IDIAP RESEARCH REPORT

SMARTPHONE MULTI-MODAL BIOMETRIC AUTHENTICATION: DATABASE AND EVALUATION

Ramachandra Raghavendra Martin Stokkenes
Amir Mohammadi Sushma Venkatesh
Kiran B. Raja Pankaj Wasnik Eric Poiret
Sébastien Marcel Christoph Busch

Idiap-RR-17-2020

JULY 2020

Smartphone Multi-modal Biometric Authentication: Database and Evaluation

Raghavendra Ramachandra, Martin Stokkenes, Amir Mohammadi, Sushma Venkatesh, Kiran Raja, Pankaj Wasnik, Eric Poiret, Sébastien Marcel, Christoph Busch

***Abstract*—Biometric-based verification is widely employed on the smartphones for various applications, including financial transactions. In this work, we present a new multimodal biometric dataset (face, voice, and periocular) acquired using a smartphone. The new dataset is comprised of 150 subjects that are captured in six different sessions reflecting real-life scenarios of smartphone assisted authentication. One of the unique features of this dataset is that it is collected in four different geographic locations representing a diverse population and ethnicity. Additionally, we also present a multimodal Presentation Attack (PA) or spoofing dataset using a low-cost Presentation Attack Instrument (PAI) such as print and electronic display attacks. The novel acquisition protocols and the diversity of the data subjects collected from different geographic locations will allow developing a novel algorithm for either unimodal or multimodal biometrics.Further, we also report the performance evaluation of the baseline biometric verification and Presentation Attack Detection (PAD) on the newly collected dataset.**

***Index Terms*—Biometrics,Smartphone, Spoofing, Presentation attacks, Database.**

I. Introduction

Secure and reliable access control using biometrics are deployed in various applications that include border control, smartphone unlocking, banking transactions, financial services, attendance system and etc. The use of biometrics in access control applications not only improves the reliability but also improves the user experience by authenticating the user based on who they are. Thus, the user is neither required to remember passcode nor need to possess any smart cards to gain access to the control process. Based on ISO/IEC 2382-37, the concept of biometrics is defined as: automated recognition of individuals based on their behavioural and biological characteristics [1]. The biometric characteristics can be either a physical (face, iris, fingerprint, etc.) or a behavioral (keystroke, gait, etc.) trait that can be used to recognize the data subject.

Evolution of the biometric technology has resulted in several consumer applications including smartphone biometrics. According to the Acuity Intelligence Forecast [2], smartphone-based biometrics can surpass a revenue of 50.6 billion dollars by 2022 that also includes the revenue of biometric applications for financial transactions. Further, it is also estimated from TrendForce [3] that, there will be 1.4 billion smartphone devices as of 2022. These factors indicate the evolution of different types of biometric-based applications for smartphones including banking applications in the lines of already available services from different vendors like Apple Pay, Samsung pay, Google Wallet. The majority of commercial smartphones available today provide only uni-modal biometrics and the most popular biometrics that are used by smartphone vendors include fingerprint, iris and face.

Even though the utility of biometrics on smartphones has enabled several advantages, there exist several challenges in real-life applications to take the full advantage of the biometrics authentication process on the consumer smartphone. Among the many challenges, vulnerability towards attacks and interoperability challenges are key problems that limit reliable and secure applications of smartphones for financial services. The vulnerability of smartphones to both direct (Presentation Attacks, aka., spoofing attacks) and indirect attacks are well exemplified in the literature [4] [5]. Further, it is also demonstrated in [6] that, the biometric template can be retrieved from the smartphone hardware chip, which further indicates another vulnerability. Recent works based on the master prints [7] demonstrates the vulnerability of the fingerprint recognition system itself, irrespective of the manufacturer. The second key challenge is the interoperability, as smartphone system uses proprietary biometric solutions, it is very challenging to use them with the traditional biometric systems. This limits the user by locking oneself to the particular smartphone to enable applications. These factors motivated the research towards smartphone biometrics that can be independent of the devices. To this extent, several attempts have been made to develop both uni-modal and multimodal biometric solution. The most popular biometric characteristics investigated include face [8], visible iris [9], soft-biometrics [10], finger photo [11].

The crucial aspect hindering the advances in the area of smartphone biometrics is the availability of suitable biometric data to benchmark the recognition performance of the newly developed algorithms and the reproducible evaluation. There exists a limited number of smartphone dataset which are publicly available and which are particularly addresssing biometric characteristics such as: face [12], fingerphoto [11], iris [9], soft-biometrics [10] and multi-modality [13]. However, the collection of biometric data is a time and resource consuming tedious process that demands additional efforts in selecting the capture device, design of data collection protocols, post-

Raghavendra Ramachandra, Martin Stokkenes, Sushma Venkatesh, Kiran Raja, Pankaj Wasnik and Christoph Busch are with Norwegian Biometrics Laboratory, Norwegian University of Science and Technology (NTNU), Norway. e-mail: (raghavendra.ramachandra@ntnu.no).

Amir Mohammadi and Sebastien Marcel are with Idiap Research Institute, Martigny, Switzerland. e-mail: ({amir.mohammadi,marcel}@idiap.ch).

Eric Poiret is with IDEMIA, France.

Manuscript received september 19, 2019; revised August 26, 2015.

processing of captured data, annotation of captured data, data anonymization and the collection itself. Further, one also needs to consider the legal regulations to obtain the data subject's consent and also to respect the data protection regulations.

This paper presents a recently collected smartphone based multimodal biometric dataset within the framework of the SWAN project[14]. The SWAN Multimodal Biometric Dataset (SWAN-MBD) was collected between April 2016 till December 2017 with a collective effort from three different institutions: The Norwegian University of Science and Technology (NTNU), Norway, Idiap Research Institute, Switzerland and IDEMIA in France and India. The data collection is carried out in six different sessions with a time gap between sessions of 1 week to 3 weeks. The capture environment included both indoor and outdoor scenarios in assisted/supervised and unsupervised capture settings. The data capture includes both self and assisted capture processes replicating the real-life applications such as banking transactions. Three different biometric characteristics such as face, periocular and voice corresponding to 150 subjects were captured and the data subjects are living in 4 different geographic locations such as Norway, Switzerland, France and India. Further, we also present a new SWAN Multimodal Presentation Attack Dataset (SWAN-MPAD) using two different types of Presentation Attack Instruments (PAIs) such as high-quality print and display attack (using iPhone 6 and iPad PRO) for the face and periocular characteristics. In the case of voice biometrics, two different quality loudspeakers were used to record the replay attacks on iPhone 6. Being a new smartphone multimodal biometric dataset with presentation attack samples, it will allow one to develop and benchmark both verification and Presentation Attack Detection (PAD) algorithms.

The following are the main contribution of this paper:

- New multimodal dataset collected using smartphone in 6 different session from 150 subjects.
- New multimodal presentation attack dataset collected using two different PAI on face, periocular, and voice biometrics.
- Performance evaluation protocols to benchmark the accuracy of both verification and PAD algorithms.
- Quantitative result of the baseline algorithms are reported using ISO/IEC SC 37 metrics on both verification and PAD.

The rest of the paper is organised as follows: Section II presents the related wok on the available multimodal biometric datasets collected using smartphone, Section III presents the SWAN multimodal biometric and presentation attack dataset, Section IV presents the evaluation protocols to benchmark the algorithms, Section V discuss the different baseline systems used in benchmark the performance on SWAN multimodal biometric dataset. Section VI discuss the quantitative results of the baseline systems, Section VII discuss the potential use of newly collected dataset for research and development tasks, Section VIII provides the information on data distribution and Section IX draws the conclusion.

