MobiBits: Multimodal Mobile Biometric Database

Ewelina Bartuzi, Katarzyna Roszczewska, Mateusz Trokielewicz, Radosław Białobrzeski

Biometrics and Machine Inteligence Laboratory
Research and Academic Computer Network
Kolska 12, 01045 Warsaw, Poland
corresponding author: ewelina.bartuzi@nask.pl

Abstract—This paper presents a novel database comprising representations of five different biometric characteristics, collected in a mobile, unconstrained or semi-constrained setting with three different mobile devices, including characteristics previously unavailable in existing datasets, namely hand images, thermal hand images, and thermal face images, all acquired with a mobile, off-the-shelf device. In addition to this collection of data we perform an extensive set of experiments providing insight on benchmark recognition performance that can be achieved with these data, carried out with existing commercial and academic biometric solutions. This is the first known to us mobile biometric database introducing samples of biometric traits such as thermal hand images and thermal face images. We hope that this contribution will make a valuable addition to the already existing databases and enable new experiments and studies in the field of mobile authentication. The *MobiBits* database is made publicly available to the research community at no cost for non-commercial purposes.

Index Terms—biometrics, recognition, multimodal, database, signatures, voice, speaker, face, hand, iris, thermal imaging, visible spectrum, mobile devices, smartphones.

I. INTRODUCTION

Implementation of biometric authentication on mobile devices in the recent years has lead to an almost universally ubiquitous presence of biometrics in our daily lives. Starting with the fingerprint sensor of the iPhone 5s in 2013, more than 700 mobile devices have since supported this kind of authentication, with many manufacturers employing other biometric characteristics: iris in Samsung Galaxy S8 or Microsoft Lumia 950, and recently also face in 3D with Apple iPhone X's Face ID. Today, biometric authentication is used for unlocking a phone, but also for protecting financial assets in payment systems, *e.g.,* Apple Pay. With this reliance upon mobile identity, secure and convenient authentication methods are crucial, hence the recently increasing interest in biometric authentication for mobile. Biometric traits such as face and fingerprint have already been given considerable attention in the scientific community, therefore, the main contributions of our paper include:
(1) a multimodal biometric database – *MobiBits*, comprising representations of five different biometric characteristics, acquired with sensors embedded in commercially available, consumer-grade mobile phones,
(2) a novel use of thermal imaging enabled smartphone, adding a new dimension to typical, visible spectrum images of face and hand, offering additional cues that can be
utilized, *e.g.,* for Presentation Attack Detection,
(3) experiments employing commercial and academic methods, which provide benchmark results of recognition performance for the introduced datasets.
According to our knowledge, our dataset offers the largest number of biometric characteristics represented in a single database, and employs the richest collection of data acquisition scenarios, which are trying to mimic those of potential real-world applications.

II. RELATED WORK

Iris. Iris recognition, usually carried out with specialized devices operating in near infrared (NIR), poses several challenges when implemented on mobile – most importantly, the need to utilize visible light images. Raja *et al.* investigated visible spectrum iris recognition on mobile devices by employing deep sparse filtering [1] and K-means clustering [2], being able to achieve equal error rate (EER) of 0.31%. Trokielewicz *et al.* showed that high quality images acquired with an iPhone 5s can be successfully used with existing commercial and academic iris matchers, with FTE of 0%, and close-to-perfect correct recognition rates, and introduced a dataset of high quality iris images [3]. Cross-spectral iris recognition between mobile phone samples and typical NIR images is studied in [4], with EER$\simeq 2\%$.

Handwritten signatures. With advantages such as ease of use, familiarity, and social acceptability, signature biometrics has the potential of mobile use. On a smartphone, one can acquire on-line handwritten signatures (X,Y) coordinates in time, as well as pressure (selected phones). On-line signature recognition methods can be divided into two groups: non-parametric and parametric. The most popular method from the first group is Dynamic Time Warping (DTW) [5], [6], which normalizes signatures and matches the two samples. EERs achieved in experiments with drawing tablets ranged from 0.6 % to 3.4% for random forgeries and from 5.4% to 17.18% for skilled forgeries. The most common parametric method are Hidden Markov Models [7], [8]. These methods determine the probability of sample belonging to the learned model of transitions between observation states. EERs range from 4.6% to 7.3% for skilled forgeries, whereas the average for random forgeries EER$\simeq 3\%$.

