Unobtrusive Gait Recognition using Smartwatches

Neamah Al-Naffakh[1,2], Nathan Clarke[1,3], Fudong Li[1], Paul Haskell-Dowland[1,3]

Abstract: Gait recognition is a technique that identifies or verifies people based upon their walking patterns. Smartwatches, which contain an accelerometer and gyroscope have recently been used to implement gait-based biometrics. However, this prior work relied upon data from single sessions for both training and testing, which is not realistic and can lead to overly optimistic performance results. This paper aims to remedy some of these problems by training and evaluating a smartwatch-based biometric system on data obtained from different days. Also, it proposes an advanced feature selection approach to identify optimal features for each user. Two experiments are presented under three different scenarios: Same-Day, Mixed-Day, and Cross-Day. Competitive results were achieved (best EERs of 0.13% and 3.12% by using the Same day data for accelerometer and gyroscope respectively and 0.69% and 7.97% for the same sensors under the Cross-Day evaluation. The results show that the technology is sufficiently capable and the signals captured sufficiently discriminative to be useful in performing gait recognition.

Keywords: mobile authentication, gait biomtrics; accelerometer; smartwatch authentication

1 Introduction

Billions of mobile devices are being used globally having a wide variety of applications (e.g., e-commerce and banking). The use of mobile devices has inherently raised security concerns and there exists a prevalent requirement to secure these devices. Smartwatches have been steadily increasing in popularity and this trend is expected to continue as the technology improves. Therefore, wearables could be used to enhance mobile security in a more effective way. Recent studies have demonstrated that both smartphones [DNBB10, MM14, NWB12] and smartwatches [JW15, SMS16, ACDL16] can provide gait-based biometric authentication service by using various sensors. However, the majority of prior research either used a limited dataset or trained and tested the system on data that was collected on the same day (which is not a realistic model for a real world application as the user would be required to enroll on the system every day). To this end, this paper explores the use of smartwatches for transparent authentication based upon gait recogntion. The main contributions of this study are demonstrated as follows

- To the best of the author's knowledge, this is the biggest dataset for smartwatch-based gait authentication, which contains gait data of 60 users over multiple days.
- A comprehensive feature set was extracted in the time and frequency domains and analyzed to highlight their impact on system performance.
- The novel feature selection method utilised a dynamic feature vector for each user and successfully reduced the feature vector size with better performance.
- Identifying the optimal source sensor for the authentication task.
- The results of this study outperform the prior biometric accelerometer –based studies.

[1] School of Computing, Electronics and Mathematics, Plymouth University, UK,
[2] Computer Science and Mathematics College, Kufa University, Najaf, Iraq
[3]Security Research Institute, Edith Cowan University, Perth, Western Australia
{Neamah.Al-Naffakh, N.Clarke, Fudong.Li}@plymouth.ac.uk; p.haskelldowland@ecu.edu.au

The rest of the paper is organized as follows: Section 2 reviews the state of the art in transparent and continuous authentication that specifically uses accelerometer (Acc) and gyroscope (Gyr) sensors. Data collection and feature extraction are outlined in Section 3. Sections 4, 5 and 6 present the experiment design, feature selection approach, results and discussion. Section 7 presents the conclusions and future research directions.

2 Related Work

Gait-based biometric systems have an advantage over password-based systems in that impersonation is much more difficult to accomplish even video footage of someone walking on a treadmill (to match the victim's pace) is not sufficient to mimic a user [GSB07]. Verifying people based on their walking patterns is an unobtrusive mechanism that does not require explicit user interaction and provide continuous authentication. Recently, increased interests are shown in mobile gait authentication; and performance rates vary considerably depending upon feature extraction methods and types of classifiers utilised. A comprehensive analysis of the prior studies on gait authentication using smartwatch and mobile sensors is summarized in Table 1.

