Seeing Red: PPG Biometrics Using Smartphone Cameras

Giulio Lovisotto, Henry Turner, Simon Eberz and Ivan Martinovic
University of Oxford, UK

`name.surname@cs.ox.ac.uk`

Abstract

In this paper, we propose a system that enables photoplethysmogram (PPG)-based authentication by using a smartphone camera. PPG signals are obtained by recording a video from the camera as users are resting their finger on top of the camera lens. The signals can be extracted based on subtle changes in the video that are due to changes in the light reflection properties of the skin as the blood flows through the finger. We collect a dataset of PPG measurements from a set of 15 users over the course of 6-11 sessions per user using an iPhone X for the measurements. We design an authentication pipeline that leverages the uniqueness of each individual's cardiovascular system, identifying a set of distinctive features from each heartbeat. We conduct a set of experiments to evaluate the recognition performance of the PPG biometric trait, including cross-session scenarios which have been disregarded in previous work. We found that when aggregating sufficient samples for the decision we achieve an EER as low as 8%, but that the performance greatly decreases in the cross-session scenario, with an average EER of 20%.

1. Introduction

Biometric authentication is a popular approach to effortless user authentication, especially on mobile devices such as smartphones or tablets. While both fingerprint scanning and face recognition offer quick and accurate authentication, their modalities are also easy for attackers to observe. At the same time, reliably detecting forged biometric samples (such as latex fingers) remains an arms race with attackers crafting increasingly sophisticated forgeries. In recent years, the photoplethysmogram (PPG) has become particularly interesting in the context of mobile devices. PPG is a method to optically detect changes in blood flow volume and is often used in optical heart rate sensor of smartwatches. In the medical domain, PPG is used to monitor patients' heart rates and blood oxygen levels through pulse oximeters. Similarly to the electrocardiogram (ECG), PPG offers a range of distinctive biometric features derived from
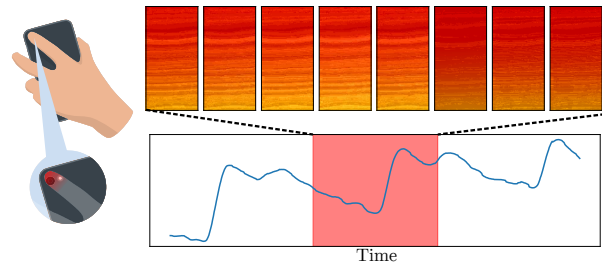

Figure 1: System Overview. Users place their finger on the smartphone camera. The blood flow to the fingertip changes the light reflection properties of the skin creating subtle changes in the color of the video frames (luma component).

unique characteristics of an individual's cardiovascular system. Previous research in the field has relied on medical datasets or purpose-built PPG sensors. However, using these systems is not practical due to the specialized hardware required (for example, users are unlikely to clamp a pulse oximeter on their finger for authentication).

In this work, we are proposing the first system to collect a user's PPG through a mobile phone camera and authenticate them through the extracted signal. During a measurement, the user places their finger on the phone's camera while the finger is illuminated by the phone camera flashlight (see Figure 1). Smartphone-based PPG authentication has two main advantages over established biometrics: Unlike fingerprints or facial images, PPG is hard to observe from a distance, making it less susceptible to presentation attacks. In addition, the hardware requirements are low and no specialized sensors are needed. This makes the technique even suitable for low-cost feature phones. We validate our approach through a user study with 15 participants and capture their PPG over 6-11 sessions. The multi-session nature of our experiment allows us to explore the time stability of the biometric, a component largely overlooked by previous work. In order to facilitate future work in the field, we make our code and the dataset available online[1].

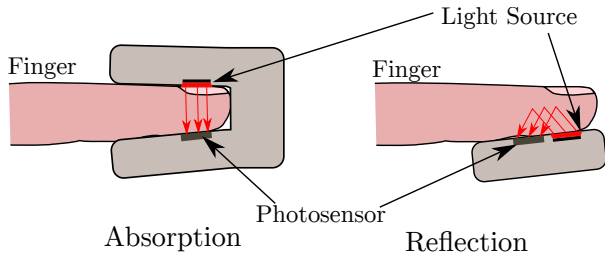[1] https://github.com/ssloxford/seeing-red/

Figure 2: The two modes of PPG collection. The absorption mode is usually used by devices in a medical setting, such as pulse oximeters, whereas the reflection mode is used in smartwatches and other consumer systems. Our system operates in reflection mode.

2. Background and Related Work

In this section, we will present the physiological foundations of PPG-based authentication, discuss different measurement methods and summarize related work.

2.1. Physiological foundations

A photoplethysmogram (PPG) is an optically obtained plethysmogram that can be used to detect blood volume changes in the microvascular bed of tissue [29]. Having first been proposed in 1938, PPG is a relatively simple technique that only relies on a light source to illuminate the skin tissue, and a photo detector to measure the small variations in light intensity associated with changes in perfusion in the catchment volume [21]. Within each heartbeat three distinct fiducial points can be identified that relate to different stages of the cardiac cycle (see Figure 4). Some biometric features used in this work (see Section 4.7) are directly based on these fiducial points and derive their distinctiveness from the uniqueness of the person's cardiovascular system.