Face. In 2012, the bi-modal MOBIO database was published [9], including experiments for face and speaker bio-

	MobiBits	*MobBIO*	*MBMA*	*BioSecure*	*FTV*	*I-Am*	*MOBIO*
Volunteers	53	105	100	713	50	100	150
Acquisition sessions	3	1	2	2	1	1–3	12
Devices used	Huawei Mate S Huawei P9 Lite CAT S60	Asus Transformer Pad TF 300	iPhone 4s Galaxy S2 iPad 2 Motorola Xoom	Samsung Q1 Philips SPC900NC Webcam HP iPAQ hx2790 PDA	HP iPAQ rw6100	Samsung Galaxy S4	Nokia N93i MacBook
Biometric characteristics included	signatures voice face iris hand	voice face iris	voice signatures face	voice signatures face fingerprints	face teeth voice	arm gestures ear shape	face voice

TABLE I
COMPARISON BETWEEN EXISTING MOBILE MULTIMODAL BIOMETRIC DATABASES THE *MobiBits* DATABASE.

metrics running on a Nokia N93i smartphone. Fusing both characteristics resulted in EER≃10%. Half error rates were achieved using the same database and a range of different methods in the ICB 2013 Face Recognition Evaluation [10]. Mobile face recognition with locally executed implementations has recently gained interest due to the iPhone X, which employs 3D face recognition.

Voice. Due to its convenience in some applications and increasing accuracy, text-independent speaker recognition can be a natural choice for mobile authentication systems. The bi-modal MOBIO database [9], contains a collection of voice samples captured with a Nokia N93i. This database was also used in the ICB 2013 Speaker Recognition Evaluation [11], resulting in encouraging error rates of 5% EER.

Hand. Hand images can supply information on both hand geometry and texture. The authors of [12] developed a biometric recognition method based on features extracted from hand silhouette and its contour, obtaining EER=3.7%. According to the works of Zhang *et al.* and Sun *et al.* [13], [14], large areas of the palmar side of the hand provide enough personal information for the identity authentication, even with low-resolution images of less than 100 dpi. Hand could be a useful for biometric authentication thanks to simple acquisition process using the built-in camera on a mobile phone without an additional sensor. The most popular feature extraction methods utilize a bank of Gabor filters [15], [16], and SIFT descriptors [17], with best EER=0.79% obtained by Kim *et al.* [15].

Available mobile biometric datasets. The **MobBIO** database by Sequeira *et al.* [18] contains data acquired from 105 volunteers with an Asus Transformer Pad TF 300 tablet. It includes 16 voice samples, 8 eye images and 16 face images captured in different lighting conditions. The **BioSecure** database by Ortega-Garcia *et al.* has been collected at several universities participating in the BioSecure project [19]. It contains data collected in two acquisition sessions from 713 volunteers, each of them providing 4 frontal face images, 12 fingerprint samples, 18 voice recordings, and 25 handwritten signatures (genuine and skilled forgeries). The **MBMA** database [20] contains three characteristics (voice, face, signature) collected in two sessions with four 4 different mobile devices. The first session was used for system development, with samples acquired from 100 volunteers. Each

volunteer provided 5 face images, 3 voice records and 6 handwritten signatures. Data from the second session was acquired from 32 people and it contains 4 samples of voice recordings, 3 face images and 8 signatures. The **Face-Teeth-Voice** collection introduced in [21] gathers 1000 biometric samples: face images, teeth images, and voice recordings in an experiment involving 50 people, hence 20 samples per person, all obtained in a single session. Finally, the **I-Am** database from [22] comprises a combination of physical and behavioral biometrics in the form of ear shape images and accelerometer/gyroscope data, obtained from 100 volunteers in 1 to 3 acquisition sessions. The database collects 300 ear shape samples, and 600 accelerometer/gyroscope data samples. These datasets, including the **MOBIO** dataset described in the previous paragraphs, are summarized in Table I.