II. Related Work

In this section, we discuss the publicly available smartphone based multimodal datasets. There are limited smartphones based multimodal biometric databases currently available for researchers. The majority of these available datasets are based on the two modalities that can be captured together, for example, face (talking) and voice, iris and periocular. Table I presents an overview of the different smartphone based multimodal datasets that are publicly available.

The **BioSecure** dataset (DS-3) [20] is one of the earlier and large scale publicly available datasets (available upon license fee payment). The BioSecure dataset is comprised of four different faces, voice, signature and fingerprint collected from 713 subjects in 2 sessions. Only face and voice are collected using the mobile device Samsung Q1.

The **MOBIO** database [12] is comprised of 153 subjects from which the biometric characteristics (face and voice) are collected using Nokia N93i and MacBook laptop. The complete dataset is collected in 12 sessions by capturing the face (talking) together with voice. Voice samples are collected based on both pre-defined text and free text, and the face biometrics is recorded while the subject is talking.

The **CSIP** database [15] is comprising of 50 subjects with iris and periocular biometric characteristics captured using four different smartphone devices. The entire dataset is collected with different backgrounds to reflect the real-life scenario.

The **FTV** dataset [16] consists of face, teeth and voice biometric characteristics captured from 50 data subjects using a smartphone HP iPAQ rw6100. The face and teeth samples are collected using the smartphone camera while the microphone of smartphone used to collect the voice samples.

The **MobBIO** dataset [17] is based on the three different biometric characteristics such as the face, voice and periocular using a tablet Asus Transformer Pad TF 300T. Voice samples are collected using a microphone of Asus Transformer Pad TF 300T in which data subject was asked to readout 16 sentences in Portuguese. The face and periocular samples are collected using the 8MP camera from Asus Transformer Pad TF 300T. This dataset is comprised of 105 data subjects collected in two different lighting conditions.

The **UMDAA** dataset [18] is collected using the Nexus 5 from 48 data subjects. This database has a collection of both physical and behavioral biometric characteristics. The data collection sensors include the front-facing camera, touchscreen, gyroscope, magnetometer, light sensor, GPS, Bluetooth, accelerometer, WiFi, proximity sensor, temperature sensor and pressure sensor. The entire dataset is collected for two months and provides face and other related behavior patterns suitable for continuous authentication.

The **MobiBits** dataset [19] consists of five different biometric characteristics namely voice, face, iris, hand and signature that are collected from 55 data subjects. Three different smartphones are used to collect the biometric characteristics such as Huawei Mate smartphone is used to collect signature and periocular samples, CAT S60 smartphone is used to collect hand and face samples, Huawei P9 Lite smartphone is used to collect voice samples.

TABLE I: Publicly available smartphone based multimodal datasets

Dataset	Year	Devices	No. of subjects	Biometric	Availability
MOBIO [12]	2012	Nokia N93i Mac-book	152	Face, Voice	Free
CSIP [15]	2015	Sony Xperia Apple IPhone 4	50	Iris, Periocular	Free
FTV [16]	2010	HP iPAQ rw6100	50	Face, Teeth, Voice	Free
MobBIO [17]	2014	ASUS PAD TF 300	105	Voice, Face, periocular	Free
UMDAA [18]	2016	Nexus 5	48	Face, Behavior patterns	Free
MobiBits [19]	2018	Huawei Mate S Huawei P9 Lite CAT S60	53	Signature, Voice, Face, Periocular, Hand	Free
BioSecure-DS3 [20]	2010	Samsung Q1 Philips SP900NC Webcam HP iPAQ hx2790 PDA	713	Voice, Signature, Face, Fingerprint	Paid
SWAN Dataset	2019	iPhone 6S	150	Face, Periocular, Multilingual Voice Presentation Attack dataset	Free

A. Features of the SWAN Multimodal Biometric Dataset

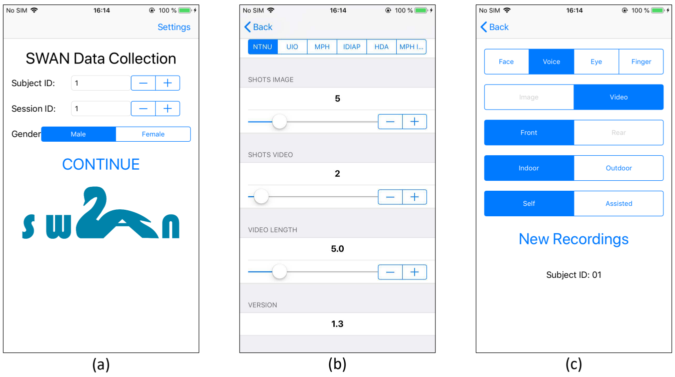
SWAN Multimodal Biometric Dataset (SWAN-MBD) is collected to complement the existing datasets and the data collection protocols are designed to meet the real-life scenario such as banking transactions. The following are the main features of the newly introduced SWAN Multimodal Biometric Dataset:

- The data collection protocol is designed to capture data in 6 different sessions. Each session reflects both time variation and capturing conditions (outdoor and indoor).
- The data collection application is developed to make data collection consistent with ease of use in regards to the user interaction such that self-capture is facilitated. This is the unique feature of this dataset in which the data is self-captured by the participants.
- The data collection is carried out in four different geographic locations such as India, Norway, France and Switzerland with multiple ethnicity representation.
- Three different biometric characteristics: face, periocular and voice are captured from 150 data subjects in all 6 different sessions. The voice samples are captured based on the question and answers which is text-dependent. The voice samples are collected in four different languages: English, Norwegian, French and Hindi.
- The whole dataset is collected using iPhone 6 smartphone and iPad-PRO.
- In addition, we also present a new SWAN Multimodal Presentation Attack dataset (SWAN-MPAD) for all three biometric characteristics (face, periocular and voice).

III. SWAN Multimodal Biometric Dataset

A. Database Acquisition

To facilitate the data collection at different locations, a smartphone application is developed for the iOS platform (version 9) that can be installed in the data capture devices (both iPhone and iPad Pro). This application has a Graphical User Interface that allows the data collection moderator to select session number, biometric characteristics, location ID, subject ID and other relevant information for data collection. Figure 1 shows the GUI of SWAN data collection application while Figure 2 shows the example images during biometric data collection. Thus, the application is designed to make sure the data collection process can be easily carried out such that data subjects can use it seamlessly during the self-capture protocol. The data collection process is broadly divided into two phases as explained below.



SWAN Data Collection

Subject ID: 1

Session ID: 1

Gender: Male

SHOTS IMAGE 5

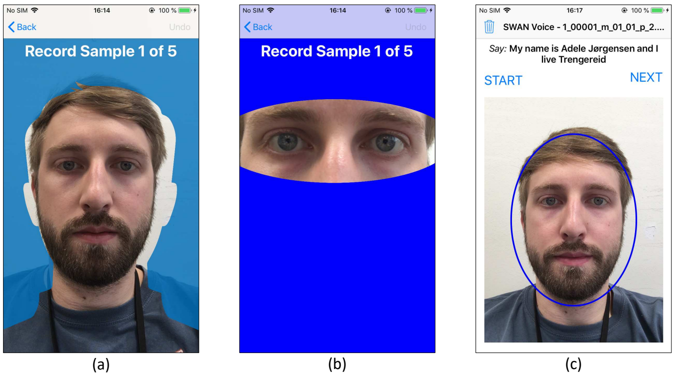
SHOTS VIDEO 2

VIDEO LENGTH 5.0

VERSION 1.3

Subject ID: 01

Fig. 1: Screen shot from SWAN multimodal biometric data capture application



SWAN Voice - 1_00001_m_01_01_p_2....

