Mobile Sensor-based Biometrics using Common Daily Activities

Kenichi Yoneda and Gary M. Weiss
Department of Computer and Information Science, WISDM Laboratory
Fordham University
441 East Fordham Road, Bronx NY 10458
{kyoneda, gaweiss}@fordham.edu

***Abstract*— Research on mobile sensor biometrics increased when mobile devices with powerful sensors, such as smartphones, became ubiquitous. However, existing studies are quite limited, especially with regard to the physical activities that are used to provide the biometric signature—many studies only consider a single activity. In this study, we provide the most comprehensive analysis of mobile biometrics to date. We evaluate eighteen physical activities and nine sensor combinations for their biometric efficacy (the accelerometer and gyroscope sensors from a smartphone and smartwatch are used). Our mobile biometric models are evaluated with respect to identification and authentication performance and are shown to achieve excellent results in both cases. Furthermore, our models perform well even when built from all eighteen activities without activity labels, which represents a big step towards achieving the goal of continuous biometrics using only a smartwatch and smartphone.**

Keywords—Mobile biometrics, Multibiometrics, Sensors

I. Introduction

This paper provides a comprehensive evaluation of motion-based mobile biometrics. The accelerometer and gyroscope sensors on commercially available Android smartphones and smartwatches [1] are used to capture a user's movements, while the user performs eighteen common daily activities. Classification algorithms are then used to induce biometric identification and authentication models from the sensor data from each of the eighteen individual activities—and from the combined activities—and these models are evaluated. Our results show that these models achieve good overall performance. Of particular note, the model induced from eighteen activities, even without activity labels, performs well, which represents a significant step toward achieving continuous biometrics in an unstructured environment. Our results also allow us to evaluate the relative effectiveness of the eighteen activities for their utility as biometric signatures.

This study extends prior research by our lab on gait-based biometrics [2]. Our previous work looked at a single activity (walking) and used data from three sensors. However, this study looks at eighteen activities using data from four sensors. The activities were originally selected for activity recognition research that we conducted, which we leveraged for this biometrics research [3, 4].

This paper makes several very significant contributions. First, it evaluates a large number of daily activities (18) to assess their ability to serve as biometric signatures. It also evaluates the combination of these activities as a biometric signature, which is a key step toward achieving the very ambitious goal of continuous biometrics, where a user is identified by their motion while performing normal daily tasks. Our study also evaluates more sensors and devices than most other research, by considering the accelerometer and gyroscope sensors on both the smartphone and smartwatch, separately and together (a total of nine sensor configurations are evaluated). The research in this paper also relies only on commonly available commercial devices, which means that this work can be used to build cost-effective biometric systems. Our study also includes more test subjects (51) than many prior research studies [5]. In summary, this paper provides one of the most comprehensive studies of mobile biometrics to date.

II. Data Collection and Transformation

This section describes the data collection process, as well as the data transformation process used to convert the raw time-series data into labeled examples, which can be used to build our biometric identification and authentication models. The study includes 51 test subjects. Each subject spent three minutes performing each of the 18 activities. Accelerometer and gyroscope data were collected from the smartwatch and smartphone that were worn by the subjects.

A. Overview of Data Collection Process

The majority of the test subjects were undergraduate or graduate university students. Because the study used human subjects, it was approved by the university's Institutional Review Board (IRB) and each subject was required to give written informed consent before participating in the study.

The 51 test subjects were asked to perform 18 routine activities for 3 minutes each. Prior to performing the activities, the subject placed an Android smartphone in their front right pocket and strapped a Bluetooth-paired Android smartwatch on their dominant hand. The study used the Google Nexus 5/5X and Samsung Galaxy S5 smartphones running Android 6.0 (Marshmallow) and the

LG G Watch running Android Wear 1.5. For consistency, the participant was instructed to have the phone oriented upright with the screen facing away from them. Data collection was stopped between each activity, and it took the subjects about 70 minutes to complete the entire data collection process.

The time-series sensor data was collected by an Android application that our research group developed, which collected accelerometer and gyroscope data from both the smartphone and smartwatch at a rate of 20 Hz. At the end of a data collection session the raw time-series sensor data was transferred from the smartphone to our lab machine via a USB connection.

