Sensing Movement on Smartphone Devices to Assess User Interaction for Face Verification

Chiara Lunerti
School of Engineering & Digital Arts
University of Kent
Canterbury, UK
c.lunerti@kent.ac.uk

Richard Guest
School of Engineering & Digital Arts
University of Kent
Canterbury, UK
r.m.guest@kent.ac.uk

Jon Baker
School of Engineering & Digital Arts
University of Kent
Canterbury, UK
j.baker@kent.ac.uk

Pablo Fernandez-Lopez
University Group for Identification Technologies
University Carlos III of Madrid
Leganés, Spain
pablofer@ing.uc3m.es

Raul Sanchez-Reillo
University Group for Identification Technologies
University Carlos III of Madrid
Leganés, Spain
rsreillo@ing.uc3m.es

***Abstract*— Unlocking and protecting smartphone devices has become easier with the introduction of biometric face verification, as it has the promise of a secure and quick authentication solution to prevent unauthorised access. However, there are still many challenges for this biometric modality in a mobile context, where the user's posture and capture device are not constrained. This research proposes a method to assess user interaction by analysing sensor data collected in the background of smartphone devices during verification sample capture. From accelerometer data, we have extracted magnitude variations and angular acceleration for pitch, roll, and yaw (angles around the x-axis, y-axis, and z-axis of the smartphone respectively) as features to describe the amplitude and number of movements during a facial image capture process. Results obtained from this experiment demonstrate that it can be possible to ensure good sample quality and high biometric performance by applying an appropriate threshold that will regulate the amplitude on variations of the smartphone movements during facial image capture. Moreover, the results suggest that better quality images are obtained when users spend more time positioning the smartphone before taking an image.**

Keywords—biometrics, face verification, mobile devices, sensing data, user interaction

I. Introduction

Face verification has the advantage of being a quick, non-intrusive authentication method that allows users to protect their private and personal information stored on smartphone devices. However, the adoption of facial recognition in a mobile context also brings a challenge: smartphone cameras can be freely moved during the acquisition process, creating unpredictable noise on facial images, which can lower the verification system's performance.

Accelerometer data has been widely used for several applications, particularly related to mobile devices for biometric continuous authentication [1], and behavioural analysis [2] [3]. Since accelerometers are fully integrated into a smartphone, it can embody a useful instrument to monitor the user interaction with the device. As a user is interacting directly with the capturing sensor(s), it is fundamental to monitor their interaction in order to understand how their behaviour might affect sample quality and system performance.

The aim of this research is to present a methodology to assess a user's interaction with a smartphone device during a facial image capture process that may subsequently be used for biometric verification purposes. Our first objective is to understand what type of features can be extracted from the accelerometer data that may give useful information regarding the user's interaction with the smartphone. Furthermore, we analyse how accelerometer data influences biometric scores and quality metrics obtained from the images taken for the authentication.

Our study conducted a data collection, recording sensor information from 53 participants while they were capturing "selfies" suitable for biometric verification on an Android smartphone device. Subsequently we have performed a statistical analysis to investigate how sensed movements can affect sample quality and biometric scores.

Results from the analysis may be used for biometric application developments. For example, the information extracted from an accelerometer can be a prediction, in real-time, of low-quality images due to excessive movement. This, in turn, may lead to a request to re-present the biometrics due to a poor score. Furthermore, it can be used to adjust a biometric match acceptance threshold, which can be lowered depending on the amount of noise expected in the images.

This paper is organised as follow: a brief illustration of previous work on this area is presented in Section II. The data collection and methodology are described respectively in Section III and Section IV. Results are presented in Section V,

978-1-5386-7931-9/18/$31.00 ©2018 IEEE

and finally, conclusions and future work are indicated in Section VI.

II. Background

To enhance the performance of a face recognition system, many systems consider rejecting low-quality images from the ones collected for the verification that present good sample quality in terms of brightness and pose symmetry. In the study presented by Boontua et al. [4], the authors analysed the performances of facial recognition system under different light conditions and considering three different user poses: frontal, left side and right side. The authors compared different features and classifiers to establish which methodology was more accurate depending on the pose and light variations. Their results showed that similar illumination conditions for enrolment and verification result in a higher accuracy. In a situation where the enrolment and verification had been taken into a different environment and different user pose, Local Binary Patterns features, and a Support Vector Machine classifier recorded higher accuracy compared to the other combinations of feature extraction and classification.