Say: My name is Adele Jørgensen and I live Trengereid

Fig. 2: Illustration of SWAN multimodal biometric data capture application during (a) Face capture (b) Eye region capture (c) Voice and talking face capture

B. SWAN Multimodal Biometric Dataset

The SWAN multimodal biometric dataset was jointly collected within SWAN project by partners at four different geographic locations: Norway, Switzerland, France and India. The dataset is comprised of 150 subjects such that 50 data subjects collected in Norway, 50 data subjects are collected in Switzerland, 5 data subjects in France and 45 data subjects in India. The age distribution of the Data subjects are between 20 to 60 years. The dataset was collected using Apple iPhone6S and Apple iPad Pro (12.9inch iPad Pro). Three biometric modalities are collected: (1) Face: both Ultra HD images and HD (including slow motion) video recordings of faces from data subjects (2) Voice: HD audio-visual recordings of talking faces from data subjects (3) Eye: both Ultra HD images and HD (including slow motion) video recordings of eyes from data subjects. The whole dataset is collected in 6 different sessions such that, **Session 1** is captured in an indoor environment with uniform illumination of the face, quiet environment reflecting high quality supervised enrolment. **Session 2** is captured in indoor with natural illumination on the face and semi-quite environment. **Session 3** is captured in outdoor with uncontrolled illumination, natural noise environment. **Session 4** is captured in indoor with uncontrolled illumination e.g., side light from windows, natural noise environment. **Session 5** is captured in indoor with uncontrolled illumination, natural noise environment in crowded place. **Session 6** is captured in indoor with uncontrolled illumination e.g., side light from windows, natural noise environment. The average time duration between sessions varies from 1 week to 3 weeks. The multimodal biometric samples are captured in both assisted capture mode (where data subjects are assisted during capturing) and self-capture mode (here data subject control the capture process on their own). Multimodal biometrics are collected using both rear and front cameras from iPhone and iPad Pro. The iPhone 6S is used as the primary device to collect the dataset, while the iPad Pro is used to collect the data only in session-1 to capture a high-quality image that can be used to generate a Presentation Attack Instrument.

Table II indicates the data collection protocol and the sample collection (both images and videos) for the facial biometric characteristic. The biometric face capture indicated in Table II corresponds to one data subject. iPhone 6S is used to capture the face data in all six session and iPad Pro is used only in session 1. During each acquisition of session-1, the face data is captured in both assisted and self-captured mode using rear and front camera from iPhone6S and iPad Pro respectively. The data collected using the iPad Pro is used to generate presentation attacks while data collected from iPhone6S is used to perform the biometric verification. Thus, in total there are: 150× Subjects ×20 images = 3000 image and 150× Subjects ×18 videos = 2700 videos. Table III presents the data statistics corresponding to eye region biometric characteristic, which is in both self-capture and assisted mode. Since the goal of the eye region data collection is to get high quality images that can be used for both periocular and visible iris recognition,

TABLE II: SWAN multimodal dataset: Face biometrics data collection

Modality	Session	Capture Mode	Device /camera	Data capture per data subject
Face	S1	Assisted	iPhone6S/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 240fps, 5s, MP4)
			iPad Pro/ Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 120fps, 5s, MP4)
		Self-Capture	iPhone6S/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
			iPad Pro/ Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
	S2	Self-Capture	iPhone6S/Front	2 videos (1280x720, 30fps, 5s, MP4)
	S3	Self-Capture	iPhone6S/Front	2 videos (1280x720, 30fps, 5s, MP4)
	S4	Self-Capture	iPhone6S/Front	2 videos (1280x720, 30fps, 5s, MP4)
	S5	Self-Capture	iPhone6S/Front	2 videos (1280x720, 30fps, 5s, MP4)
	S6	Self-Capture	iPhone6S/Front	2 videos (1280x720, 30fps, 5s, MP4)

TABLE III: SWAN multimodal dataset: Eye region biometrics data collection

Modality	Session	Capture Mode	Device /camera	Data capture per data subject
Eye	S1	Assisted	iPhone6S/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 240fps, 5s, MP4)
			iPad Pro/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 120fps, 5s, MP4)
		Self-Capture	iPhone6S/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
			iPad Pro/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
	S2	Assisted	iPhone6S/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 240fps, 5s, MP4)
		Self-Capture	iPhone6S/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
	S3	Assisted	iPhone6S/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 240fps, 5s, MP4)
		Self-Capture	iPhone6S/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
	S4	Assisted	iPhone6S/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 240fps, 5s, MP4)
		Self-Capture	iPhone6S/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
	S5	Assisted	iPhone6S/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 240fps, 5s, MP4)
		Self-Capture	iPhone6S/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)
	S6	Assisted	iPhone6S/Rear	5 Images (4032x3024, PNG) 2 videos (1280x720, 240fps, 5s, MP4)
		Self-Capture	iPhone6S/Front	5 Images (2576x1932, PNG) 2 videos (1280x720, 30fps, 5s, MP4)

TABLE IV: SWAN multimodal dataset: Voice and Talking face biometrics data collection

Modality	Session	Capture Mode	Device/Camera	Data capture per data subject
Voice and Talking faces	s1-6	Self-capture	iPhone6S/Front	8 videos: 4 videos in English and 4 videos in native language (1280x720, 30fps, variable length, MP4)

TABLE V: SWAN Presentation attack data collection details

Modality	PA name	PA source selection			PAI	Biometric capture sensor
		Session	Sensor	type		
Face	PA.F.1	1	iPhone 6s back camera	photos	Epson Expression Photo XP-860 Epson Photo Paper Glossy, PN: S041271	iPhone 6s front camera (video)
	PA.F.5	1	iPhone 6s front camera	videos	iPhone 6s display	iPhone 6s front camera (video)
	PA.F.6	1	iPad-Pro front camera	videos	iPad-Pro display	iPhone 6s front camera (video)
Eye	PA.EI.1	1	iPhone 6s back camera	photos	Epson Expression Photo XP-860 Epson Photo Paper Glossy, PN: S041271	iPhone 6s front camera (video)
	PA.EI.4	1	iPad-Pro front camera	photos	iPad-Pro display	iPhone 6s front camera (video)
	PA.EI.5	1	iPhone 6s front camera	videos	iPhone 6s display	iPhone 6s front camera (video)
Voice	PA.V.4	1	iPad-Pro microphone	audio	Logitech high quality loudspeaker	iPhone 6s microphone (audio)
	PA.V.7	1	iPhone 6s microphone	audio	iPhone 6s speakers	iPhone 6s microphone (audio)

we collected the eye region dataset in both capture modes across all 6 sessions. Therefore the assisted mode is captured using the rear camera of iPhone 6S with 12mega pixels that provides good quality images for visible iris recognition. However, the images collected from the self-capture process using the frontal camera can be used to develop the periocular verification systems. Similar to face capture process, the iPad Pro is used only in session 1 to capture a good quality eye region image that is used to generate a presentation attack instrument, to be used against the periocular biometric system. The whole dataset consists of 150× Subjects ×70 images = 10500 image and 150× Subjects ×28 videos = 4200 videos.