The systolic peak is a result of the direct pressure wave traveling from the left ventricle to the periphery of the body [5]. Its amplitude relates to stroke volume [23] and is used to estimate continuous blood pressure [7]. The diastolic peak (or inflection) is a result of reflections of the pressure wave by arteries of the lower body [5]. The time difference between the systolic peak and diastolic peak has been previously used as a measure of large artery stiffness [22] The dicrotic notch is a small and brief increase in arterial blood pressure that appears when the aortic valve closes and is commonly used as an equivalent of end-systolic left ventricular pressure [8]. A more in-depth review of the PPG signal morphology and the medical reasons for the distinctiveness of individual components can be found in [11].

2.2. Measuring PPG

In a clinical setting, PPG is typically measured with a pulse oximeter, usually clamped to a person's finger, and consists of a light source on the top and a photosensor on the bottom, as shown in Figure 2. The sensor then measures changes in light absorption, allowing it to distinguish between oxygenated and deoxygenated blood. A PPG sensor used in absorption mode must be located on the body at a site where transmitted light can be detected [25]. Consequently, measurement sites are limited to the extremities of the body, such as the fingertips or earlobes.

In recent years, PPG sensors have become increasingly popular in wearable devices (smartwatches in particular) to allow continuous heart rate measurements without the need for a chest strap. PPG-enabled watches use a set of LEDs to illuminate the skin tissue and a photosensor to capture the resulting reflection, as seen in Figure 2. The key difference to pulse oximeter measurements is that the light source and photosensor are adjacent. This is required as the absorption mode is not feasible on the wrist.

2.3. PPG for Authentication

While PPG was first proposed in 1938, it has only been used for authentication since 2003. Work in this area differs in terms of measurement device (particularly reflection vs absorption mode), sample size, feature types and consideration of feature stability over time. A summary of the field with regard to these properties can be found in Table 1.

Most early work in the field uses relatively simple, low-dimensional feature sets [12, 13], focusing on the number of peaks and the slope of parts of the waveform. Bao et al. use heart rate variability (HRV) derived from PPG as a sole feature [2].

More recent work has significantly extended the set of features to more comprehensively capture the distinctiveness of each constituent part of the PPG waveform and incorporates non-fiducial methods in order to avoid the difficulty of peak detection in noisy signals [16]. The latter is particularly useful for the reflective measurement mode where the the diastolic peak and dicrotic notch are often hard to identify.

Unlike earlier work, which collected data from a single measurement session, Kavsaoğlu et al. conducted three recording sessions with 30 participants [18]. In order to compare intra-session and cross-session performance, they report individual results for each of the sessions and for a combined dataset consisting of both sessions. The authors achieve identification rates of 90.44% and 94.44% for single-session tests and 87.22% for the combined dataset. However data from the sessions have not been separated in the evaluation, which positively biases the results compared to a real-world scenario [10].

The most important difference in our work is the measurement device. Previous work relies on either purpose-built hardware or medical-grade devices, whereas we use readily available consumer hardware. In addition, most re-

Ref	Measurement	Subjects	Dataset	Sessions
[13]	reflection	17	own	1
[12]	reflection	17	own	1
[2]	unknown	12	own	1
[33]	absorption	3	own	1
[30]	absorption	14/15	[1]	1
[4]	absorption	44	own	1
[18]	reflection	30	own	3
[28]	absorption	23	[19]	1
[16]	absorption	42	[17]	1
[15]	absorption	42	[17]	1
[31]	absorption	20	own	1
Our work	reflection	15	own	6-11

Table 1: Related work on PPG-based authentication.

search uses single-session data and performs training and testing within the same session. Due to the single-session nature of most experiments, it is possible that systems distinguish measurement sessions, rather than individuals. For example, PPG measurements are affected by skin temperature [14] and sensor contact pressure and position [6], both of which can easily introduce artifacts into the signal that would only be present in this recording.

3. Experiment Design

In this section we principally present our data collection method, which we subsequently use for developing our authentication systems.

3.1. Collection Apparatus

As previously mentioned, a key contribution of our work is the ability to use PPG for authentication without purpose-built hardware. As such, our system makes use of a smartphone's camera and camera flash to act as the photosensor and light source respectively. One challenge in this setup is presented by the camera light: commercial PPG sensors in reflection mode use green light as its wave length leads to ideal penetration depth [21]. Conversely, the phone camera uses white light, which leads to noisier measurements. We collect the PPG data using a custom iOS application, running on an Apple iPhone X.

Users are instructed verbally to hold the phone in their dominant hand, and place their finger lightly over both the camera and camera flash. They are instructed to sit down whilst performing the capture, and to remain as still as possible throughout; the application automatically stops recording if too much movement is detected i.e total acceleration exceeds 1.3g ($12.75m/s^2$).

Additionally, we fix the camera ISO and exposure time to be the minimum possible, given the camera FPS and resolution settings. Likewise, we set the white balance gain to be the maximum in the red channel, and minimum in the blue and green channels. By fixing these settings we ensure consistency between capture sessions, reducing noise in the signal and increasing the information captured about the color changes.