B. Monitored Activities

This study includes 18 routine activities. The purpose of using these activities is to determine the efficacy of specific activities for biometrics, although we also wanted to determine if a collection of normal daily activities could collectively form a useful biometric signature—and ultimately be used to implement continuous biometric monitoring.

Table 1 lists the 18 activities included in this study. They are organized into three categories. The divisions are based on the belief that the smartwatch sensor data will be especially beneficial for the hand-oriented activities. Actual food was used for the eating activities.

Table 1. Eighteen Monitored Activities

General Activities (non hand-oriented)
• Walking
• Jogging
• Stairs (ascending & descending)
• Sitting
• Standing
• Kicking a Soccer Ball (two people)
General Activities (hand-oriented)
• Dribbling a Basketball
• Catch with a Tennis Ball (two people, underhand)
• Typing
• Writing
• Clapping
• Brushing Teeth
• Folding Clothes
Eating Activities (hand-oriented)
• Eating Pasta
• Eating Soup
• Eating a Sandwich
• Eating Chips
• Drinking from a Cup

Most of the activities are self-explanatory and precise details are not presented due to space considerations. For the *stairs* activity, the test subject went up and down a single flight of indoor stairs continuously for three minutes. For the *catch* and *kicking* activities the test subject conducted the activity with a researcher. For the *typing* and *writing* activities, a specific prompt was given to the test subject. For all eating activities, the researcher ensured that sufficient food/drink was available for the test subject, so that they could continue the activity without interruption.

C. Data Transformation

This section describes the process for transforming the raw time-series sensor data into examples that can be used by conventional classification algorithms. Sensor readings, for both the accelerometer and gyroscope, are recorded in the following format:

<timestamp, x, y, z>

The timestamp is measured in nanoseconds and the x, y, z values correspond to the three spatial axes. The x, y, and z values are measured in m/s^2 for the accelerometer and in rad/s (radians per second) for the gyroscope.

We converted the time-series data into examples using a sliding-window approach, without overlap, such that the data stream is divided into 10-second segments. Since every activity was performed for 3 minutes this yielded approximately 18 examples per activity. A 10-second interval was chosen because prior activity recognition and biometrics studies have used the same interval and demonstrated good results [2, 6]. Moreover, 10 seconds is a practical length for performing activities for the purpose of biometrics.

Once the data is divided into 10-second segments, the low-level sensor data is transformed into 43 descriptive, high-level features. The exact same set of features are used for both the accelerometer and gyroscope data. The full list of features is provided below. The value in the square brackets indicates the number of features generated. When three features are generated they correspond to the three spatial axes.

- Average[3]: Average sensor value (each axis)
- Standard Deviation[3]: Standard deviation (each axis)
- Average Absolute Difference[3]: Average absolute difference between the 200 values and the mean of these values (each axis)
- Time Between Peaks[3]: Time between peaks in the sinusoidal waves formed by the data as determined by a simple algorithm (each axis)
- Average Resultant Acceleration[1]: For each of the sensor samples in the window, take the square root of the sum of the square of the x, y, z axis values, and then average them.
- Binned Distribution[30]: The range of values is determined (maximum - minimum), 10 equal-sized bins are formed, and the fraction of the 200 values

within each bin is recorded (each axis)

Finally, each example is appended with a label that indicates the activity the participant was performing and a numerical ID that uniquely identifies the participant.

D. Dataset

This section describes the transformed data set that is used to build the biometric classification models. Each sensor generates over 16,200 examples, all with 43 features, which corresponds to over 45 hours of data per sensor. With 51 test subjects, a complete data set with no missing data would consist of almost 184 hours of data (51 users × 18 activities × 4 sensors × 3 minutes). Our data collection efforts came close to achieving a perfect data set, with a 99% successful collection rate. Table 2 provides a summary of the data. Currently, the data is available upon request but will be submitted to a public dataset repository by the end of the year.

Table 2. Summary of Collected Data (in hours)

Sensor	**Phone**		**Watch**	
	Accel.	*Gyro.*	*Accel.*	*Gyro.*
Collected	45.5	45.4	45.6	45.3
Missing	0.4	0.5	0.3	0.6

III. EXPERIMENT METHODOLOGY

This section describes the methodology for building and evaluating our biometric models. Section IIIA describes the classification algorithms used to build the models, Section IIIB describes the combinations of sensors used to build each model and how the sensor data is fused, and Sections IIIC and IIID describe the methodology for constructing the identification and authentication models.