In mobile scenarios, the acquisition process of facial images for authentication is complex, as the users move as well as the smartphone camera. Background accelerometer data can be employed to analyse the variation in movements of smartphone and user. Researchers have adopted the accelerometer to study user's behaviour in real life scenarios for many applications such as home security and healthcare. The authors in [2] designed a model to recognise a set of daily activity studying the three-axial accelerometer data collected from four volunteers under real-world conditions. The classification method described by the authors recognised the activity tasks with an accuracy up to 91.15%, independently on the smartphone positions, whether the smartphone was kept in the users' hand or in their pockets.

Accelerometer data can be used to continuously authenticate the user on their smartphone in combinations with other behavioural biometrics, such as swiping and keystroke dynamics, or with traditional modalities such as voice and face recognition. In the study presented by Crouse et al. [5], accelerometer data had been employed in fusion with face recognition system to unobtrusively authenticate the user on the device and to enhance the matching performances. Image uprightness correction was performed to the images that were taken by the users before performing the authentication, resulting in 6% higher performances at 0.1% False Accept Rate (FAR) compared to the authentication performed with the original images. When using the continuous authentication application, 24 subjects took part in the experiment carried out by the authors, and they retained the access on the smartphones for over 96% of the trials for the whole duration of the test (15 minutes).

In our study, we aimed to analyse user's behaviour using the three-axial accelerometer data recorded during the presentation of facial images for verification under different environmental conditions to enhance the experience of the user and the performances of the verification system in a mobile context.

Fig. 1. A visual representation of pitch, roll, and yaw around the smartphone's axis.

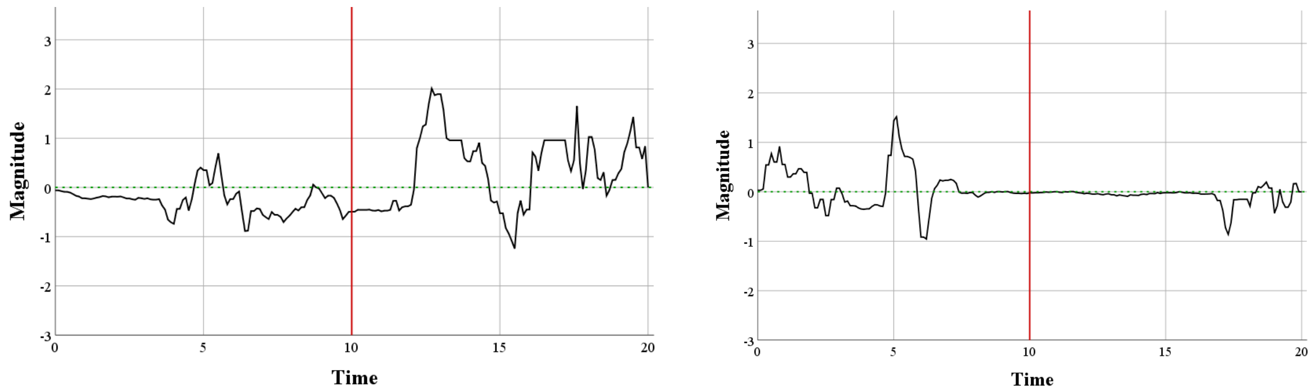
III. Data Collection

We conducted a data collection that recorded smartphone sensor data in the background for a total of 53 participants, while they were taking images for the purpose of face verification, using the front facing camera. The experiment consisted of 3 sessions, where accelerometer data was recorded in the background for the entire duration. For each 30 minute session, participants were provided with a smartphone (a Google Nexus 5) and a map which contained 10 locations in a sequential order which participants had to follow by walking between locations. At each location, participants were instructed to proceed with the acquisition of facial images on their smartphone device, which would be suitable for biometric authentication. There was a minimum requirement of 5 images for each location, but participants were free to take more images if they wanted. Locations were both indoors and outdoors and were selected to represent real-life scenarios where smartphones are commonly used. Moreover, the locations were selected due to their unique and diverse environmental attributes, where lighting conditions and other background noise would vary. To collect the images and the sensing data in the background, we developed an Android app which connected to a sensing API and automated the data collection process.