A single capture session, as taken by the application, consists of a video recorded in 1280 x 720 resolution at 240 frames per second, for a duration of 30 seconds. The video is resized to 360 x 240 and uploaded to a server for validation. After each capture session, the researcher supervising the data collection takes the phone away from the user and returns it to them for any subsequent capture sessions, leading to users placing their finger on the camera differently each time. A validation function analyses the red light channel of the video once it has been uploaded to the server, and rejects the video if there are sudden jumps or a lack of red light, as these are likely caused by the user lifting their finger from the camera.

3.2. Collection Procedure

We collect data for a total of 15 participants, performing between 6 and 11 capture sessions (measurements) per participant. There were 13 male and 2 female participants. Of the participants, 13 identify as white, one as Indian and one as South-East Asian. Each participant is handed the device running the application and given the above instructions on how to use the application, as well as being given a visual prompt towards correct usage on the screen. Captures for each participant were taken across multiple sittings, with no more than three capture sessions occurring in the same sitting. Sittings were taken approximately a day apart, with a median of 32.1 hours between any pair of capture sessions.

3.3. Research Ethics

Before the data collection, we obtained informed consent from the participants. Our data collection process and experiment took place with ethical approval from our institution: reference SSD/CUREC1A CS_C1A_19_032.

4. Method

Our method consists of a pipeline of steps which transform the raw video captured by the camera into biometric samples which can be used for authentication. We design the pipeline with the aim of running the system in real-time, meaning that all processing steps can be run as the video is being captured (i.e., we never use processing that requires looking at a whole video file, only at the video recorded up to that point).

4.1. Signal Extraction

To obtain the signal from the raw video, we compute the mean of the pixel-wise luma component from the pixels in each video frame, so that if F is a video composed by a
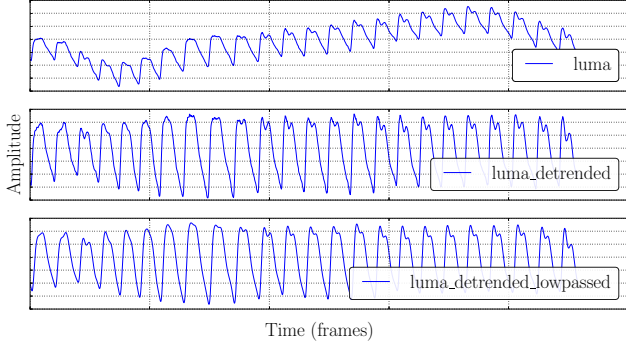
Figure 3: Signal over the course of the preprocessing steps. The first plot shows the luma component extracted by the video frames (Eq.1). The second plot shows the signal after de-trending with rolling average and the third plot shows the signal after the low-pass filtering, which is used in the rest of the pipeline.

sequence of frames $\{f_1, ..., f_m\}$, then the signal originating from F is:

$$S = \{Y(f_1), ..., Y(f_m)\}, \text{where}$$

$$Y(f) = \frac{1}{n} \sum_{i,j \in f} [0.299 f_{i,j}^{(r)} + 0.587 f_{i,j}^{(g)} + 0.114 f_{i,j}^{(b)}]. \quad (1)$$

In Equation 1, i and j iterate over the pixels of the image and the superscripts $_{(r)}$ indicate the considered RGB channel of the frame, either red green or blue. The channel co-efficients of Equation 1 are taken from the ITU-R BT.601 standard.[2]

4.2. Signal Preprocessing

After the signal from a video is extracted, we apply the following preprocessing steps to remove noise. First, in order to remove trends from the signal we compute and subtract a rolling average of the signal with the signal itself. We use a window size of 1 second for the rolling average. Then, we use a low pass filter to remove high frequency noise, with cutoff frequency at 4Hz, i.e., 240beats per minute. See Figure 3 for a visualization of the processed signals.

4.3. Beat Separation

As mentioned above, in our approach we want to perform single beat authentication, therefore we design a beat detection algorithm to separate individual beats in the signal. We report in Algorithm 1 the procedure to obtain individual beats from a preprocessed signal. Alg. 1 first finds the minimums of a smoothed version of the signal (moving average, Line 3), and then matches those minimums to the minimums in the original signal. The $\arg \mathrm{rel\ min}$ instruction extracts all relative minimums indexes from the signal.

Algorithm 1 - Beat Separation. Given a signal detects individual heartbeats boundaries.

1: **Input:** preprocessed signal S, max_bpm U, sampling rate fps, smoothing window size ws.
2: $g = 60 \cdot \frac{fps}{U}$ ▷ minimum inter-beat gap
3: $A = \mathrm{moving\_average}(S, ws)$
4: $mins^{(S)} = \arg \mathrm{rel\ min}\ S$
5: $mins^{(A)} = \arg \mathrm{rel\ min}\ A$
6: $B = \{\}$ ▷ Individual Beats
7: **for** $i \in mins^{(A)}$ **do**
8:      $S_i = \{s_{i-g}, ..., s_{i+g}\}$
9:      $j = \underset{i-g,...,i+g}{\arg \min}\ S_i$
10:      $B = B + \{j\}$
11: **end for**
12: **return** B

We found that this algorithm leads to a more reliable beat separation, as the signals often presents random noise which survives the preprocessing above.