A. Classification Algorithms

Python's scikit-learn module, an open source library for data mining and analysis [7], provides the three classification algorithms that are used in this study: k-Neighbors, Decision Tree, and Random Forest. The default parameters were used for the three algorithms. For the k-Neighbors classifier, the number of neighbors was set to 5, using uniform weights and the Minkowski distance metric. For the Random Forest classifier the maximum number of features considered was the square root of the number of features in the data, and the number of decision trees in the forest was set to 10.

B. Sensor Combinations

The basic models in this study rely on a single sensor, yielding 4 models, since the phone and watch both have an accelerometer and gyroscope sensor. Since multiple sensors may yield superior results, we also consider the following 5 sensor combinations:

- Phone: Phone Accel + Phone Gyro
- Watch: Watch Accel + Watch Gyro
- Accel: Phone Accel + Watch Accel
- Gyro: Phone Gyro + Watch Gyro
- All: Phone Accel + Phone Gyro + Watch Accel + Watch Gyro

The sensor data is fused by concatenating the features. For example, to create the fused "phone" data set, the 43 phone accelerometer features and 43 phone gyroscope features are concatenated. In the sensor combinations listed above, the first four contain 86 features while the last one contains 172 features. Each combination of sensors essentially yields a different data set. Thus, we have a total of 9 data sets (4 individual sensors and 5 fused sensors).

C. Identification Experiments

The identification task is to identify a user from a sample of their motion sensor data. For this task, we build two types of models. The first type of model builds a classifier for each of the 18 individual activities. This yields 486 different experiment configurations (18 activities × 9 sensor combinations × 3 algorithms).

The second type of model is trained from the aggregation of data from all 18 activities. We execute three variations of this experiment, based on whether the activity label is 1) not provided, 2) provided, or 3) not provided but predicted by an induced activity recognition model. The case where the activity label is provided corresponds to the case where both the training data and test data are generated from a carefully organized sequence of activities (like in this study). The case where the activity labels are not provided more closely correspond to the case of continuous biometrics, where no manual labeling of the activities occurs. The case where we predict the activity label, which results in a two-step classification process, is an attempt to provide the benefits associated with activity labels without the effort of manual labeling the activities. We have 81 total "aggregate" experiment configurations (3 variations × 3 algorithms × 9 sensor combinations).

All experiments use stratified 10-fold cross-validation to build and evaluate the models, to ensure that each fold contains the same distribution of users. Overall, a total of 567 (486 + 81) distinct identification experiments were conducted, with 10 runs per distinct experiment.

D. Authentication Experiments

An authentication model can distinguish a specific user from an imposter. For authentication, each test subject requires their own model, which means we must construct 51 models. Each model must be trained using data from the user to be authenticated. Training data from "other" users is also required, but in real world situations data from all potential imposters will not be available. Thus, the "imposters" in the test set should not be represented in

the training set. Given that we have 51 test subjects, we can partition the 50 "other" users into two sets, one set to be used in the training set and the other set to be used in the test set.

Since authentication is a binary classification problem and the positive class (the user we are trying to authenticate) is rare in comparison to the negative class (imposters we are trying to reject), we under-sampled the "other" users to create a training set that has a ratio of 1:3 (1 user to 3 imposters). We experimented using several different class ratios, including 1:1, and 1:2, but somewhat surprisingly this did not notably alter the results.

We decided to use the 1:3 ratio for three reasons, and we would argue that these reasons generally favor a training set that contains more "other" users than the user to be authenticated. First, this ratio is used in other biometric studies [2, 6]. Second, by using more imposter data than data for the user to be authenticated, we are biasing the classifier to predict "imposter", meaning that it will be very conservative when authenticating a user. Lastly, by using more "imposter" data we are more able to represent a variety of imposters—which is important since the "other" class must represent all other users.