III. DATABASE CHARACTERISTICS

The *MobiBits* multimodal mobile biometric database includes data collected from 53 volunteers (20 female and 33 male), using three different smartphones. The age of subjects ranged from 14 to 71 years. The data were collected in 3 sessions organized during 3 following months. Data collection took place twice in each acquisition session, separated by approximately 15 minutes. All samples were acquired in typical office conditions with air conditioning set to 24°C. Other factors that might have influenced the results, such as gender, age, health condition, environmental conditions before the measurement, time from last meal, were recorded in the metadata. For further information on how to get access to the data please contact the authors at: mobibits@nask.pl.

Three smartphones equipped with custom data acquisition software and different technical specifications and capabilities were used for data collection: Huawei Mate S, Huawei P9 Lite, and CAT s60, with the following accessories: an Adonit Dash 2 stylus for signatures to get a more precise contact with the display, hot pillows, and cooling gel compresses were used to simulate different environmental conditions.

ON-LINE HANDWRITTEN SIGNATURES. The on-line handwritten signatures were recorded in real time and the order of the signature components is recorded by Huawei Mate S (with pressure-sensitive screen). Their representations contain
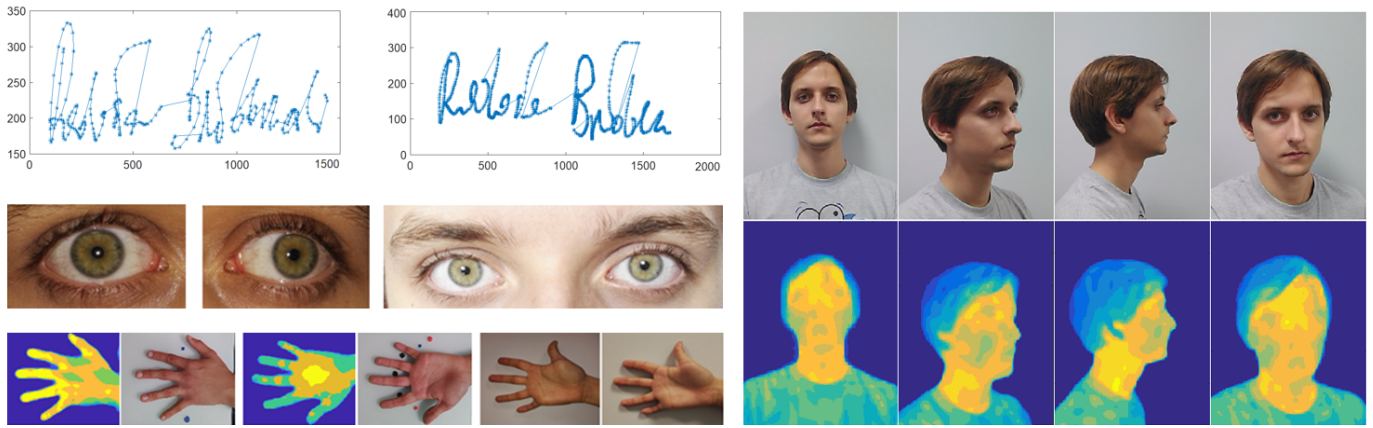
Fig. 1. Selected samples from the MobiBits database. **Left:** genuine signatures and skilled forgeries, iris images, and hand images (thermal and visible light). **Right:** visible light and thermal face images.

(x, y) plane coordinates and pressure values for several points on the display surface, all as a function of time. The database contains 40 genuine signatures and 20 skilled forgeries per person.

VOICE RECORDINGS. The voice samples were captured using a Huawei P9 Lite smartphone in a regular office environment, with no artificial noise added. All speakers were speaking Polish language. Samples were recorded during three sessions of unconstrained speech and reading (sample rate: $44.1 kHz$, duration: 0:01 - 1:10 min).

FACE IMAGES. Face images were collected in three different acquisition sessions using a CAT s60 mobile phone. This smartphone allowed collection of two types of images: photos in visible light with a resolution of 480×640 pixels and thermal images of 240×320 pixels. Each session includes at least two photos for each of s six poses: frontal images, left and right semi-profile images, left and right profile images and a 'selfie' image.

