database. It consists of real access and attack videos corresponding to 55 subjects. The videos were recorded using six different mobile devices in three different illumination conditions and background scenes. For the spoofing attacks, we considered two types of attacks: print attacks and video-replay attacks. Both of these attacks were created using two presentation attack instruments (two printers and two displays). To evaluate the robustness of the developed face PAD methods, we designed four protocols. These protocols study the effect of the environmental conditions (namely illumination and background scene), PAI and camera device variations on the generalization abilities. The results of a face PAD method based on color texture analysis were reported as a baseline. We invite the research community to consider this new database for the evaluation of new PAD methods.

## References

[1] A. Anjos, J. Komulainen, S. Marcel, A. Hadid, and M. Pietikäinen. Face anti-spoofing: Visual approach. In S. Marcel, M. Nixon, and S. Z.Li, editors, *Handbook of Biometric Anti-Spoofing*, pages 65–82. Springer-Verlag, 2014.

[2] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face anti-spoofing based on color texture analysis. In *IEEE International Conference on Image Processing (ICIP)*, 2015.

[3] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11(8):1818–1830, 2016.

[4] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7, 2012.

[5] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel. The replay-mobile face presentation-attack database. In *International Conference of the Biometrics Special Interests Group (BIOSIG)*, 2016.

[6] T. de Freitas Pereira, A. Anjos, J. De Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *International Conference on Biometrics (ICB)*, pages 1–8, 2013.

TABLE III: The performance of the of color LBP method under different illumination conditions.

| Train | Test | Dev | Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | EER | Video-replay | | | Print | | | Overall | | |
| | | | APCER | BPCER | ACER | APCER | BPCER | ACER | APCER | BPCER | ACER |
| Session 3 | Session 3 | 2.9 | 2.6 | 2.4 | 2.5 | 2.9 | 2.4 | 2.7 | 2.9 | 2.4 | 2.7 |
| Session 1, 2 | Session 3 | 4.7 | 5.8 | 21.3 | 13.5 | 1.7 | 21.3 | 11.5 | 5.8 | 21.3 | 13.5 |

TABLE IV: The performance of the color LBP method under PAI variation.

| Train | Test | Dev | Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | EER | Video-replay | | | Print | | | Overall | | |
| | | | APCER | BPCER | ACER | APCER | BPCER | ACER | APCER | BPCER | ACER |
| Pr 2, D 2 | Pr 2, D 2 | 4.9 | 10.3 | 4.0 | 7.2 | 8.2 | 4.0 | 6.1 | 10.3 | 4.0 | 7.2 |
| Pr 1, D 1 | Pr 2, D 2 | 4.3 | 11.4 | 7.0 | 9.2 | 21.5 | 7.0 | 14.2 | 21.5 | 7.0 | 14.2 |

TABLE V: The performance of the color LBP method without camera device variation.

| Train | Test | Dev | Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | EER | Video-replay | | | Print | | | Overall | | |
| | | | APCER | BPCER | ACER | APCER | BPCER | ACER | APCER | BPCER | ACER |
| P=1 | P= 1 | 4.8 | 3.9 | 9.2 | 6.5 | 2.7 | 9.2 | 6.0 | 3.9 | 9.2 | 6.5 |
| P=2 | P= 2 | 3.8 | 5.0 | 3.8 | 4.4 | 8.6 | 3.8 | 6.2 | 8.6 | 3.8 | 6.2 |
| P=3 | P= 3 | 1.8 | 4.7 | 9.0 | 6.8 | 0.1 | 9.0 | 4.6 | 4.7 | 9.0 | 6.8 |
| P=4 | P= 4 | 7.7 | 8.9 | 7.8 | 8.4 | 10.3 | 7.8 | 9.1 | 10.3 | 7.8 | 9.1 |
| P=5 | P= 5 | 4.2 | 7.8 | 3.5 | 5.6 | 8.9 | 3.5 | 6.2 | 8.9 | 3.5 | 6.2 |
| P=6 | P= 6 | 1.9 | 4.2 | 1.8 | 3.0 | 3.8 | 1.8 | 2.8 | 4.2 | 1.8 | 3.0 |
| Avg± std | | 4.0 ± 2.2 | 5.7± 2.1 | 5.9± 3.2 | 5.8± 1.9 | 5.8 ± 4.1 | 5.9 ± 3.2 | 5.8 ± 2.1 | 6.8 ± 2.8 | 5.9 ± 3.2 | 6.3 ± 1.9 |

TABLE VI: The performance of the color LBP method under camera device variation.