IV. Methodology

To understand the effect of user interaction over quality and system performance, we extracted features from the accelerometer data: we combined the information from the acceleration on each axes and we calculated the magnitude and angular accelerations for pitch, roll and yaw (Fig.1).

We also selected three quality metrics from the ISO/IEC Technical Report 29794-5:2010 [6] and we calculated them for all the images collected during the experiment. We then selected from all the images collected, two random sets of 5 images for enrolment in two different scenarios (indoors and outdoors). We consider all the remaining images as verification dataset and obtained biometric scores from the comparisons between the verification images and the images from the two enrolment sets.



(a) (b)

Fig. 2. Gait movements in a 10-second window before and after an image was taken. The graphs shown two different scenarios, one where a user was still moving or had not stopped completely before taking an image (a) and the other one where a user had stopped or recorded little movement while taking an image (b).

A. Accelerometer Data

Time-stamp and accelerometer data was recorded with a sampling frequency of 10 Hz for each image collected. We applied a low-pass filter of 0.8 Hz and segmented the signal using three window sizes of 1, 3, and 5 seconds before and after each image was taken. We then extracted features that could be used to analyse user and smartphone movements. First, we calculated the magnitude for each image using (1), where M is Magnitude, and, A_x, A_y, and A_z are the directional accelerations on each smartphone's axes [7].

$$M = \sqrt{A_x^2 + A_y^2 + A_z^2} \tag{1}$$

Magnitude can give us information about gait movements, whether the users were walking or still moving when capturing the images. When the signal does not present any variation, it means that the user stopped walking and is not moving, or is performing minimal movements with the smartphone before the authentication. An example of the two scenarios is presented in Fig.2. From the three different selected time windows, we observed that the overall trend of the magnitude presented peak-to-peak amplitudes within the range of ±3 m/s^2. We empirically selected three thresholds: 0.5 m/s^2, 0.75 m/s^2, and 1 m/s^2. We considered as magnitude features the number of peaks in the signals and the amplitude of their variations when they were over the three selected thresholds. We then combined accelerations from each axis to obtain pitch ρ (the angle around the x-axis) using (2), roll φ (the angle around the y-axis) using (3), and yaw θ (the angle around the z-axis) using (4), where A_x, A_y and A_z are the measured acceleration data from the accelerometer module for the specific axes [8].

$$\rho = \tan^{-1}\left(\frac{A_x}{\sqrt{{A_y}^2 + {A_z}^2}}\right) \tag{2}$$

$$\varphi = \tan^{-1}\left(\frac{A_y}{\sqrt{{A_x}^2 + {A_z}^2}}\right) \tag{3}$$

$$\theta = \tan^{-1}\left(\frac{\sqrt{{A_x}^2 + {A_y}^2}}{A_z}\right) \tag{4}$$

These features were selected to represent the movements which a user performed while adjusting themselves in front of the smartphone in certain lighting conditions.

B. Image Quality

To assess quality, we segmented each face from the captured images and we calculated quality metrics following recommendations in ISO/IEC Technical Report (TR) 29794-5:2010 [6]. We selected three facial image metrics among the ones that are commonly used in the state of the art: brightness, contrast, and blurriness. Image brightness was calculated as the mean of the pixels intensity values. Contrast was calculated from the histogram as the difference in luminance of the object in the image as described in the TR. The level of blurriness was calculated with a range of 0 to 1 where 0 means sharp and 1 means blurry following the indication in [9].

C. Biometric outcomes

To simulate real-life scenarios, all images were taken in an unconstrained environment [10]. For each participant, we considered enrolment under two different environmental conditions: enrolment selecting 5 random images from all the ones taken indoors and enrolment selecting 5 random images taken when outdoors. All the remaining images were considered as the participant's verification dataset (indoors and outdoors verification scenarios). We performed a biometric verification using a commercial state of the art recognition system [11] and for each comparison, we obtained a biometric score (ranging from 0 to 500), as an average of the scores obtained when comparing the verification image with the 5 enrolment images.