Tab 1. Comprehensive Analysis on Gait Authentication using Mobile and Smartwatch Sensors

Study	Approach	Features Type	Classification methods	Accuracy %	Users	Duration (Seconds)	Device
[DNBB10]	C	TD	DTW	20.1 (EER)	51	120/MD	M
[NDBB11]	C	TD	DTW	21.7 (EER)	48	1200/CD	M
[MM14]	C	TD	DTW	19 (EER)	35	240/CD	M
[KWM10]	S	TD	NN	100 (CCR)	10	300-600/SD	M
[NB11]	S	FD	HMM	6.15 (EER)	48	1200/CD	M
[HN12]	S	FD	SVM	10 (EER)	36	1200/CD	M
[NWB12]	S	FD	KNN	8.24 (EER)	36	1200/CD	M
[CLB15]	S	TD	BN	96.27(CCR)	44	400/SD	M
[NBB11]	S	TD	SVM	10 (EER)	51	120/CD	M
[JW15]	S	TD	RF	1.4 (EER)	59	300-600/SD	SW
[WTGYS16]	S	TD	RF	94.2 (CCR)	17	2160/SD	SW
[CAM16]	S	TD	KNN	2.9 (EER)	15	- / SD	SW
[SMS16]	S	TD	RF	2.6 (EER)	18	350/CD	SW
[KPR16]	S	TD & FD	KNN	95 (CCR)	40	240/SD	SW
[ZYCWS17]	S	TD & FD	NN	0.5 (EER)	9	- / SD	SW

Legend: **C**: Cycle-based; **S**: Segment-based; **TD**: Time Domain; **FD**: Freqency Domain; **DTW**: Dynamic Time Warping; **HMM**: Hidden Markov Model; **SVM**: Support Vector Machine; **KNN**: k-nearest neighbours; **RF**: Random Forest; **NN** Neural Network; **EER**: Equal Error Rate; **CCR**: Correct Classification Rate; **M**: Mobile; **SW**: Smartwatch; **SD**: Same Day; **, CD**: Cross Day

Two main approaches can be used to extract gait features, namely cycle and segment-based. Cycle extraction attempts to segment the data into pairs of steps. This offers a very exciting opportunity where if such a system is implemented effectively. Howeover, the literature shows high EERs (ranging from 19% [MM14] to 21.7% [NDBB11]). This is most likely the result of the complicated and unclear nature of cycle extraction. In

contrast, the performance of the segment based methods, which focus on fixed-length blocks of data, appearing to be more effective and stable, with studies reporting EERs between 1.4% and 10 % [JW15, NBB11]. With respect to features, several studies in the domain have used both time domain (TD) and frequency domain (FD) features but little attention has given to measure the impact of these features on the system performance.
As illustrated in Table 1, the most recent studies used a smartwatch device to collect the Acc and Gyr gait data for transparent authentication systems (TAS). However, in [JW15, WTGYS16, CAM16, ZYCWS17] the gait data was obtained on the SD and the dataset is considered limited ranging from 9 to 18 users (apart from JW15). In addition, the authors did not carry out any particular study on feature selection in order to identify the most discriminative features. In contrast, a feature selection mechanism was conducted by [KPR16] and reported 95% CCR by using the SD scenario. However, the system performance was reduced to 86.8% CCR (with a limited dataset of 13 users only) when the CD scenario was applied. This can be attributed that the proposed approach is not sophisticated enough to identify a unique feature set for individuals that work over time.

3 Data Collection and Feature Extraction

The Acc and Gyr data was captured from the Microsoft Band 2 at a rate of 32 samples per second for the x, y and z-axes and automatically sent to a smartphone residing in the user's pocket via Bluetooth. In total, 60 users participated in the data collection; each user was required to walk on a predefined route in two sessions on two different days (within a time frame of 3 weeks between the sessions). Every session consisted of three walks trails from each user. In each trail, the user was asked to walk at a natural speed on flat ground for 2 minutes with few turns. For a more realistic scenario, the subject had to stop in order to open a door. Moreover, no other variables, such as type of footwear or clothing, are controlled. Once the data collection was completed, the signal processing phase was undertaken- a brief description of the steps is as follows

- Time interpolation: as the Microsoft Band 2 sensors were not able to record data at a fixed sample rate, time interpolation was required to make sure that the time period between two successive data points was always equal
- Filtering: a low pass filter was designed in order to enhance the accuracy of the signal. This was carried out with several settings (i.e. 0.1, 0.2, 0.3, 0.4 and 0.5) and the cut-off frequency of 0.2Hz achieved the best accuracy.
- Segmentation: the tri-axial raw format for both Acc and Gyr signals were segmented into 10-second segments by using a sliding window approach with no overlapping. Therefore, in total 36 samples were collected for each user per day.