4.4. Fiducial Points Detection

After individual beats have been extracted, we detect fiducial points in the signal, which are used to extract features. We focus on three points in particular: (i) systolic peak, (ii) dicrotic notch and (iii) diastolic peak. In low-noise signals, these points are easily found by finding maximums and minimums in the signal and its first derivative, see Figure 4. However we found that this procedure needs to account for noisy signals, so we craft a more robust algorithm to detect them which falls back to best guess points.

4.5. Beat Signal Quality

Oftentimes, slight hand or finger movements may cause noise in the signal which bypasses our filtering pipeline. We therefore design a set of quality criteria for individual beats with the goal of excluding noisy beats from further processing. These criteria are the following:

- **max bpm**: all beats which correspond to > 120 bpm, as these are caused by noise which bypasses the separation of Alg. 1;

- **number of peaks**: all beats with more than three distinct maximums;

- **distance from reference**: all beats that are too different from a reference expected beat wave. We compute this difference using dynamic time warping [27] between the current beat and the reference and we set a threshold of 2.0, i.e., beats with distance higher than 2.0 are discarded. The reference wave is the average beat wave in our dataset.
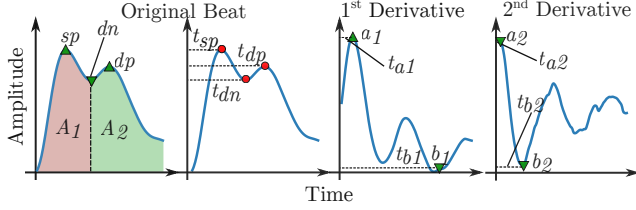
Figure 4: Visualization of fiducial point based features for a single beat. Time span features are based on time from start of an individual heartbeat, and are normalized to account for different bpm.

All beats that match at least one of these conditions are flagged as failed to acquire (FTA). Using these quality filters leads to better authentication accuracy at the cost of slightly longer authentication time. In the analysis, we report the results including and excluding these beats.

4.6. Feature Extraction

We consider four groups of features: statistical, curve widths, frequency domain and fiducial points based. To make features independent of a user's bpm at the time of capture, after the physiological features are computed, we re-sample beats onto a fixed sampling rate of 1,000 Hz and we normalize them so that the amplitude values lie in [0, 1]. Features are computed on a beat-basis, so that each beat leads to a sample.

Statistical. These are straightforward statistical features that are based on the physiological traits of the users and for a single beat include: (i) maximum value, (ii) minimum value, (iii) the difference max-min and (iv) the length of the beat. Given that our camera settings are fixed, features (i)-(iii) are a result of the absolute values found in the luma component of the video frames, which are related to skin tone and amount of blood flowing through the finger. Feature (iv) is just related to the individual's heart rate at rest, which has some variability, but we found it to be a distinctive features in our case.

Curve Widths. These features are the widths of the curve defined by the beat at a set of pre-defined heights. For a certain height $h \in [0, 1]$, where 1 is the maximum height and 0 the minimum, we consider the curve that lies above the threshold h and compute the distance between the extremes of this curve, see Figure 4. We choose 18 different heights evenly spaced in [0.05, 0.95] and compute one width feature for each of them.

Frequency Domain. For each beat we compute the discrete Fourier transform and use it as features. Section 4.7 discusses how we reduce the dimensionality of these features with feature selection.

Fiducial Points. We add a set of features based on the loca-

tion of the fiducial points discussed above based on related work, see Figure 4 for examples of the features used.

4.7. Feature Selection

After feature extraction, each beat is represented by a total of 541 features, 4 physiological, 18 curve widths, 500 frequency domain and 19 from the fiducial points. To reduce dimensionality and remove redundant features, we apply multi step feature selection.

Step 1. We use principal component analysis (PCA) [32] to reduce the number of features in the frequency domain and the curve width groups. For the frequency features, we fit a PCA with 100 components and retain only the first n components so that these n components describe 99% of the space variance. For the curve width features, we do the same but fit a PCA model with 15 components. We find that roughly 5 components are necessary for the frequency group (depending on the dataset) while 9 are sufficient for the width group.

Step 2. In this step, we combine two different techniques for feature selection on the remaining features. At first, we compute the pairwise correlation coefficient between features and for a pair of feature distributions (f_1, f_2) we drop one feature from the dataset at random if the correlation coefficient between them is $r_{f_1, f_2} > .95$. Then, to avoid the effect of outliers, we remove all feature values which lie outside the 1st and 99th percentile of the feature distribution. Then we use minimum redundancy maximum relevance (mRMR) feature selection [24] to select the top 60% performing features (differently from the original paper, we choose the Mutual Information Quotient criteria in the algorithm). Afterwards, we select the top 60% features ranked on their relative mutual information (RMI). We use the non-parametric approach to estimate RMI introduced in [26] to avoid imprecisions arising from modeling continuous distributions with discrete bins. Finally, we choose all the features that pass both selections (RMI and mRMR) to create our final feature set. Section 5 presents and discusses the selected features set.

5. Results

In this section we present the results for a set of authentication use-cases.