V. RESULTS

From the analysis, we noticed that there was not a substantial difference between the variations of magnitude smaller than 0.5 and 0.75 m/s2 and when the variations were calculated under the threshold set to 1 m/s2. For this reason, we decided to present our result only in the case of selecting 1 m\s2 as a threshold as an example of a more permissive approach.

Firstly, we observed how the quality metrics selected are correlated with the variations recorded with the accelerometer data for gait movements and the angular rotations. Image brightness and contrast did not appear to have a correlation with the number of peaks presented in the gait signals, nor with the amplitude of these variations. On the contrary, images taken with a lot of variations in the magnitude resulted in less blurriness of the images. This was particularly observed in the scenario in which the images have been taken in indoor locations presenting correlation coefficients of $r = -0.059$ when the variations were calculated in an interval of 1 second before taking the image, $r = -0.076$ when considering 3 seconds before, and $r = -0.075$ when considering 5 seconds before ($n = 3438$, $p < 0.001$).

When assessing the amplitude of the acceleration movements recorded in the magnitude that was bigger than 1 m/s2, blurriness presented the strongest correlations when the images were taken indoors presenting $r = 0.228$ when considering 1 second before the image was taken, $r = 0.216$ with a 3 seconds window before taking the image, and $r = 0.195$ with a window of 5 seconds before taking the image ($n = 9246$, $p < .001$). Recording a number of small frequent movements in the magnitude resulted in better quality performance compared to recording a few movements that were bigger in magnitude.

We also observed a linear negative correlation between the movements recorded for the angular rotations of the smartphone pitch, yaw and roll and the level of brightness of the images. The correlation was noted especially from movements recorded for roll angles, in particular when images were taken indoors, although they were not significantly strong. Despite pitch, roll, and yaw movements affecting the level of blurriness of the images, it appeared that there is no correlation between angular rotations and the biometric scores. We performed a Spearman's correlation between the amplitudes recorded in the movements in the magnitude and the number of variations with the biometric scores for different environmental conditions for both types of verification and enrolment.

When considering the number of variations, the correlation appears to be positive and it is stronger for the verification images taken outdoors for both of the enrolment scenarios. Taking a window image of only 1 second before and after, the correlation result as $r = 0.082$ for enrolment indoors and $r = 0.128$ for enrolment taken outdoors ($n = 4590$, $p < 0.001$). There is a smaller difference between the 3 seconds and 5 seconds windows, where the coefficient it around $r = 0.106$ for both when the enrolment is indoors and around $r = 0.174$ when enrolment is taken outdoors ($n = 4590$, $p < 0.001$). There was a positive correlation, but not particularly strong for when the verification images were taken indoors. Significant correlation coefficients for the amplitude of movements when considering 1 second (*A_1s*), 3 seconds (*A_3s*), and 5 seconds (*A_5s*) before and after the image acquisition and biometric scores are reported in Table I.

TABLE I. SPEARMAN'S CORRELATIONS

Enrolment	Verification	Correlation			
		Spearman's rho	***A_1S***	***A_3S***	***A_5S***
Indoor	Indoors	Correlation Coefficient	-.288**	-.293**	-.280**
		Sig. (2-tailed)	.000	.000	.000
		N	2983	2984	2984
Outdoor	Indoors	Correlation Coefficient	-.246**	-.228**	-.213**
		Sig. (2-tailed)	.000	.000	.000
		N	2983	2984	2984
Indoor	Outdoors	Correlation Coefficient	-.169**	-.154**	-.135**
		Sig. (2-tailed)	.000	.000	.000
		N	4590	4590	4590
Outdoor	Outdoors	Correlation Coefficient	-.305**	-.310**	-.270**
		Sig. (2-tailed)	.000	.000	.000
		N	4590	4590	4590

**. Correlation is significant at the 0.01 level (2-tailed).

Regardless of indoor or outdoor environment, there is a strong negative correlation between the amplitude of the movements and the biometric score for each enrolment scenario: enrolment images taken in a controlled environment ($r = -0.644$, $p = 0.001$), taken indoors ($r = -0.538$, $p = 0.006$) and outdoors ($r = -0.581$, $p = 0.002$).