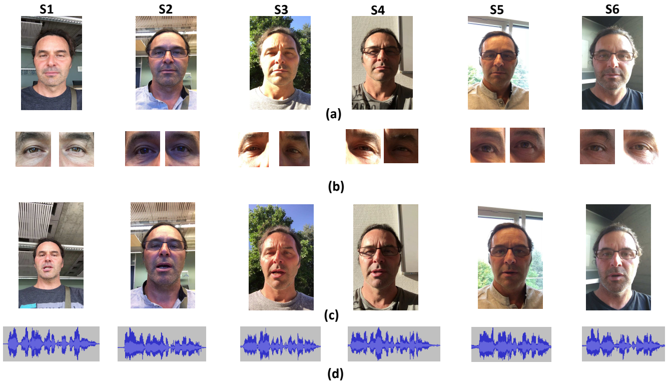


Fig. 3: Illustration of SWAN multimodal biometric dataset images (a) Face capture (b) Eye region capture (c) Talking face capture (one only frame is included for the simplicity) (d) voice sample

Table IV indicates the statistics and protocol of the voice and talking face data collection. To cover both text independent and text dependent modes, the audio recordings (actually audio-video recordings) are captured with the data subjects pronouncing these 4 utterances in English followed by these 4 utterances in a national language depending on the site (Norwegian, French and Hindi). The four sentences include: *Sentence #1*: ”My name is `FAKE_FIRSTNAME FAKE_NAME` and I live `FAKE_ADDRESS`” the address is a short string that is limited to the street name. Thus no street number, no zip code, no city and no country information is recorded. *Sentence #2*: ”My account number is `FAKE_ACCOUNTNUMBER`”. The account numbers are presented by groups of digits (eg. ”5354” ”8745”). The data subject is free to pronounce the groups of digits the way he/she wants (digits by digits, as a single number or as a combination). *Sentence #3:* ”The limit of my card is 5000 euros”. *Sentence #4:* ”The code is 9 8 7 6 5 4 3 2 1 0”. The 10 digits are presented one by one and the data subjects were asked to pronounce the digits separately. The Audio-visual data with voice and talking faces are collected in the self-capture mode using the frontal camera of iPhone6S in all 6 sessions. Thus, 150× Subjects ×48 videos = 7200 videos corresponding to audio-visual data. Figure 3 illustrates the example images from SWAN multimodal dataset collected in all six session indicating an inter-session variability.

C. SWAN-Presentation Attack Dataset

The SWAN presentation attack dataset is comprised of three different types of presentation attacks that are generated for three different biometric characteristics, namely: face, eye and voice. The presentation attack (PA) database generation generally requires obtaining biometric samples ”PA source selection”, generating artefacts, and presenting attack artefacts to the biometric capture sensor. Session 1 recordings from SWAN multibiometric dataset is used to generate the presentation attack dataset. The artefact generation is carried out using five different Presentation Attack Instruments (PAI) such as: (1)High quality photo on paper generated using the photo printer (face & Eye). The print artefacts on face and eye are generated using Epson Expression Photo XP-860 with high quality paper Epson Photo Paper Glossy (Product Number: S041271; Basis Weight: 52 lb.(200g/m^2); Size: A4; Thickness: 8.1 mil.). (2) Electronic display of artefacts using iPad-Pro (face & Eye) (3) Electronic display of artefacts using iPhone 6S (face & Eye) (4) Logitech high quality loudspeaker (Voice) (5) iPad-Pro loudspeaker (Voice). We have used these PAI by considering the cost for generation versus attack potential and thus selected PAIs, which are of low-cost (reasonably) and at the same time indicate a high vulnerability for the biometric system under attack.

Table V presents the PAD data collection procedure with presentation attack source, PAI and biometric capture sensor. All the presentation attack data for face and eye are collected using a frontal camera of iPhone 6S and all recordings were videos and at least 5 seconds long for each PA and in the *.mp4* format. Both the biometric capture device (iPhone 6s) and the PAI (paper or display) is mounted on stands. In the case of voice samples, the audio files collected using iPhone6S with the same compression specifications as bona fide samples.

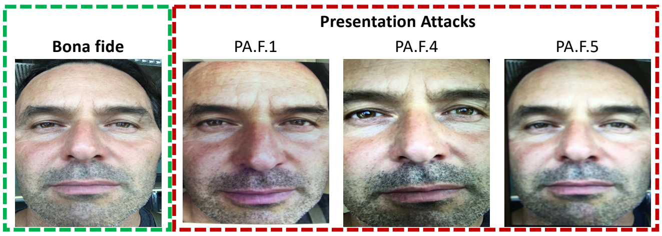


Fig. 4: Illustration of SWAN Presentation Attack dataset (a) Bona fide (b) Presentation Attack (PA.F.1) (c) Presentation Attack (PA.F.4) (d) Presentation Attack (PA.F.5)

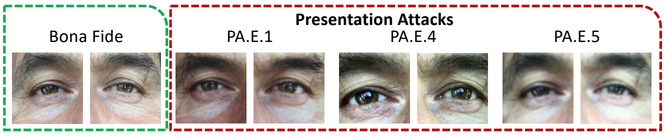


Fig. 5: Illustration of SWAN Presentation Attack dataset (a) Bona fide (b) Presentation Attack (PA.E.1) (c) Presentation Attack (PA.E.4) (d) Presentation Attack (PA.E.5)

Figure 4 & Figure 5 shows the example images for face and eye presentation attacks respectively.

IV. Experimental Performance Evaluation Protocols

In this section, we discuss the evaluation protocol that is used to report the performance of the baseline algorithm for both verification and Presentation Attack Detection (PAD) subsystems.

A. *Biometric verification performance evaluation protocol*

To evaluate the biometric verification performance, we propose two protocols: **Protocol-1:** designed to study the individual verification performance from independent sessions. Thus, we enrol images from session 2 as reference samples and then compute the verification performance using samples from session 3 to session 6 individually as probe samples. **Protocol-2:** this protocol is designed to evaluate the performance of the biometric system across all session. Thus, in this protocol, we enrol images from session-2 and probe biometric samples from all session 3 to session 6, to evaluate the verification performance. However, for simplicity, we have used only self-captured data to report the performance of the baseline verification systems.

B. *Presentation Attack Detection protocol*

To evaluate the performance of the baseline presentation attack detection techniques we propose two different evaluation protocols. **Protocol-1**: This protocol is to evaluate the performance of the PAD techniques independently on each PAI, thus training and testing of the PAD techniques are carried out independently on each PAI. **Protocol-2**: This protocol is designed to evaluate the performance of the PAD algorithms when trained and tested with all PAIs. To effectively evaluate the PAD algorithms, we divide the whole dataset to have three independent partitions such that training set has 90 subjects, the development set has 30 subjects and also the testing set has 30 subjects. Further, we perform the cross validation by randomising the selection of training, development and testing set for $N = 5$ times and average the results, which are reported with standard deviation.

V. Baseline Systems

In this section, we discuss baseline biometric systems used to benchmark the performance of biometric verification and presentation attack detection.