Preliminaries. In each use-case, we mainly monitor the equal error rate (EER) of the authentication, that is the point at which False Reject Rate (FRR) and False Accept Rate (FAR) are equal. In each experiment, rather than using the classification output from the classifier as the decision (accept/reject), we use the distance from the decision boundary (or prediction probability) to make the classification decision, i.e., we list all distances coming from the positive (the

Feature	mRMR (MIQ)	RMI
A_2/A_1	1.29	0.14
t_{b_1}	1.11	0.13
b/a_2	1.11	0.10
b_2	1.02	0.13
max	1.02	0.17
min	0.88	0.23
a_2	0.86	0.12
t_{b_2}	0.82	0.12
t_{sp}	0.77	0.12
fft_{PCA_0}	0.75	0.09
length	0.72	0.12
fft_{PCA_1}	0.70	0.08
t_{dn}	0.69	0.15
t_{a_2}	0.67	0.08
fft_{PCA_5}	0.66	0.08
b_1	0.64	0.12
t_{a_1}	0.64	0.10
A_1	0.61	0.11

Table 2: Features retained and their importances after the feature selection step, for both ALL and post-FTA datasets. The importances reported are computed in the ALL dataset.
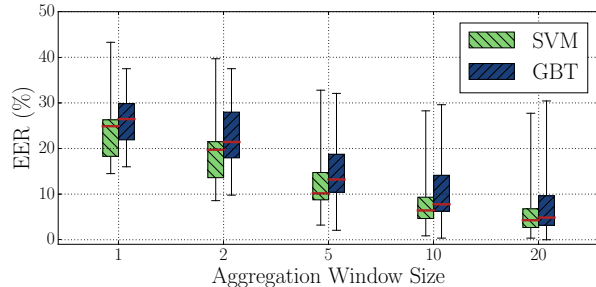

Figure 5: EERs in the multi-class case, for various classifiers and aggregation window sizes. Values are averaged over the two folds before plotting. Whiskers show the range of the data, horizontal lines in boxes show the median.

user) and negative (other users) samples, and find the EER by changing the threshold.

Aggregation. We report results for aggregating the decision on multiple samples, i.e., we choose an aggregation window size n, collect n samples and compute the decision based on the aggregated samples. The aggregation function we use is the mean.[3] To gather more data points, whenever we aggregate samples together, we randomly draw sample sequences of length n for 100 times from the positive (genuine) class. For a certain user, to construct the negative (impostor) class, we randomly draw sample sequences of length n for 10 times for each other user. This way, each EER value for a user is computed using $100 + 10 \times (|U| - 1) = 240$ authentication attempts.

At the end of our data collection, we obtain a dataset of a total of 3,836 samples (one sample equals one beat) with at least 178 samples per user. Of the samples, 3,529 (92%) pass the beat signal quality checks discussed in Section 4.5. Throughout the evaluation, we refer to these two separate datasets as **ALL** and **Post-FTA** and we report results for both. After feature selection, we retained the top 18 features for both ALL and Post-FTA, see Table 2 for details.

5.1. Multi-class Case

Setup. We randomly split the data into two stratified folds and train two different classifiers, support vector machine (SVM) with a radial basis function kernel and gradient boosted trees (GBT). For SVM we use a standard scaler to

---

[3]We also tried median but it performed worse on average.

normalize the data before feeding it into the classifier, and only use the training part to fit the scaler.

Results. We report in Figure 5 the obtained EERs. We found that single sample EER lies around 15%, but decreases quickly to less than 5% when using an aggregation window size > 10. Out of the tested models we found that SVM performs the best, with an EER of $<1\%$ at aggregation window size of 20.

5.2. One-class Case

Setup. As a PPG-based authentication system leveraging the smartphone camera would reside on the user's device, the system would likely only have access to the legitimate user's data (provided at enrolment). We therefore consider the case where the training algorithms do not have access to any negative class data. We use two algorithms which can be fit only to a single class: Isolation Forest (IsFrst) [20] and a one class SVM (OSVM). The training data is therefore only provided at enrolment, when the user first uses the system, and as such we vary the number of samples provided at enrolment. This helps us provide insights on how many samples are necessary to capture sufficient intra-user variation. The rest of the experiment follows the same method described at the beginning of Section 5, with the difference that we repeat the enrolment 10 times to ensure that the enrolment samples do not bias the results (selecting particularly good/bad samples for enrolment will greatly influence the results). To compute the EER, we use the distance from the decision boundary for OSVM and the anomaly score for IsFrst. The results are averaged over the number of repeated enrolments first and then over the individual users.