For each user's model, the collected activity data associated with that user is randomly divided into two equal portions. One portion is placed into the training set and the other is placed into the test set, which gives 90 seconds of data for each set. Then eighteen random users are chosen from the data set and 30 seconds of data randomly selected for each of these users. Nine of the users are placed in the training set while the remaining nine are placed in the test set, resulting in 270 seconds (9 users × 30 seconds) of data for each set. Looking at the ratio of data from the user and data from other users, it is clear that the above methodology yields a 1:3 ratio. As with the identification experiments, the process is repeated for each of the 9 sensor combinations and 3 algorithms, giving a total of 1377 experiments (51 users × 9 × 3).

IV. Results

Section IVA presents the results for the identification experiments while Section IVB presents the results for the authentication experiments. Due to space limitations, we cannot provide the detailed results for all of the variations of the experiments, so in some cases we only provide summary results. For example, all experiments were run using the k-Neighbors, Decision Tree, and Random Forest algorithms, but we only provide the most granular results for the Random Forest algorithm, since our results indicate that this algorithm performs best.

As discussed in Section IIIC, some of the identification models utilize data from only a single activity, while other models utilize data from all 18 activities. We begin by describing the results for the models using single activities and then proceed to the results that use all 18 activities.

A. Identification Results using Single Activities

The results in this section are based on a single activity. The most granular results are based on a single test example, which corresponds to 10 seconds of data. However, we can assume that sensor data streaming from a phone or watch comes from a single person—at least over small time frames—so making a prediction based on a single 10-second sample of data is unnecessarily restrictive. To improve prediction accuracy we use 5 examples, or 50 seconds worth of data, and use majority voting to predict the identity of the subject. In the case of a tie, the first user in the list was arbitrarily selected. The results using this strategy are reported in Table 4 (next page).

Based on the average performance over all activities in Table 4 (last row), the best sensor combination to use is "Accel," which corresponds to the accelerometer for both the watch and phone, followed closely by the combinations of all sensors ("All"). The two relevant averages are highlighted in bold. However, they may have suboptimal performance for specific biometric activities. If we look at each of the specific activities, we see that the accelerometer sensor combination performs best (if we include ties) for 16 of the 18 activities, while the combination of all sensors performs best for 14 of the 18 activities. Therefore, from this table, for results on Random Forest with voting, we conclude that the accelerometer sensor combination should be used. Due to space considerations we cannot provide results for all three algorithms, with and without voting, at the level of granularity reflected in Table 4. Thus we provide only the summary results for all three algorithms in Table 3.

The identification accuracies displayed in Table 3 are based on the "Accel" sensor combination, since that combination was shown to perform best overall. The results are also aggregated over all 18 individual activities. Note that the value in Table 3 for Random Forest with voting corresponds to the value of 99.7 highlighted in bold in Table 4.

Table 3. Average Identification Performance (%) using "Accel"

Algorithm	Without Voting	With Voting
k-Neighbors	77.8	88.8
Decision Tree	91.7	98.0
Random Forest	96.4	99.7

The results from Table 3 shows that Random Forest performs best, Decision Trees perform second best, and k-Neighbors performs worst. The superiority of Random Forest for such identification tasks is consistent with some prior studies [2, 6]. More importantly, these results indicate that a majority-voting scheme improves identification accuracy regardless of the classification algorithm used.

Table 4. Identification Accuracy (%) using a Single Activity with Random Forest and Voting