A feature extraction process is carried out on both the Acc and Gyr data segments of each user. In total, 140 features were extracted based upon prior work identified in gait recognition studies. Features were extracted from both the time and frequency domains on Acc or Gyr data. Since most features are generated on a per-axis basis and each sensor has 3 axes, most features are represented by a multiple of three values. The number of generated features and their types are presented in Table 2. Details of these features (e.g., how they are calculated) can be found in [KWM10, JW15].

Tab 2. List of the extracted TD and FD features

Feature Type	NF	TD	FD	Feature Type	NF	TD	FD
Difference	3	√	√	Skewness	3	√	√
Variance	3	√	√	Average	3	√	√
Median	3	√	√	Kurtosis	3	√	√
Maximum	3	√	√	Minimum	3	√	√
Energy	3	-	√	Entropy	3	-	√
Time Between Peaks	3	√	-	Standard Deviation	3	√	√
Correlation Coefficients	3	√	√	Root Mean square	3	√	√
Cosine Similarity	3	√	-	Covariance	3	√	-
Interquartile range	3	√	√	Binned histogram	30	√	-
Peaks Occurrence	3	√	-	Percentile 25,50	6	√	√
Average Absolute Difference	3	√	√	Average Resultant Acceleration	1	√	√

4 Experimental Methodology

Biometric authentication or verification is a binary classification problem, where the aim is to determine if a system can identify a genuine user correctly or as an imposter. A separate model is generated for each user. The reference and testing templates were created under three different scenarios for SD, MD and CD. For SD and MD, the data was divided into two sets: 60% of the data for training and the remaining 40% for testing; also training samples were extracted from both days for the MD scenario. For the CD scenario, the first day's data was used for training and the second day data was employed for testing. Also, the Feedforward Multi-layer Perceptron (FF MLP) neural network was used as the default classifier due to its reliable performance [KWM10].

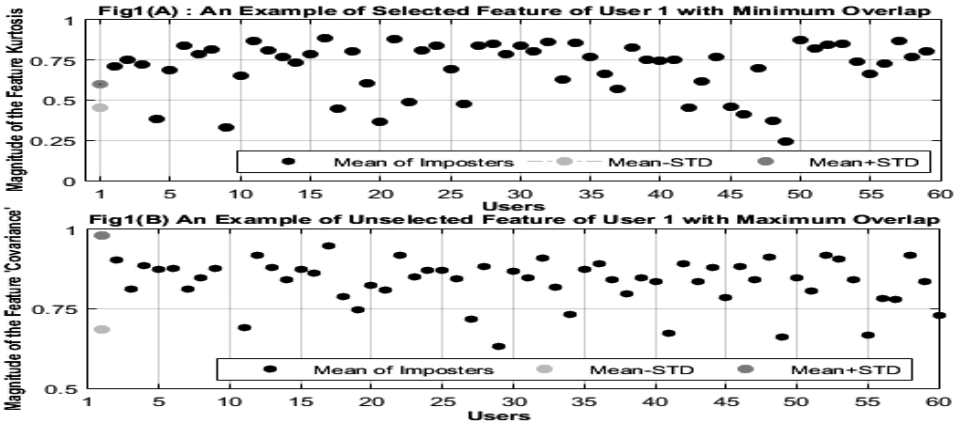
The feature selection step is important for biometrics based studies in order to reduce the potentially large dimensionality of input data. By selecting an optimal feature set for individuals, the system performance could be potentially enhanced. Also, it will be easier to manipulate and calculate smaller feature subsets on digital devices. Majority of gait recognition systems select common features for all the population; this could be useful if the system is based on identifying the genuine user only. However, a balance between security and usability needs to be taken. Therefore, this study focused on creating a dynamic feature vector that contains distinctive features for each user. As a result, the feature subset for each user very different from each other (e.g., the reference templates could be created by using features 1, 2, and 7 for user 1 while features 3, 4, and 5 for user 2). This can be achieved by calculating the mean and Standard Deviation (STD) for each feature of all users and then compares the authorized user's results against impostors to select the feature set with the minimal overlap. In other words, for each feature, a score is calculated based upon the following condition:

- If the mean of imposter's activity is not within the range of the mean +/- STD of genuine, add 1 to the total score.
- Dynamically select the features according to their score order from high to low. The highest score means less overlap between imposters and genuine user (see Fig 1 (A)).

Fig 1 shows an example of applying the proposed feature selection method on two different features for user 1. Based upon the overlap percentage, it is clear from Fig 1 that the Kurtosis feature has lowest overlap score compared to the Covariance feature.

As a result, the Kurtosis feature was selected to form the feature vector of user1, while the second feature (i.e. Covariance) was neglected. This procedure is repeated for each individual and each feature resulting in a bespoke and prioritized feature set.

Fig 1: the effect of the dynamic feature selection approach



In order to evaluate the proposed system, several consecutive experiments were undertaken that include:

- Analysis and highlighting the impact of the time and frequency domain features on the system performance.
- The discriminative features were evaluated and the reference and test templates were created by selecting an optimal feature set for each user independently.
- The results cover the three evaluation scenarios (SD, MD, and CD), the two different sensors (Acc and Gyr), and one classification algorithm (FF MLP neural network).

5 Results

According to the plan, the first expermient was to highlight the impact of the time and frequency domains features on the system performance and the results are presented in Table 3(using the SD scenario).

Table 3: EER of Using All Features, Time and Frequency Domains

Feature type	NF	EER (%)	
		Acc	Gyr
All Features	140	**0.13**	3.37
Time domain	88	**0.15**	3.73
Frequency domain	52	3.09	12.69

It is clear that good performances were achieved by using the TD features and all feature sets; and little difference in results is observed between the two sets. By using the FD features alone, reasonable performance is obtained; but its performance is far less promising in comparison with the results of using TD features alone, suggesting FD features add little contribution towards the classification process. Given the fact that detecting redundancies features makes the system more efficient, therefore, only the TD features (i.e. 88 features) were used in subsequent experiments as it shows low EER.

Further analysis was conducted to reduce the extracted TD features by applying the proposed dynamic feature selection method. Table 4 shows the impact of feature selection under the SD, MD and CD scenarios and two sensors. It can be concluded that the feature selection mechanism has a positive effect on the performance by minimizing the number of features and maximizing the discriminative information. In addition, as expected the system performance of the SD and MD scenarios exceeded the CD evaluation for both sensors.

Tab 4: Impact of the dynamic feature selection technique upon the performance in detail.

Evaluation Scenario	Sensor	Number of Selected Features								
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**88**
SD	Acc	1.13	0.78	0.24	0.26	0.27	**0.13**	0.20	0.16	0.15
SD	Gyr	6.6	4.88	3.63	3.74	**3.12**	3.58	3.48	3.43	3.73
MD	Acc	2.22	0.82	0.42	0.22	0.25	0.20	0.22	**0.16**	0.28
MD	Gyr	7.63	4.81	3.85	3.80	3.53	3.51	**3.24**	3.25	3.35
CD	Acc	4.68	2.39	1.43	0.9	0.84	0.83	**0.69**	0.77	0.93
CD	Gyr	11.09	9.76	8.62	8.49	8.94	8.53	8.42	**7.97**	8.29