Results. We report the results of the analysis in Figure 6. The figure shows the distributions of EERs across: (i) ALL or Post-FTA dataset, (ii) number of enrolment samples, (iii) aggregation window size. The number of enrolment samples represents a small fraction of the available user data
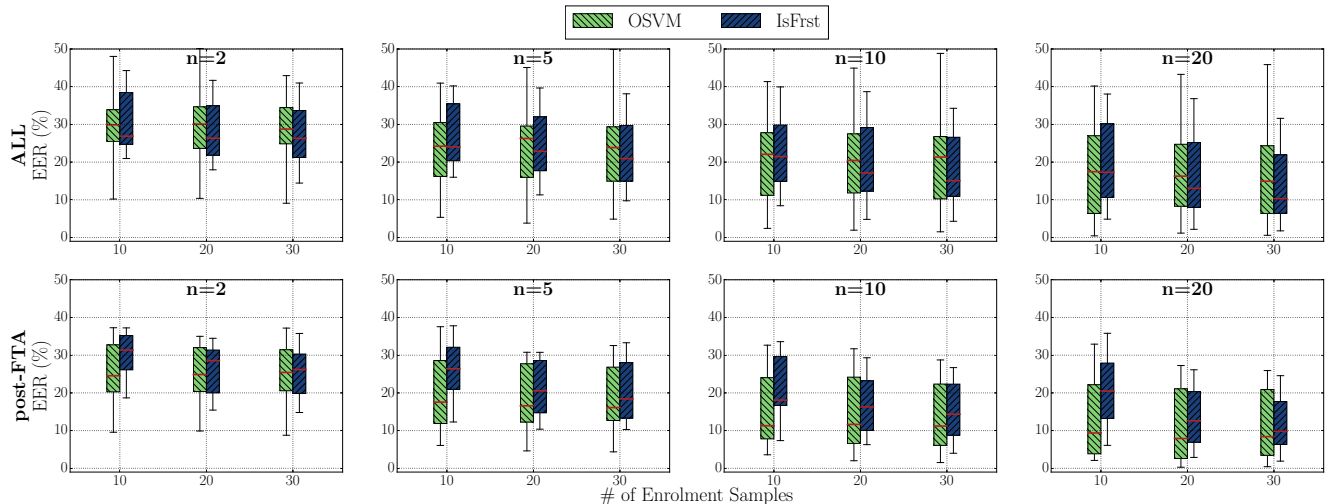
Figure 6: EERs in the one-class case, for various classifier, aggregation window sizes and no. of enrollment samples. Values are averaged over the number of random enrollments first and then across users. The first row of plots reports the results for the ALL dataset, while the second row reports the post-FTA dataset. Each plot reports the aggregation window size n used at the top, i.e., either 2, 5, 10 or 20. Whiskers show the 95% confidence intervals.

(for each user we have at least 178 samples, and we use as few as 10 for training). We found that OSVM performs slightly better on average but not in a consistent manner. Using the Post-FTA dataset leads to better recognition results (as low as 4% at 20 aggregated samples) and smaller variations in the EER distributions, while the average EER for the ALL dataset is above 10% for most configurations with the exception of using IsFrst and aggregating 20 samples.

5.3. One-class Cross-Session

Setup. The evaluation methodology is identical to the one described in Section 5.2, with the difference that now we select enrolment data by choosing a number of sessions (we have at least 6 sessions per user, with a mean of 9.3 sessions and median of 10 sessions) and test on the remaining sessions. A key difference to the work by Kavsaouglu et al. [18] is that we draw training and testing samples from separate sessions. This makes our analysis more realistic, as our results are not positively biased by features that are not stable between sessions.

Results. We report the results of the analysis in Figure 7. As in the previous section, the figure shows the distributions of EERs across: (i) ALL or Post-FTA dataset, (ii) number of enrolment sessions, (iii) aggregation window size. As mentioned in Section 3.2, we have 6-11 measurement sessions per user, the number of enrolment sessions (which varies between 1-3) represents between 10%-50% of available data per user. We found that the EERs are much higher in this case, with most scenarios scoring over 20% EER on

average and high variance in the results. This suggests that each PPG measurement is not only unique to the user, but also to the way the measurement is being taken, leading to most of our features being inconsistent over separate sessions.

5.4. EER User Distribution

We investigate how EER distributes over different users, showing the results in Figure 8. The figure shows how the distribution is uneven across users, with the worst performing user contributing the most to the overall EER. User 6 has over 20% EER in the One-class case and above 30% in the cross session case, while the authentication is relatively stable for a subset of users even in the cross-session scenario (e.g., User 3, 5, 11, 12 have below 5% EER). We found that such results are highly susceptible to the quality of samples used for enrolment, suggesting that stricter quality metrics in this phase could further improve robustness. Equally this effect could be caused due to specific, unknown characteristics of given individuals, as has been shown to be the case in biometric systems previously [9].