VI. CONCLUSIONS AND FUTURE WORK

The aim of this study was to assess a user's interaction with a smartphone face verification system using the three-axial accelerometer data. From the analysis, it is possible to investigate the variation in magnitude and angular rotations to assess whether a person is walking or moving to set the smartphone for facial image capture. The number of the variations and their amplitude can be screened with a threshold.

From the results it is shown that a high amplitude of variations in the magnitude and the angular rotations lower the sharpness of the image among the metrics considered to assess facial image quality. Equally, the amplitude of movements recorded by the accelerometer result in a lower biometric performance. When considering the number of variations in the magnitude, the greater the number of variations recorded by the accelerometer data, the better the performance and the sharpness of the image.

Our intuition is that a number of movements registered few seconds before taking the images in the magnitude represent the movements that participants performed with the device to find the right positioning for the smartphone, and subsequently take better quality images for verification. Future research will assess the different type of movements recorded by the accelerometer data to find a possible way to filter the movements produced by the walking gait of the users while taking the images from the movements recorded by the accelerometer to position the smartphone when taking the images.

Assessing a user's interaction with a smartphone in unconstrained environments allows the enhancement of a biometric system's performance for face verification, particularly when using an accelerometer sensor as the primary data source. The applicability of this study can be used to provide feedback in real-time to a user during the verification process to enhance biometric system performance.

Pitch, roll, and yaw can give a prediction of the quality of captured images. To be able to use the angular rotation, the best way is to combine them with other sensor data, such as the magnetometer. Future research will use the outcome of this study to compare evaluation scenarios, and consider the implementation of accelerometer data with a fusion of other sensor data to get a more accurate user interaction.

ACKNOWLEDGMENTS

This work was carried out with funding from the University of Kent 50th anniversary scholarship.

REFERENCES

[1] Nickel, C. and Busch, C., "Classifying accelerometer data via hidden markov models to authenticate people by the way they walk," IEEE Aerospace and Electronic Systems Magazine, vol. 28, no. 10, pp. 29-35, 2013.

[2] Bayat, A., Pomplun, M. and Tran, D.A., "A study on human activity recognition using accelerometer data from smartphones," Procedia Computer Science, vol. 34, pp. 450-457, 2014.

[3] Ellavarason, E., Guest, R., Deravi, F., "A Framework for Assessing Factors Influencing User Interaction for Touch-based Biometrics," in Signal Processing Conference (EUSIPCO), 2018 26th European, in press.

[4] Boontua, M., Nam-asa, P., Arwatchananukul, S. and Aunsri, N., "A study of features and classifiers for multiple environment face recognition system," Digital Arts, Media and Technology (ICDAMT), 2018 International Conference on, pp. 249-253, 2018.

[5] Crouse, D., Han, H., Chandra, D., Barbello, B. and Jain, A.K., "Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data," Biometrics (ICB), 2015 International Conference on, pp. 135-142, 2015.

[6] "ISO/IEC TR 29794-5:2010 - Information technology -- Biometric sample quality -- Part 5: Face image data," [Online]. Available: https://www.iso.org/standard/50912.html.

[7] Fernandez-Lopez, P., Sanchez-Casanova, J., Tirado-Martín, P. and Liu-Jimenez, J., "Optimizing resources on smartphone gait recognition," in Biometrics (IJCB), 2017 IEEE International Joint Conference on, 2017.

[8] K. Tuck, "Tilt Sensing Using Linear Accelerometers," in Freescale Semiconductor Application Note AN3107.

[9] Crete, F., Dolmiere, T., Ladret, P. and Nicolas, M.,, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in Human vision and electronic imaging XII, 2007.

[10] Lunerti, C., Guest, R.M., Blanco-Gonzalo, R., Sanchez-Reillo, R. and Baker, J., "Environmental effects on face recognition in smartphones," Security Technology (ICCST), 2017 International Carnahan Conference on, no. IEEE, pp. 1-6, 2017.

[11] "NEUROtechnology VeriLook 10.0 Standard SDK," [Online]. Available: http://www.neurotechnology.com/verilook.html.