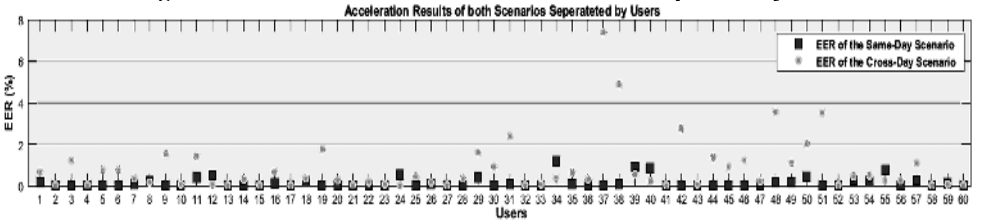
As shown in the Table 4 vastly good results were achieved with best EERs of 0.13% for Acc and 3.12% for Gyr by utilizing the SD scenario (compared to 2.9%, 1.4% and 0.5% of EERs by [CAM16, JW15, ZYCWS17] and CCR of 95% and 94% by [WTGYS16, KPR16]. Moreover, high performances with EERs of 0.78% and 4.88% can still be achieved by using only 20 features for Acc and Gyr accordingly. Comparing to the SD scenario, no significant difference was found in the MD scenario where the best EERs are 0.16% for Acc and 3.24% for Gyr, as the training set contained samples from both days. However, these results outperform the outputs (i.e. EER ranging from 6.1% to 21.7%) of previous studies [NB11, NBB11] under the MD scenario.

As shown in Table 4, the best performance of the CD scenario are EERs of 0.69% (for Acc) and 7.97% (for Gyr). As expected the system performance is droped under the CD test as the human's behaviour does change over time. Nonetheless, the presented CD results are still very promising (i.e. 0.69% EER) in comparison with the prior work that reported EERs in the range of 2.6% - 21.7% [NDBB1, MM14, NB11, HN12, NWB12, NBB11, SMS16] . In addition, the CD test does not require the user to re-enrol in the system on a daily basis.

With the aim to understand how individual user performed, results on each user's Acc for both SD and CD scenarios are presented in Figure 2. As shown in Figure 2 high level of performance (i.e. in the range of 0-2% EER) were obtained for 90% of users, (apart from users 31, 37, 38, 42, 48, and 51) for both SD and CD scenarios. This suggests that users have a consistent and distinctive set of Acc pattern characteristics.

With respect to the feature subset size, as shown in Table 4 the SD test requires less features (i.e., 60 features) than the CD (i.e., 80 features) to produce the lowest EER. This could be explained because the user's gait pattern could vary or be inconsistent over time due to many factors (e.g., shoes, clothes, and mood), hence more features are required for individual to be identified. Moreover, creating a dynamic feature vector size for each user independently might greatly reduce the EER (e.g., the refrence template can be constructed by using 20 features for user 1 while 40 features will be used for user 2)

Fig 2: The Acceleration Results of Both Scenarios Separated by Users



6 Discussion

As shown in the previous section, the presented results reveal that smartwatch based gait recognition is highly efficient and recommended to be used for verifying users in a transparent and continuous manner. The best results were EERs of 0.13% and 0.69% for SD and CD scenarios respectively by using Acc signals. However, the results were obtained in controlled conditions, so, further investigation is required by collecting the user's data during the entire day over multiple days in order to find the influence of collecting real life data on the system performance. Although features were extracted from both time and frequency domains, the findings in Table 2 support the use of time domain features alone as a better decision especially for mobile devices. For the realistic test, the EER was slightly increased from 0.13% to 0.69% when the Acc reference and test templates were created from the data of two different days. Because the obtained Acc results were very strong, the fusion of data from both sensors was not necessary. Further influencing factors on the biometric system performance is the selected feature subset; selecting unique features for each user would improve the results and reduce the complex computations on the smart devices which have limited processing resources. Therefore, a feature selection approach of any mobile-based biometric system needs to be sophisticated enough before the classification phase takes place. As expected, the proposed feature selection approach in this study, which was based on creating a dynamic feature vector for each user, successfully reduced the user's feature vector size and resulted in lower EER's of 0.13% and 0.69% for the SD and CD tests respectively (compared to 0.15% and 0.93% when the whole features were used). However, further investigation is required to reduce the number of the optimal features for each user independently which might offer better accuracy/error rates.