A. *Biometric verification*

a) Face Biometrics: Two deep learning based face biometric systems are evaluated: VGG-Face[1] [21] and FaceNet [22], [23]. The choice of networks is based on the obtained accuracy on the LFW dataset - challenging FR dataset [24] where FaceNet reported an accuracy of 99.2% and VGG-Face reported an accuracy of 98.95%.

b) Eye Biometrics: To evaluate the periocular biometric system, we have used five different methods that include: Coupled-Autoencoder [25], Collaborative representation of Deep-sparse Features [26], Deep Convolution Neural Network (DCNN) features from pre-trained networks such as: AlexNet, VGG16 and ResNet50. These baseline systems are selected based on the reported verification performance especially on the smartphone environment.

c) Audio-visual biometrics: Inter-Session Variability (ISV) [27], [28] and a deep learning based network (ResNet) [29] is used for evaluating the voice biometrics. ISV is a GMM-UBM [30] based system which explicitly models the session variability. We have used an extended ResNet implementation of [29] named dilated residual network (DRN) which is publicly available[2]. The DRN model is one of the state-of-the-art systems on the Voxceleb 1 database [31] evaluations achieving 4.8% EER on the dataset.

The audio-visual system is the result of score fusion between the FaceNet and the DRN models which are the best performing algorithm for face and voice subsystems, respectively.

B. *Presentation Attack Detection Algorithms*

a) Face and Eye Attack Instruments: We report the detection performance of five different PAD algorithms on both eye and face attack instruments. The baseline algorithms include Local Binary Pattern (LBP) [32], Local Phase Quantisation (LPQ)[32], Binarised Statistical Image Features (BSIF)[33], Image Distortion Analysis (IDA) [34] and Color texture features [35]. We use linear SVM as the classifier that is trained and tested with these features following different evaluation protocols as discussed in Section IV. We have selected these five PAD algorithms as the baseline methods to cover the spectrum of both micro-texture and image quality based attack detection techniques.

[1] Website: www.robots.ox.ac.uk/~vgg/software/vgg_face

[2] https://www.idiap.ch/software/bob/docs/bob/bob.learn.pytorch/v0.0.4/guide_audio_extractor.html

b) Audio-visual attack instruments: We use three features: MFCC,SCFC,LFCC with two classifiers (SVM and GMM) [36] to develop voice PAD systems. The same PAD systems that were used for still faces were also used here. The final audio-visual PAD system is the result of score-level fusion of the best algorithms of face and audio PAD systems (Color Textures-SVM and LFCC-GMM).

VI. Experimental Results

In this section, we present the baseline algorithms performance evaluation on both biometric verification and presentation attack detection. The performance of the baseline algorithms is presented using the Equal Error Rate (%) following the experimental protocol presented in Section IV. The performance of the baseline PAD algorithms is presented using *Bona fide Presentation Classification Error Rate (BPCER) and Attack Presentation Classification Error Rate (APCER)*. **BPCER** is defined as the proportion of bona fide presentations incorrectly classified as attacks while **APCER** is defined as the proportion of attack presentations incorrectly classified as bona fide presentations. In particular, we report the performance of the baseline PAD techniques by reporting the value of BPCER while fixing the APCER to 5% and 10% according to the recommendation from IS0/IEC 30107-3 [37]. The PAD evaluation protocol is presented in Section IV.

A. Biometric Verification Results

Table VI indicates the performance of the uni-modal biometric systems with Eye, 2D Face and Audio-visual. For the simplicity, we have used the self capture images from each session to present the quantitative results on eye verification using baseline methods (see Section IV for evaluation protocols). Based on the obtained results as presented in Table VI, the following can be observed (1) among the five different methods, the DeepSparse-CRC [26] method shows the marginal improvement over other techniques in both Protocol-1 and Protocol-2 (c). (2) Both DeepSparse-CRC [26] and Deep Autoencoder [25] algorithms indicate a degraded performance for Protocol-2 when compared to that of the Protocol-1. This can be attributed to the capture data quality reflected from all four sessions (S3-S6).

Both face biometric systems showed similar performance across all sessions. The FaceNet system outperformed the VGG-Face system by a large margin. However, the same FaceNet system's performance was degraded when evaluated on audio-visual data. This can be attributed to the fact the camera was held at a reading position (lower down) during the audio-visual data capture process compared to holding device higher up when taking still face videos. The DRN speaker verification system performed well compared to the ISV system (not shown for brevity, the ISV's EER on all sessions was 13.1%). The worst performance was achieved on session 3 which can be mainly attributed to the noise (especially from wind) outdoor. The final audio-visual biometric system was the average score of FaceNet and DRN systems. Performing this fusion improved the results (except for session 4) which showed that the information from face and voice were mainly complementary in performing verification.

TABLE VI: Baseline performance of uni-modal biometric system

Modality	Algorithms	Enrolment	Probe	EER(%)
Eye biometrics	Deep Autoencoder [25]	Session 2	Session 3	25.59
			Session 4	23.88
			Session 5	27.26
			Session 6	23.07
			All Session	25.67
	DeepSparse-CRC [26]	Session 2	Session 3	21.32
			Session 4	22.30
			Session 5	22.55
			Session 6	23.54
			All Session	22.41
	AlexNet features	Session 2	Session 3	24.23
			Session 4	28.72
			Session 5	27.21
			Session 6	21.46
			All Session	25.46
	ResNet features	Session 2	Session 3	29.59
			Session 4	24.78
			Session 5	24.17
			Session 6	20.79
			All Session	24.77
	VGG16 features	Session 2	Session 3	34.62
			Session 4	27.41
			Session 5	27.45
			Session 6	24.82
			All Session	28.62
Face biometrics	FaceNet (InceptionResNetV1) [23]	Session 2	Session 3	4.3
			Session 4	3.3
			Session 5	5.3
			Session 6	2.7
			All Session	4.2
	VGG-Face [38]	Session 2	Session 3	17.2
			Session 4	17.3
			Session 5	16.6
			Session 6	17.2
			All Session	17.0
Audio-Visual	FaceNet (InceptionResNetV1) [23]	Session 2	Session 3	14.2
			Session 4	13.1
			Session 5	13.9
			Session 6	13.0
			All Session	13.5
	DRN	Session 2	Session 3	4.3
			Session 4	2.1
			Session 5	3.2
			Session 6	3.3
			All Session	3.2
	FaceNet-DRN-score-mean-fusion	Session 2	Session 3	3.4
			Session 4	3.0
			Session 5	3.1
			Session 6	2.9
			All Session	3.1

B. Biometric vulnerability assessment

Figure 6 illustrates the vulnerability analysis on the uni-modal biometrics from the SWAN multimodal biometric dataset. The vulnerability analysis is performed using the baseline uni-modal biometric in which, the bona fide samples are enrolled and presentation attack samples are used as the probe. Finally, the comparisons scores obtained on the probe samples are compared against the operational threshold set, and finally, the quantitative results of the vulnerability are presented using an Impostor Attack Presentation Match Rate (IAPMR (%)). For vulnerability analysis with eye biometric system we have used the DeepSparse-CRC [26] system, for 2D face biometrics we have used FaceNet (InceptionResNetV1) [23] and for audio-visual biometric system we have employed