IRIS AND PERIOCULAR. Rear camera (13 Mpx) of the Huawei Mate S device has been used to capture images of irises (Session 1) and face focused on the eyes region (Session 2 and 3). Data collection was carried out in a typical office setting with artificial light sources. Flash was enabled during the first and the third session, and disabled during the second session to simulate different, real-world conditions of authentication.

HAND IMAGES. For hand data, images of palmar and dorsal side of both left and right hand were collected. CAT s60 phone with built-in thermal camera was used. According to the current state of the art, thermal images have never been used in mobile biometrics. Due to this fact, hand images were collected according to scenario which allows the temperature dynamics analyses to stimulate rich research ideas regarding this fascinating, yet largely unstudied biometric characteristic. We collected hand images: supported by glass stand and without, after cooling, after warning and with no temperature influence. One session - *Extra Session*, used 13 Mpx camera, for more accurate images of the hand texture.

IV. DATABASE EVALUATION: TOOLS, PROTOCOL, AND PERFORMANCE METRICS

For the evaluation of the *MobiBits* database we performed benchmark calculations of the recognition accuracy for all characteristics, using a proprietary PLDA/i-vector solution for voice recordings, DTW method for signatures, VeriLook for face images, IriCore for the iris images and VGG-based CNN approach for hand images. **Dynamic Time Warping** (DTW) algorithm calculates an optimal warping path between two given time series is calculated, and then used for comparison [6]. **VeriLook** is a face recognition technology offered by Neurotechnology [23], which uses a set of robust image processing algorithms, including deep neural networks. It allows face recognition with non-frontal images, however, the default allowed rotation angles are modest. **IriCore** is an iris recognition SDK developed by IriTech [24], with the exact algorithm not disclosed. As this method is fine-tuned to work with iris images compliant with the ISO/IEC 19794:2011 standard [25], pre-processing of photos was performed with 640×480 cropping and grayscale conversion of the red channel, which is said improve the visibility of the iris texture in heavily pigmented irises [3], [4]. **PLDA/i-vector voice recognition** is our proprietary solution based on state-of-the-art PLDA method [26], [27] with 40 mean and variance normalized MFCC features extracted from every signal frame, a 256-component, gender-independent background model, 64-dimensional total variability subspace, and 64-dimensional PLDA model. The results were normalized using cohort symmetric score normalization [28]. Prior to feature extraction, additional spectral subtraction voice activity detection was performed to remove non-speech periods [29]. The system was trained exclusively on the MOBIO dataset. **VGG-16-Hand** is a method based on the well-known convolutional network VGG-16 model [30], with weights fine-tuned on a dataset of hand images from the *Extra Session* with flash enabled (cf. Section III), and its bottleneck layers modified to reflect the number of output classes.

IriCore and VeriLook engines performed quality checks

	VOICE			SIGNATURE			FACE			IRIS			HAND		
	classes	samples	FTE	classes	samples	FTE	classes	samples	FTE	classes	samples	FTE	classes	samples	FTE
Session 1	51	406	0.00%	51	1020	0.00%	51	778	45.89%	49	497	14.69%	49	2255	0.00%
Session 2	46	190	0.00%	42	420	0.00%	47	1005	42.00%	46	525	10.48%	49	1508	0.00%
Session 3	48	624	0.00%	50	500	0.00%	47	1105	40.90%	47	2082	12.92%	49	1740	0.00%

TABLE II
A SUMMARY OF THE NUMBER OF SUBJECTS, IMAGES AND FTE RATE FOR EACH CHARACTERISTIC.

prior to feature extraction and biometric template creation. Since some samples were not accepted, *failure-to-enroll rates (FTEs)* were calculated to express the ratio of rejected samples to all samples. Each biometric characteristic and each fusion model was tested in a verification scenario. Averaged Receiver Operating Characteristics (ROCs) and mean Equal Error Rates (EERs) were used to present the verification accuracy. These performance metrics were calculated for comparison scores obtained by performing all possible comparisons within the data, excluding the within-session ones, as these would involve comparing highly correlated data.