6. Conclusion

In this paper, we proposed a system that enables PPG-based authentication by using a smartphone camera. The PPG signal is collected by recording a video from the camera as the user is resting their finger on top of the lens. We extract the signal based on subtle changes in the video that are due to changes in the light absorption properties of the skin as the blood flows through the finger. Using an iPhone
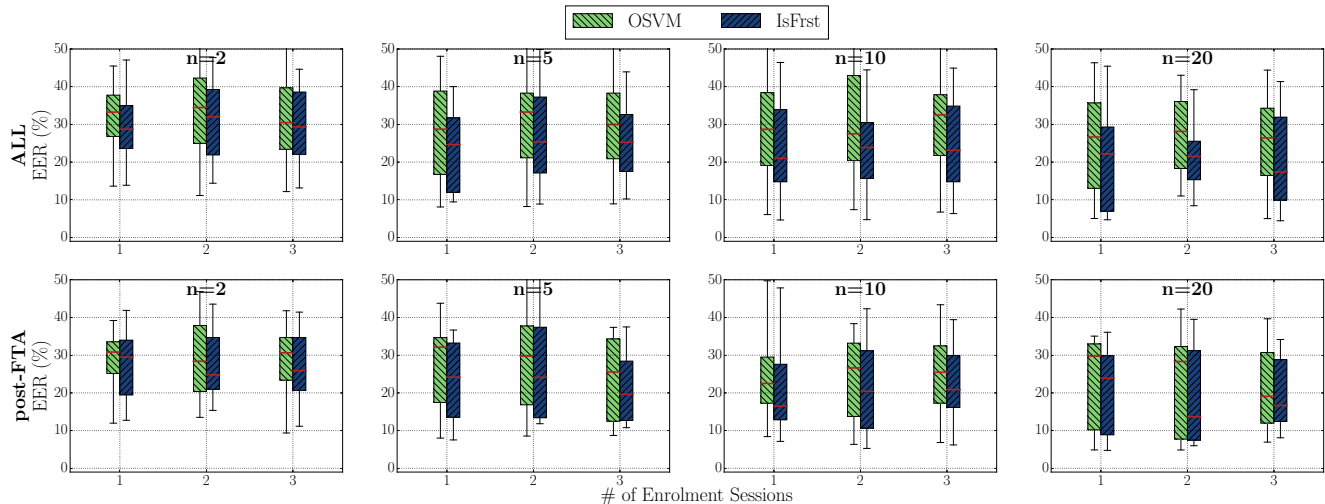
Figure 7: EERs in the one-class cross-session case, for various classifier, aggregation window sizes and no. of enrolment sessions. Values are averaged over the number of random enrolments first and then across users. The first row of plots reports the results for the ALL dataset, while the second row reports the post-FTA dataset. Each plot reports the aggregation window size n used at the top, i.e., either 2, 5, 10 or 20. Whiskers show the 95% confidence intervals.
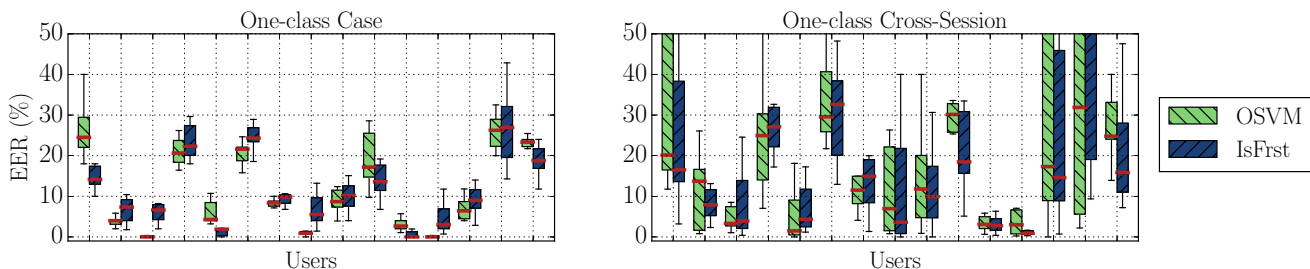

Figure 8: Distribution of EERs across users in our dataset, for the one-class and cross-session case, the Post-FTA dataset and an aggregation window size of 20. Whiskers show the 95% confidence intervals.

X, we collect a dataset of PPG measurements from a set of 15 users, over the course of 6-11 capture sessions per user. We design a pipeline for the data analysis and conduct a set of experiments to evaluate the recognition performance based on a set of features extracted from individual heartbeats. We found that when aggregating sufficient samples for a decision we can achieve equal error rates as low as 8%, but that the performance greatly decreases in the cross-session scenario, with EERs around 20% on average across all users.

Future Work. We observed that a set of factors greatly affect the fidelity of the signal. At first, we found that the warmth of the fingertip changes the amount of blood flow, leading to slightly different measurements. Secondly, breathing has an effect on the signal: heartbeats are stronger when inhaling compared to exhaling [3]. Additionally, slight hand and/or finger movements generate noise that can be (in part) corrected by using accelerometer and gy-roscope measurements. We also plan to better understand the device capabilities needed for this technique to be used successfully. Firstly, evaluating the hardware layout impact on the signal quality, in particular distance between light source and camera. Secondly, using lower frame rates and lower resolution recordings, as using more coarse-grained measurements will enable the use of PPG authentication on lower-end devices. Finally, a larger dataset, both in terms of users and sessions will help us draw more insights on the time stability of the PPG biometric trait.