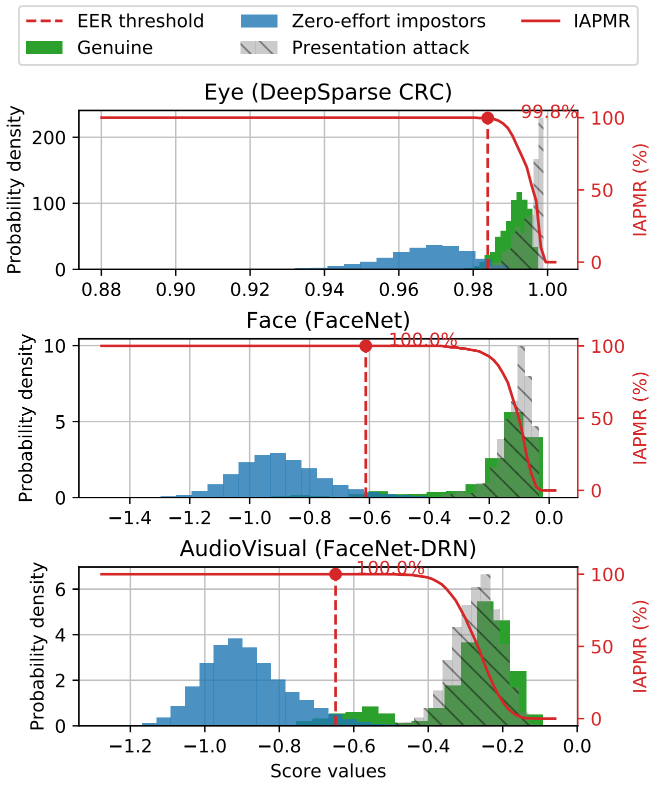


Fig. 6: Vulnerability analysis of uni-modal biometric systems

FaceNet (InceptionResNetV1) [23] for face and DRN [29] for voice whose scores are fused at comparisons core level. All three types of presentation attack samples are used as the probe samples to compute the IAMPR. As indicated in the Figure 6, all three modalities have indicated IAMPR of 100% on both face and audiovisual biometrics and 99.8% on eye biometrics. The obtained results further justify the quality of the presentation attacks generated in this work and also the need for Presentation Attack Detection (PAD) techniques to mitigate the presentation attacks.

C. Biometric PAD results

Table VII indicates the quantitative performance of the baseline PAD algorithms on the eye recognition subsystem on Protocol-1 and Protocol-2. Based on the obtained results, the following can be observed: (1) Among the three different PAI, the detection of PA.E.4 (iPad Pro front camera) indicates an excellent detection accuracy on all five baseline methods employed in this work. (2) Among five different baseline PAD algorithms, the PAD technique based on Color Textures-SVM [35] has indicated the best performance on both Protocol-1 and Protocol-2. (3) The performance of the PAD algorithms in Protocol-2 shows a degraded performance when compared to that of the Protocol-1.

Table VIII and IX indicates the quantitative performance of the baseline PAD algorithms on 2D face modality on Protocol-1 and Protocol-2. Based on the obtained results, it is noted that (1) the performance of the PAD techniques are degraded in the Protocol-2 when compared to Protocol-1. (2) Among

TABLE VII: Baseline performance of PAD techniques on Eye biometrics

Evaluation Protocol	Algorithms	Development Database	Testing Database		
				BPCER@ APCER	
		D-EER(%)	D-EER(%)	= 5(%)	= 10(%)
Protocol-1 (PA.E.1)	**BSIF-SVM [33]**	10.14 ± 1.16	11.78 ±2.62	57.95 ± 40.16	37.54 ± 29.97
	Color Textures-SVM [35]	3.65 ± 2.13	1.07 ± 1.47	17.59 ± 16.49	10.51±11.54
	IDA-SVM [34]	19.42 ± 17.46	24.23 ± 19.80	44.87 ± 26.86	37.77 ± 25.83
	LBP-SVM [32]	5.52 ± 1.17	9.55 ± 2.15	17.70 ± 14.90	16.86 ± 16.41
	LPQ-SVM [32]	13.14 ± 3.74	12.63 ± 2.99	53.80 ± 32.60	37.34 ± 28.74
Protocol-1 (PA.E.4)	**BSIF-SVM [33]**	7.51 ± 2.04	8.53 ± 1.92	34.41 ± 13.80	18.82 ± 5.59
	Color Textures-SVM [35]	0.06 ± 0.06	0.85 ± 0.57	0.31 ± 0.58	0.11+ 0.22
	IDA-SVM [34]	20.41 ± 8.24	24.35 ± 13.09	61.88 ± 32.19	55.44 ± 37.87
	LBP-SVM [32]	4.47 ± 0.92	7.26 ± 3.31	8.13 ± 7.44	5.51 ± 6.74
	LPQ-SVM [32]	5.22 ± 1.44	5.18 ± 1.25	22.07 ± 10.27	9.84 ± 6.54
Protocol-1 (PA.E.5)	**BSIF-SVM [33]**	10.33 ± 2.03	13.60 ± 2.98	31.96 ± 17.10	21.80 ± 11.40
	Color Textures-SVM [35]	1.35 ± 1.33	0.97 ± 0.52	1.64 ± 3.49	0.76 ± 1.66
	IDA-SVM [34]	17.81 ± 9.22	21.50 ± 6.89	29.18 ± 19.26	20.97 ± 14.33
	LBP-SVM [32]	1.12 ± 0.48	1.52 ± 0.65	2.14 ± 2.13	1.02 ± 1.24
	LPQ-SVM [32]	1.34 ± 0.65	1.91 ± 0.29	11.65 ± 24.25	4.76 ± 10.35
Protocol-2	**BSIF-SVM [33]**	10.02 ± 1.34	12.04 ± 2.89	42.02 ± 21.27	24.60 ± 10.51
	Color Textures-SVM [35]	2.01 ± 0.98	2.68 ± 0.72	4.79 ± 8.37	1.93 ± 3.50
	IDA-SVM [34]	35.40 ± 9.78	38.16 ± 8.37	80.35 ± 13.96	73.29 ± 16.75
	LBP-SVM [32]	6.34 ± 0.63	8.95 ± 2.11	22.02 ± 25.93	14.75 ± 19.41
	LPQ-SVM [32]	6.92 ± 2.77	7.20 ± 1.73	22.69 ± 12.80	9.57 ± 5.79

TABLE VIII: Baseline performance of PAD techniques on face biometrics

Evaluation Protocol	Algorithms	Development Database	Testing Database		
				BPCER@ APCER	
		D-EER(%)	D-EER(%)	=5(%)	=10(%)
Protocol-1 (PA.F.1)	**BSIF-SVM [33]**	5.20±8.75	10.13±2.58	38.92±29.24	30.16± 24.20
	Color Textures-SVM [35]	2.20±1.25	8.71±1.28	24.87± 29.68	18.90±22.16
	IDA-SVM [34]	34.50±22.41	37.61±22.32	63.78±30.58	56.77±36.27
	LBP-SVM [32]	3.08±1.91	13.63±2.06	39.23±27.97	33.78±24.86
	LPQ-SVM [32]	5.67±2.67	13.87±2.69	30.52±12.52	22.54±9.47
Protocol-1 (PA.F.5)	**BSIF-SVM [33]**	17.16±5.11	20.30±1.19	67.48±30.14	50.32±23.73
	Color Textures-SVM [35]	11.57±4.37	12.46±3.21	45.76±19.67	27.96±9.17
	IDA-SVM [34]	24.55±3.11	29.08±13.40	56.53±26.66	46.87±28.95
	LBP-SVM [32]	15.83±3.65	17.13±4.38	42.05±18.31	30.96±16.13
	LPQ-SVM [32]	24.25±2.75	18.15±3.07	63.63±25.64	53.83±26.49
Protocol-1 (PA.F.6)	**BSIF-SVM [33]**	17.76±4.27	27.74±1.55	70.50±16.30	59.48±15.43
	Color Textures-SVM [35]	4.21±1.79	5.66±2.13	12.29±7.79	8.08±4.85
	IDA-SVM [34]	31.78±12.69	30.99±12.99	71.09±29.95	62.75±37.14
	LBP-SVM [32]	6.70±0.98	9.41±2.59	32.21±22.19	19.30±16.50
	LPQ-SVM [32]	17.07±3.18	17.04±2.41	51.63±26.95	38.36±25.62
Protocol-2	**BSIF-SVM [33]**	20.05±3.97	25.55±1.51	58.12±13.61	46.06±12.76
	Color Textures-SVM [35]	7.07±1.13	14.45±1.56	47.31±7.66	36.06±6.33
	IDA-SVM [34]	32.66±19.22	33.49±16.17	66.54±15.45	58.79±16.99
	LBP-SVM [32]	27.67±11.90	35.95±7.65	75.09±12.18	64.23±16.14
	LPQ-SVM [32]	18.47±3.93	20.23±3.84	48.28±10.95	39.05±9.33