Please note that thermal images of hands and those of faces are not used for comparisons here, as we are not aware of any recognition methods that would be capable of efficient processing of such samples. Instead, our intent was to provide them as datasets that are complementary to these of visible light images, so that they can be utilized for other research, such as designing robust methods for Presentation Attack Detection (PAD).

V. RESULTS AND DISCUSSION

FTEs calculated for all tested characteristics are presented in Table II. High FTE values for the face engine may be explained with difficulties in detecting and processing profile images by the VeriLook software, as the profile and semi-profile images constituted around 60% of the total image count. Iris enrollment errors can be a result of using mixed quality grayscale images converted from RGB images, instead of high quality NIR images recommended for the IriCore software. Other problems may be related to: iris image noise, reflections on the eye, lashes, squinted eyes. The voice recognition engine, the CNN for hand images and the DTW method did not perform any quality checks, hence zero FTEs.

The ROC curves together with respective EER values are shown in `Figure` 2. The best single characteristic results can be achieved with face recognition, which allows verification accuracy with EER=2.32%. The worst result was obtained for signatures - skilled forgeries, EER=18.44%. High equal error rate for the irises is related to differences in the quality of the pictures, as high quality pictures obtained with flash enabled are compared against low-quality, no-flash pictures.

VI. CONCLUSIONS

The most important deliverable of this paper is the multimodal mobile biometric database – MobiBits, comprising samples of five different biometric characteristics, collected using three different mobile devices, including a novel type
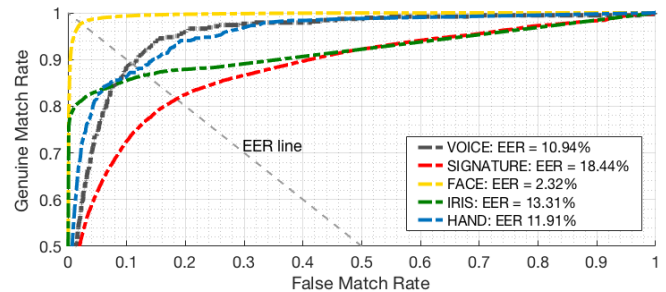

Fig. 2. ROC curves plotted for each individual biometric characteristic. EER values are also shown.

of samples: thermal images of hands and faces, acquired using a thermal sensor equipped mobile phone, which we think can be a valuable contribution for studies involving presentation attack detection methods. A comprehensive set of experiments conducted on the data is reported to show the example benchmark accuracy that can be achieved on the dataset using selected commercial and academic recognition solutions. The MobiBits database is offered publicly at no cost for non-commercial research purposes. We hope that this will constitute a worthy addition to the mobile biometric datasets available for the biometrics community, and will provide an incentive for further research in the field of mobile biometrics.

REFERENCES

[1] K. Raja, R. Raghavendra, V. Vemuri, and C. Busch, "Smartphone based visible iris recognition using deep sparse filtering," *Pattern Recognition Letters*, vol. 57, pp. 33–42, 2014.
[2] K. B. Raja, R. Raghavendra, and C. Busch, "Smartphone Based Robust Iris Recognition in Visible Spectrum using Clustered K-mean Features," *IEEE Xplore*, 2014.
[3] M. Trokielewicz, "Iris Recognition with a Database of Iris Images Obtained in Visible Light Using Smartphone Camera," *IEEE ISBA*, 2016.
[4] M. Trokielewicz and E. Bartuzi, "Cross-spectral Iris Recognition for Mobile Applications using High-quality Color Images," *Journal of Telecommunications and Information Technology*, vol. 3, pp. 91–97, 2016.
[5] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez, "Performance evaluation of handwritten signature recognition in mobile environments," *IET Biometrics*, vol. 3, no. 3, pp. 139 – 146, 2014.
[6] J. Putz-Leszczyska, "On-Line Signature Verification using Dynamic Time-Warping with positional coordinates," *Proc. of. SPIE*, 2016.