7 Conclusions and future work

Based on the performance in this study, smartwatch-based gait recognition shown to be effective and can be used with in TAS. The paper also presents an analysis of the feature set to examine the impact of features upon performance, which has resulted in proposing a dynamic feature set. The proposed system was evaluated by collecting the motion data from 60 users and analysed the feature set to determine its uniqueness. However, more experimental work should be carried out to investigate the impact of the dynamic feature vector size for each user.

Further work will also explore examining a wider range of different activities (e.g., fast

walking and typing on smartphone touch screen) to expand the technique from merely gait recognition to activity recognition. A future study will aim to remove the one factor that is explicitly controlled in all previous studies – the nature of the controlled data collection and instead look to understand what the performance of the approach is with real life data over a prolonged period of time. As the nature of the real life signals is likely to be noisy, an appraoch will be used in order to predict the user's activity.

References

[ACDL16] N. AI-Naffakh, N. Clarke, P. Dowland and F. Li, "Activity Recognition using Wearable Computing", in ICITST, Barcelona, pp. 189-195, 2016.

[CAM16] Cola, G., Avvenuti, M. and Musso, F.. Gait-based authentication using a wrist-worn device. in MOBIQUITOUS, Hiroshima, 2016.

[CLB15] N. Capela, E. Lemaire and N. Baddour, "Feature Selection for Wearable Smartphone-Based Human Activity Recognition with Able bodied, Elderly, and Stroke Patients", PLOS ONE, vol. 10, no. 4, p. e0124414, 2015.

[DNBB10] M. O. Derawi, C. Nickel, P. Bours and C. Busch, "Unobtrusive User-Authentication on Mobile Phones using Biometric Gait Recognition",in IIH-MSP, pp.306-311, 2010.

[GSB07] D. Gafurov, E. Snekkenes and P. Bours, "Spoof Attacks on Gait Authentication System",IEEE Transactions on Information Forensics and Security,pp. 491-502, 2007.

[HN12] M. Reese Hestbek and C. Nickel, "Biometric gait recognition for mobile devices using wavelet transform and support vector machines", in IWSSIP, pp. 205-210, 2012.

[JW15] A. H. Johnston and G. M. Weiss, "Smartwatch-Based Biometric Gait Recognition", in Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, USA, 2015.

[KPR16] R. Kumar, V.V Phoha, and R. Raina: Authenticating users through their arm movement patterns. arXiv preprint arXiv:1603.02211 (2016).

[KWM10] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Cell phone-based biometric identification", in *the Fourth IEEE International BTAS Conference*, 2010, pp. 1–7.

[MM14] M. Muaaz and R. Mayrhofer, "Orientation Independent Cell Phone Based Gait Authentication", in MOMM Conference, Taiwan, 2014.

[NB11] C. Nickel and C. Busch,"Classifying accelerometer data via Hidden Markov Models to authenticate people by the way they walk", in ICCST, Barcelona, pp. 1–6, 2011.

[NBB11] C. Nickel, H. Brandt and C. Busch, "Classification of Acceleration Data for Biometric Gait Recognition on Mobile Devices", in BIOSIG Conference, Darmstadt, 2011.

[NDBB11] C. Nickel, M. O. Derawi, P. Bours, and C. Busch, "Scenario test of accelerometer-based biometric gait recognition", in *the Third IWSCN* Conference, pp. 15–21, 2011.

[NWB12] C. Nickel, T. Wirtl and C. Busch, "Authentication of Smartphone Users Based on the Way They Walk Using k-NN Algorithm", in IIH-MSP, Greece, pp. 16-20, 2012.

[SMS16] Babins, S., Manar, M., Nitesh, S.: Walk-Unlock: Zero-Interaction Authentication Protected with Multi-Modal Gait Biometrics. arXiv:1605.00766 (2016).

[WTGYS16] G. M. Weiss, J. L. Timko, C. M Gallagher, K. Yoneda and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach", in BHI, Las Vegas, USA, pp. 426-429, 2016.

[ZYCWS17] ZHANG, X., YAO, L., CHEN, K., WANG, X., Z. SHENG, Q. and GU, T: DeepKey: An EEG and Gait Based Dual-Authentication System. arXiv: 1706.01606 (2017).