TABLE IX: Baseline performance of PAD techniques on Talking Faces biometrics

Evaluation Protocol	Algorithms	Development Database	Testing Database		
				BPCER@ APCER	
		D-EER (%)	D-EER (%)	=5(%)	=10(%)
Protocol-1 (PA.F.5)	**BSIF-SVM**	23.25±5.08	27.32±3.91	97.92±1.77	91.88±6.18
	Color Textures-SVM	2.67±1.57	5.45±1.50	57.93±27.78	43.67±27.25
	IDA-SVM	35.00±8.99	37.87±9.20	53.79±44.49	50.84±43.82
	LBP-SVM	17.83±12.12	24.06±11.01	52.37±37.12	39.69±33.97
	LPQ-SVM	5.25±2.86	8.66±1.38	30.87±25.64	19.46±19.70
Protocol-1 (PA.F.6)	**BSIF-SVM**	21.74±4.55	29.34±3.67	81.30±11.92	68.47±19.95
	Color Textures-SVM	5.47±2.98	15.21±2.86	48.55±40.99	43.67±43.22
	IDA-SVM	37.88±17.72	41.69±9.97	69.82±35.84	66.42±35.48
	LBP-SVM	38.51±8.25	41.13±8.85	96.36±4.05	92.79±7.52
	LPQ-SVM	19.17±3.06	25.10±1.97	63.90±33.95	54.94±37.27
Protocol-2	**BSIF-SVM**	26.48±3.31	24.56±2.95	65.17±14.69	53.78±17.40
	Color Textures-SVM	4.31±1.50	5.18±1.58	9.72±7.24	5.22±4.61
	IDA-SVM	37.07±10.02	39.15±9.48	65.94±35.50	58.39±36.43
	LBP-SVM	34.00±7.64	31.35±6.93	52.11±38.95	45.56±41.87
	LPQ-SVM	16.72±1.98	16.86±3.68	60.44±21.89	51.61±25.25

TABLE X: Baseline performance of PAD techniques on Voice biometrics

Evaluation Protocol	Algorithms	Development Database	Testing Database		
		D-EER (%)	D-EER (%)	BPCER@ APCER	
				=5(%)	=10(%)
Protocol-1 (PA.V.4)	**MFCC-SVM**	0.00±0.00	0.11±0.13	0.00±0.00	0.00±0.00
	LFCC-GMM	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
	SCFC-GMM	1.28±0.38	1.21±0.38	0.02±0.03	0.00±0.00
Protocol-1 (PA.V.7)	**MFCC-SVM**	0.00±0.00	0.12±0.10	0.00±0.00	0.00±0.00
	LFCC-GMM	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
	SCFC-GMM	1.58±0.49	1.17±0.21	0.08±0.06	0.00±0.00
Protocol-2	**MFCC-SVM**	1.52±0.43	1.67±0.17	0.14±0.16	0.03±0.04
	LFCC-GMM	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
	SCFC-GMM	2.09±0.41	1.48±0.07	0.04±0.02	0.00±0.00

TABLE XI: Baseline performance of PAD techniques on AudioVisual biometrics

Evaluation Protocol	Algorithm:	Development Database	Testing Database		
		D-EER (%)	D-EER (%)	BPCER@ APCER	
				=5(%)	=10(%)
Protocol-1 Low-Quality (PA.F.5 and PA.V.7)	**Fusion of Color Textures-SVM and LFCC-GMM**	0.08±0.17	0.09±0.11	0.65±0.68	0.24±0.15
Protocol-1 High-Quality (PA.F.6 and PA.V.4)		21.74±4.55	29.34±3.67 8	1.30±11.92 6	8.47±19.95
Protocol-2 All Attacks		26.48±3.31	24.56±2.95 6	5.17±14.69 5	3.78±17.40

five different baseline PAD algorithms, the PAD technique based on Color Textures-SVM [35] has indicated the best performance on both Protocol-1 and Protocol-2.

Table X indicates the quantitative performance of the baseline PAD algorithms on voice biometrics. In all protocols the LFCC-GMM system showed well performance (0% error rates) in detection of the presentation attacks. Among other systems, protocol 2 was the more challenging protocol in the MFCC-SVM and SCFC-GMM baselines.

Finally, Table XI reports the score mean fusion of Color Textures-SVM and LFCC-GMM baselines (best performing systems of each modality) on Audio-Visual attacks. The low-quality attacks (PA.F.5 and PA.V.7) were detected with error rates less than 1%. However, the performance is degraded significantly when high quality attacks are used with a D-EER of 29.34% on the testing database. This degradation in performance is also evident in protocol 2 where both low and high quality attacks are used.

VII. Research Potential for SWAN Multimodal Biometric Dataset

In this section, we summarize the research potential that can be anticipated using SWAN multimodal biometric dataset. As emphasized in Section II-A that, the SWAN multimodal dataset includes new challenging scenarios and evaluation protocols that are tailored for the verification experiments. This dataset is unique in terms of the number of data subjects captured in four different geographical locations using the same capture protocols. Thus, the following are research direction that can be pursued with SWAN multimodal biometric dataset:

- Developing novel algorithms for smartphone-based biometric verification on both unimodal (face, voice and periocular) and multimodal biometric characteristics.
- Study of variation in developed biometric systems (both unimodal and multimodal) due to environmental variations captured in six different sessions.
- Study on multi-lingual speaker verification as each data-subject has recorded its voice samples in multiple languages.
- Evaluation of both unimodal and multimodal Presentation Attack Detection (PAD) algorithms.
- Generation of new Presentation Attack Instrument (PAI) and also the vulnerability evaluation of potential PAIs.
- Demographic study including age, gender, or visual aids.
- Study on cross-European diversity and geographic variability in terms of both unimodal and multimodal biometrics.

VIII. Data-set distribution

The Idiap subset of the database will be distributed online on the Idiap dataset distribution portal (https://www.idiap.ch/dataset). Also we are aiming to make the complete database available through the BEAT platform[3] [39]. The BEAT platform is a European computing e-infrastructure solution for open access and scientific information sharing and re-use. Both the data and the source code of experiments are shared while protecting privacy and confidentiality. As the data on the BEAT platform is easily available for experimental research but cannot be downloaded, it makes BEAT an ideal platform for biometrics research while the privacy of users who participated in the data collection is maintained.