[7] B. Van, S. Garcia-Salicetti, and B. Dorizzi, "On using the Viterbi path along with HMM likelihood information for online signature verification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 5, p. 12371247, 2007.

[8] R. Tolosana *et al.*, "Update strategies for HMM-based dynamic signature biometric systems," *IEEE International Workshop on Information Forensics and Security*, 2015.

[9] C. McCool *et al.*, "Bi-Modal Person Recognition on a Mobile Phone: Using Mobile Phone Data," 2012, pp. 635–640.

[10] M. Gunther, A. Costa-Pazo, C. Ding *et al.*, *The 2013 Face Recognition Evaluation in Mobile Environment*, ser. International Conference on Biometrics, 2013.

[11] E. Khoury *et al.*, "The 2013 speaker recognition evaluation in mobile environment," in *2013 International Conference on Biometrics (ICB)*, 2013.

[12] A. de Santos Sierra *et al.*, "Silhouette-based hand recognition on mobile devices," *2009 International Carnahan Conference on Security Technology*, 2009.

[13] D. Zhang, S. Member, W.-K. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Trans. Pattern Anal. Mach. Intell*, pp. 1041–1050, 2003.

[14] Z. Sun, T. Tan, Y. Wang, and S. Z. Li, "Ordinal Palmprint Represention for Personal Identification," ser. CVPR '05. IEEE Computer Society, 2005, pp. 279–284. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2005.267

[15] J. S. Kim, G. Li, B. Son, and J. Kim, "An empirical study of palmprint recognition for mobile phones," *IEEE Trans on Consum Electron*, vol. 61, no. 3, pp. 311–319, 2015.

[16] M. Franzgrote *et al.*, "Palmprint Verification on Mobile Phones Using Accelerated Competitive Code," *International Conference on Hand-Based Biometrics (ICHB)*, 2011.

[17] M. Chora and R. Kozik, "Contactless palmprint and knuckle biometrics for mobile devices," *Pattern Analysis and Applications*, vol. 15, no. 1, p. 37385, 2012.

[18] A. F. Sequeira, J. C. Monteiro, and H. P. O. Ana Rebelo, "MobBIO: A multimodal database captured with a portable handheld device," *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2014.

[19] J. Ortega-Garcia, J. Fierrez, F.Alonso-Fernandez, and et. al., "The Multiscenario multienvironment BioSecure Multimodal Database (BMDB)," IEEE Trans Pattern Anal Mach Intell.2010; 32(6):1097-1111.

[20] A. F. Sequeira *et al.*, "Multi-modal biometrics for mobile authentication," *IEEE International Joint Conference on Biometrics (IJCB)*, 2014.

[21] D.-J. Kim, K.-W. Chung, and K.-S. Hong, "Person authentication using face, teeth and voice modalities for mobile device security," *IEEE Trans on Consum Electron*, vol. 56(4), 2010.

[22] A. F. Abate, M. Nappi, and S. Ricciardi, "I-Am: Implicitly Authenticate Me Person Authentication on Mobile Devices Through Ear Shape and Arm Gesture," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. PP(99), pp. 1–13, 2017.

[23] Neurotechnology, "VeriLook SDK, version 10.0: www.neurotechnology.com/verilook.html," accessed: October 11, 2017.

[24] IriTech Inc., "IriCore Software Developers Manual, version 3.6," 2013.

[25] ISO/IEC 19794-6:2011, "Information technology – Biometric data interchange formats – Part 6: Iris image data," 2011.

[26] S. J. D. Prince and J. H. Elder, "Probabilistic Linear Discriminant Analysis for Inferences About Identity," in *IEEE 11th International Conference on Computer Vision*, 2007.

[27] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector Length Normalization in Speaker Recognition Systems," in *INTERSPEECH*, 2011.

[28] Z. N. Karam, W. M. Campbell, and N. Dehak, "Towards reduced false-alarms using cohorts," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 4512–4515.

[29] M. W. Mak, "Robust Voice Activity Detection for Interview Speech in NIST Speaker Recognition Evaluation," in *Proceedings of the APSIPA ASC 2010*, 2010.

[30] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *https://arxiv.org/abs/1409.1556*, 2014.