Acknowledgements

References

[1] Opensignal ppg dataset. http://www.opensignals.net. 3

[2] Shu-Di Bao, Yuan-Ting Zhang, and Lian-Feng Shen. Physiological signal based entity authentication for body area sensor networks and mobile healthcare systems. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 2455–2458. IEEE, 2006. 2, 3

[3] Dieter Barschdorff and Wei Zhang. Respiratory rhythm detection with photoplethysmographic methods. In *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 912–913. IEEE, 1994. 8

[4] Angelo Bonissi, Ruggero Donida Labati, Luca Perico, Roberto Sassi, Fabio Scotti, and Luca Sparagino. A preliminary study on continuous authentication methods for photoplethysmographic biometrics. In *2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, pages 28–33. IEEE, 2013. 3

[5] Anne M Brumfield and Michael E Andrew. Digital pulse contour analysis: investigating age-dependent indices of arterial compliance. *Physiological measurement*, 26(5):599, 2005. 2

[6] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, Cinna Soltanpur, and Homer Nazeran. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International journal of biosensors & bioelectronics*, 4(4):195, 2018. 3

[7] Eric Chern-Pin Chua, Stephen J Redmond, Gary McDarby, and Conor Heneghan. Towards using photo-plethysmogram amplitude to measure blood pressure during sleep. *Annals of biomedical engineering*, 38(3):945–954, 2010. 2

[8] Gunnar Dahlgren, Flavio Veintemilla, Göran Settergren, and Jan Liska. Left ventricular end-systolic pressure estimated from measurements in a peripheral artery. *Journal of cardiothoracic and vascular anesthesia*, 5(6):551–553, 1991. 2

[9] George R. Doddington, Walter Liggett, Alvin F. Martin, Mark A. Przybocki, and Douglas A. Reynolds. Sheep, goats, lambs and wolves: a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. In *ICSLP*, 1998. 7

[10] Simon Eberz, Kasper B Rasmussen, Vincent Lenders, and Ivan Martinovic. Evaluating behavioral biometrics for continuous authentication: Challenges and metrics. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 386–399, 2017. 2

[11] Mohamed Elgendi. On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1):14–25, 2012. 2

[12] YY Gu and YT Zhang. Photoplethysmographic authentication through fuzzy logic. In *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003.*, pages 136–137. IEEE, 2003. 2, 3

[13] YY Gu, Y Zhang, and YT Zhang. A novel biometric approach in human verification by photoplethysmographic signals. In *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003.*, pages 13–14. IEEE, 2003. 2, 3

[14] In Choel Jeong, Hyungro Yoon, Hyunjeong Kang, and Hojun Yeom. Effects of skin surface temperature on photoplethysmograph. *Journal of healthcare engineering*, 5(4):429–438, 2014. 3

[15] Nima Karimian, Zimu Guo, Mark Tehranipoor, and Domenic Forte. Human recognition from photoplethysmography (ppg) based on non-fiducial features. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4636–4640. IEEE, 2017. 3

[16] Nima Karimian, Mark Tehranipoor, and Domenic Forte. Non-fiducial ppg-based authentication for healthcare application. In *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 429–432. IEEE, 2017. 2, 3

[17] Walter Karlen, Srinivas Raman, J Mark Ansermino, and Guy A Dumont. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7):1946–1953, 2013. 3

[18] A Reşit Kavsaoğlu, Kemal Polat, and M Recep Bozkurt. A novel feature ranking algorithm for biometric recognition with ppg signals. *Computers in biology and medicine*, 49:1–14, 2014. 2, 3, 7

[19] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011. 3

[20] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, 2008. 6

[21] Y Maeda, M Sekine, T Tamura, A Moriya, T Suzuki, and K Kameyama. Comparison of reflected green light and infrared photoplethysmography. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2270–2272. IEEE, 2008. 2, 3

[22] Sandrine C Millasseau, RP Kelly, JM Ritter, and PJ Chowienczyk. Determination of age-related increases in large artery stiffness by digital pulse contour analysis. *Clinical science*, 103(4):371–377, 2002. 2

[23] Willie Bosseau Murray and Patrick Anthony Foster. The peripheral pulse wave: information overlooked. *Journal of clinical monitoring*, 12(5):365–377, 1996. 2

[24] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005. 5

[25] Tania Pereira, Nate Tran, Kais Gadhoumi, Michele M. Pelter, Duc H. Do, Randall J. Lee, Rene Colorado, Karl Meisel, and Xiao Hu. Photoplethysmography based atrial fibrillation detection: a review. *npj Digital Medicine*, 3, 2020. 2

[26] Brian C. Ross. Mutual information between discrete and continuous data sets. *PLoS ONE*, 9(2):e87357, 2014. 5

[27] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49, 1978. 4

[28] Abhijit Sarkar, A Lynn Abbott, and Zachary Doerzaph. Biometric authentication using photoplethysmography signals. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2016. 3

[29] K. Shelley and S. Shelley. *Pulse Oximeter Waveform: Photoelectric Plethysmography*. Clinical Monitoring. Saunders Company, 2001. 2

[30] Petros Spachos, Jiexin Gao, and Dimitrios Hatzinakos. Feasibility study of photoplethysmographic signals for biometric identification. In *2011 17th International Conference on Digital Signal Processing (DSP)*, pages 1–5. IEEE, 2011. 3

[31] Jian Liu Yingying Chen Jerry Cheng Jiadi Yu Tian-ming Zhao, Yan Wang. Trueheart: Continuous authentication on wrist-worn wearables using ppg-based biometrics. In *Proceedings of IEEE International Conference on Computer Communications (IEEE INFOCOM 2020)*, 2020. 3

[32] Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999. 5

[33] Jianchu Yao, Xiaodong Sun, and Yongbo Wan. A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4576–4579. IEEE, 2007. 3