| Train | Test | Dev | Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | EER | Video-replay | | | Print | | | Overall | | |
| | | | APCER | BPCER | ACER | APCER | BPCER | ACER | APCER | BPCER | ACER |
| P={2,3,4,5,6} | P= 1 | 5.4 | 5.6 | 19.1 | 12.3 | 2.9 | 19.1 | 11.0 | 5.6 | 19.1 | 12.3 |
| P={1,3,4,5,6} | P= 2 | 5.2 | 7.0 | 6.2 | 6.6 | 17.6 | 6.2 | 11.9 | 17.6 | 6.2 | 11.9 |
| P={1,2,4,5,6} | P= 3 | 5.0 | 7.2 | 19.9 | 13.5 | 3.7 | 19.9 | 11.8 | 7.2 | 19.9 | 13.5 |
| P={1,2,3,5,6} | P= 4 | 4.3 | 15.1 | 5.8 | 10.4 | 12.8 | 5.8 | 9.3 | 15.1 | 5.8 | 10.4 |
| P={1,2,3,4,6} | P= 5 | 4.8 | 6.3 | 4.9 | 5.6 | 8.2 | 4.9 | 6.6 | 8.2 | 4.9 | 6.6 |
| P={1,2,3,4,5} | P= 6 | 4.9 | 15.8 | 10.4 | 13.1 | 25.2 | 10.4 | 17.8 | 25.2 | 10.4 | 17.8 |
| Avg± std | | 4.9 ±0.4 | 9.5 ±4.6 | 11.0 ±6.8 | 10.3 ±3.4 | 11.7 ±8.6 | 11.0± 6.8 | 11.4± 3.7 | 13.1± 7.6 | 11.0± 6.8 | 12.1 ±3.7 |

TABLE VII: The performance of the color LBP method under illumination, PAI and camera variations

| Train | Test | Dev | Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | EER | Video-replay | | | Print | | | Overall | | |
| | | | APCER | BPCER | ACER | APCER | BPCER | ACER | APCER | BPCER | ACER |
| P={2,3,4,5,6} | P= 1 | 5.0 | 13.4 | 18.7 | 16.1 | 12.0 | 18.7 | 15.4 | 13.4 | 18.7 | 16.1 |
| P={1,3,4,5,6} | P= 2 | 6.0 | 23.0 | 16.4 | 19.7 | 19.6 | 16.4 | 18.0 | 23.0 | 16.4 | 19.7 |
| P={1,2,4,5,6} | P= 3 | 5.6 | 5.8 | 38.4 | 22.1 | 0.0 | 38.4 | 19.2 | 5.8 | 38.4 | 22.1 |
| P={1,2,3,5,6} | P= 4 | 5.2 | 42.7 | 36.4 | 39.6 | 12.4 | 36.4 | 24.4 | 42.7 | 36.4 | 39.6 |
| P={1,2,3,4,6} | P= 5 | 5.2 | 10.0 | 21.1 | 15.5 | 2.9 | 21.1 | 12.0 | 10.0 | 21.1 | 15.5 |
| P={1,2,3,4,5} | P=6 | 4.4 | 64.6 | 0.3 | 32.5 | 100.0 | 0.3 | 50.2 | 100.0 | 0.3 | 50.2 |
| Avg± std | | 5.2 ±0.6 | 26.6± 22.8 | 21.9± 14.1 | 24.2± 9.7 | 24.5± 37.7 | 21.9± 14.1 | 23.2 ± 13.8 | 32.5± 35.6 | 21.9± 14.1 | 27.2± 14.3 |

[7] J. Galbally, S. Marcel, and J. Fiérrez. Biometric antispoofing methods: A survey in face recognition. *IEEE Access*, 2:1530–1552, 2014.

[8] ISO/IEC JTC 1/SC 37 Biometrics. Information technology – Biometric presentation attack detection – Part 1: Framework. International Organization for Standardization, 2016. https://www.iso.org/obp/ui/#iso:std:iso-iec:30107:-1:ed-1:v1:en.

[9] Y. Li, Y. Li, K. Xu, Q. Yan, and R. Deng. Empirical study of face authentication systems under osnfd attacks. *IEEE Transactions on Dependable and Secure Computing*, 2016.

[10] K. Patel, H. Han, and A. K. Jain. Cross-database face antispoofing with robust feature representation. In *Chinese Conference on Biometric Recognition (CCBR)*, pages 611–619, 2016.

[11] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *European Conference on Computer vision (ECCV)*, pages 504–517, 2010.

[12] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li. Face liveness detection using 3D structure recovered from a single camera. In *International Conference on Biometrics (ICB)*, 2013.

[13] D. Wen, H. Han, and A. Jain. Face spoof detection with image distortion analysis. *Transactions on Information Forensics and Security*, 10(4):746–761, 2015.

[14] D. Yambay, J. S. Doyle, K. W. Bowyer, A. Czajka, and S. Schuckers. Livdet-iris 2013 - iris liveness detection competition 2013. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8, 2014.

[15] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In *International Conference on Biometrics (ICB)*, pages 26–31, 2012.