IX. Conclusions

Smartphone-based biometric verification has gained a lot of interest among researchers from the past few years. Availability of the publicly available datasets is crucial to driving the research forward so that new algorithms are developed and benchmarked with the existing algorithms. However, the collection of biometric datasets is resource consuming, especially in collecting the multimodal biometric dataset from different geographic locations. In this work, we present a new multimodal biometric dataset captured using a smartphone together with the evaluation of the baseline techniques. This dataset is a result of the collaboration between three European partners within the framework of the SWAN project sponsored by Research Council of Norway. This new dataset is captured in the challenging conditions in four different geographic locations. The whole dataset is obtained from 150 data subjects in six different sessions that can simulate the real-life scenarios. Besides, a new presentation attack (or spoofing) dataset is also presented for multimodal biometrics. A brief description of the performance evaluation protocols and the baseline biometric verification and presentation attack detection algorithms are benchmarked. Experimental findings using the baseline algorithms are highlighted for both biometric verification and presentation attack detection.

Acknowledgement

This work is carried out under the funding of the Research Council of Norway (Grant No. IKTPLUSS 248030/O70)

[3] https://www.beat-eu.org/platform/

References

[1] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 2382-37:2017 Information Technology - Vocabulary - Part 37: Biometrics*, International Organization for Standardization, 2017.

[2] "Acuity Intelligence Forecast," http://www.acuity-mi.com, accessed: 2019-04-29.

[3] "Global Smartphone Production Volume May Decline by Up to 5% in 2019, Huawei Would Overtake Apple to Become World's Second Largest Smartphone Maker, Says TrendForce." [Online]. Available: https://press.trendforce.com/node/view/3200.html

[4] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Comput. Surv.*, vol. 50, no. 1, pp. 8:1–8:37, Mar. 2017. [Online]. Available: http://doi.acm.org/10.1145/3038924

[5] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of biometric anti-spoofing*. Springer, 2018, vol. 1.

[6] Y. Zhang, Z. Chen, H. Xue, and T. Wei, "Fingerprints on mobile devices: Abusing and leaking," in *Black Hat Conference*, 2015.

[7] A. Roy, N. Memon, and A. Ross, "Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2013–2025, Sep. 2017.

[8] A. Rattani and R. Derakhshani, "A survey of mobile face biometrics," *Computers & Electrical Engineering*, vol. 72, pp. 39 – 52, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S004579061730650X

[9] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Mobile iris challenge evaluation (miche)-i, biometric iris dataset and protocols," *Pattern Recogn. Lett.*, vol. 57, no. C, pp. 17–23, May 2015. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2015.02.009

[10] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 49–61, July 2016.

[11] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, and R. Singh, "On smartphone camera based fingerphoto authentication," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Sep. 2015, pp. 1–7.

[12] C. McCool, S. Marcel, A. Hadid, M. Pietikinen, P. Matejka, J. Cernock, N. Poh, J. Kittler, A. Larcher, C. Lvy, D. Matrouf, J. Bonastre, P. Tresadern, and T. Cootes, "Bi-modal person recognition on a mobile phone: Using mobile phone data," in *2012 IEEE International Conference on Multimedia and Expo Workshops*, July 2012, pp. 635–640.

[13] E. Bartuzi, K. Roszczewska, R. Białobrzeski *et al.*, "Mobibits: Multimodal mobile biometric database," in *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2018, pp. 1–5.

[14] "Secure Access Control over Wide Area Network (SWAN)," https://www.ntnu.edu/iik/swan/, accessed: 2019-08-09.

[15] G. Santos, E. Grancho, M. V. Bernardo, and P. T. Fiadeiro, "Fusing iris and periocular information for cross-sensor recognition," *Pattern Recognition Letters*, vol. 57, pp. 52 – 59, 2015, mobile Iris CHallenge Evaluation part I (MICHE I). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865514003006

[16] D. Kim, K. Chung, and K. Hong, "Person authentication using face, teeth and voice modalities for mobile device security," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2678–2685, November 2010.

[17] A. F. Sequeira, J. C. Monteiro, A. Rebelo, and H. P. Oliveira, "Mobbio: A multimodal database captured with a portable handheld device," in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 3, Jan 2014, pp. 133–139.

[18] U. Mahbub, S. Sarkar, V. M. Patel, and R. Chellappa, "Active user authentication for smartphones: A challenge data set and benchmark results," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Sep. 2016, pp. 1–8.

[19] E. Bartuzi, K. Roszczewska, M. rokielewicz, and R. Biaobrzeski, "Mobibits: Multimodal mobile biometric database," in *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2018, pp. 1–5.

[20] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. N. R. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S. Pavani, A. Frangi, L. Akarun, and A. Savran, "The multiscenario multienvironment biosecure multimodal database (bmdb)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1097–1111, June 2010.

[21] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, vol. 1. BMVA Press, 09 2015, pp. 41.1 – 41.12.

[22] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015, pp. 815 – 823, 00297.

[23] D. Sandberg, "facenet: Face recognition using tensorflow," https://github.com/davidsandberg/facenet, accessed: 2017-08-01.

[24] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[25] R. Raghavendra and C. Busch, "Learning deeply coupled autoencoders for smartphone based robust periocular verification," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 325–329.

[26] K. B. Raja, R. Raghavendra, and C. Busch, "Collaborative representation of deep sparse filtered features for robust verification of smartphone periocular images," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 330–334.

[27] R. Vogt and S. Sridharan, "Explicit modelling of session variability for speaker verification," *Computer Speech & Language*, vol. 22, no. 1, pp. 17 – 38, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0885230807000277

[28] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel, "Session variability modelling for face authentication," *IET Biometrics*, vol. 2, no. 3, pp. 117–129, Sep. 2013.

[29] N. Le and J.-M. Odobez, "Robust and discriminative speaker embedding via intra-class distance variance regularization," in *Proceedings Interspeech*, 2018, pp. 2257–2261.

[30] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital signal processing*, vol. 10, no. 1, pp. 19–41, 2000.

[31] A. Nagrani, J. S. Chung, and A. Zisserman, "VoxCeleb: a large-scale speaker identification dataset," *arXiv:1706.08612 [cs]*, Jun. 2017, arXiv: 1706.08612. [Online]. Available: http://arxiv.org/abs/1706.08612

[32] I. Chingovska, A. Andr, and S. Marcel, "Biometrics evaluation under spoofing attacks," *IEEE Transactions on Information Forensics and Security (T-IFS)*, vol. 9, no. 12, pp. 2264–2276, Dec 2014.

[33] R. Ramachandra and C. Busch, "Presentation attack detection algorithm for face and iris biometrics," in *22nd European Signal Processing Conference, EUSIPCO 2014, Lisbon, Portugal*, 2014, pp. 1387–1391.

[34] D. Wen, H. Han, and A. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 99, pp. 1–16, 2015.

[35] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 2636–2640.

[36] M. Sahidullah, T. Kinnunen, and C. Hanili, "A comparison of features for synthetic speech detection," in *Proc. INTERSPEECH*. Citeseer, 2015, pp. 2087–2091. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.709.5379&rep=rep1&type=pdf

[37] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*, International Organization for Standardization, 2017.

[38] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[39] A. Anjos, L. El-Shafey, and S. Marcel, "Beat: An open-science web platform," in *International Conference on Machine Learning (ICML)*, Aug. 2017. [Online]. Available: https://publications.idiap.ch/downloads/papers/2017/Anjos_ICML2017_2017.pdf