Activity	Single Sensor				Fused Sensor					Avg.
	Phone Accel	*Phone Gyro*	*Watch Accel*	*Watch Gyro*	*Phone*	*Watch*	*Accel*	*Gyro*	*All*	
Walking	100.0	100.0	94.1	80.4	100.0	90.2	100.0	100.0	100.0	96.1
Jogging	100.0	100.0	90.0	88.0	100.0	98.0	100.0	100.0	100.0	97.3
Stairs	98.0	90.0	70.0	43.8	96.0	75.0	100.0	91.7	100.0	84.9
Sitting	100.0	62.7	88.2	33.3	98.0	86.3	100.0	64.7	100.0	81.5
Standing	98.0	39.2	82.4	20.0	94.1	84.0	100.0	50.0	100.0	74.2
Typing	100.0	89.8	94.0	50.0	100.0	100.0	100.0	95.9	100.0	92.2
Teeth	98.0	82.4	94.1	62.7	100.0	96.1	100.0	94.1	100.0	91.9
Soup	100.0	66.7	88.2	62.0	100.0	88.0	100.0	80.0	100.0	87.2
Chips	100.0	76.0	82.4	41.2	98.0	82.4	98.0	80.0	100.0	84.2
Pasta	100.0	56.0	84.0	48.0	100.0	84.0	100.0	71.4	98.0	82.4
Drinking	100.0	58.8	86.3	41.2	100.0	80.4	100.0	60.8	100.0	80.8
Sandwich	98.0	68.0	84.0	38.0	100.0	82.0	100.0	73.5	98.0	82.4
Kicking	96.1	68.6	76.0	32.0	100.0	82.0	100.0	80.0	98.0	81.4
Catch	98.0	70.0	78.0	85.7	100.0	91.8	100.0	91.8	100.0	90.6
Dribbling	96.1	68.6	98.0	90.2	98.0	86.3	96.1	96.1	100.0	92.2
Writing	96.1	80.0	94.1	58.8	100.0	98.0	100.0	90.0	100.0	90.8
Clapping	100.0	86.3	96.1	90.2	100.0	98.0	100.0	100.0	98.0	96.5
Folding	100.0	76.5	64.7	39.2	96.1	86.3	100.0	78.4	100.0	82.4
Avg.	98.8	74.4	85.8	55.8	98.9	88.3	**99.7**	83.2	**99.6**	

B. Identification Results using All 18 Activities

The results for the identification experiments using data from all 18 activities, when using the Random Forest Algorithm, are summarized in Table 5. The table includes results from all 9 distinct sensor combinations (the rows) and three variations of the basic experiment, depending on whether the activity label was not provided at all ("without label"), was provided ("with label"), or predicted using a two-stage learning process ("predicted label").

Table 5. Identification Accuracy (%) using All 18 Activities

Sensors Used	Without Label Voting?		With Label Voting?		Predicted Label Voting?	
	No	*Yes*	*No*	*Yes*	*No*	*Yes*
Phone Accel	58.0	96.8	58.5	97.6	30.3	96.0
Phone Gyro	27.4	61.6	28.6	65.1	27.0	63.1
Watch Accel	27.8	76.0	28.6	77.3	62.7	75.4
Watch Gyro	12.4	39.8	13.2	43.9	51.8	42.4
Phone	61.2	97.0	62.1	97.5	32.7	96.2
Watch	28.6	77.1	29.3	77.9	66.6	80.6
Accel	64.0	**99.2**	63.9	**99.3**	64.0	**98.9**
Gyro	30.3	72.3	30.6	73.0	56.3	72.9
All	64.7	**99.1**	65.1	**99.1**	67.0	**98.9**
Avg.	41.6	79.9	42.2	81.2	43.8	80.5

Without voting (Voting = "No"), the results in Table 5 are not very impressive since the identification accuracies are generally under 70%. But it should be noted that since there are 51 subjects, the baseline strategy of guessing the identity of the subject would yield an accuracy of only about 2%. However, with voting using 5 examples, the results improve significantly. Like the results for models using a single activity, the best sensor combinations to use are the accelerometer for both the watch and phone ("Accel") and the combination of all sensors ("All"). Both combinations achieved around a 99% accuracy rate for all three variations (highlighted in bold in Table 5), again highlighting the benefits of fusing single sensor data. For the single sensors, the phone accelerometer ("Phone Accel") outperforms all other single sensors by a wide margin and actually performs close to the levels of the best sensor combinations ("Accel" and "All").

Looking at the overall average accuracies for the three variations of setting the activity labels, we see little overall difference. As expected, providing the actual activity label ("with label") performs the best while not providing any activity information ("without label") performs worst. Predicting the activity label ("predicted label") performs only slightly worse than the actual label, which is encouraging since predicting the label is more practical than providing the label in most real-world situations where we would want to implement continuous biometrics.

We also explored how the amount of training data impacts identification accuracy. Based on the average biometric performance of the 18 *individual* activities using Random Forest and the two accelerometer sensors (Accel), we found that there are diminishing returns once you have about one minute of data.

Table 6. Authentication EER for using a Single Activity with Random Forest and Voting

Activity	Single Sensor				Fused Sensor					Avg.
	Phone Accel	*Phone Gyro*	*Watch Accel*	*Watch Gyro*	*Phone*	*Watch*	*Accel*	*Gyro*	*All*	
Walking	9.4	9.8	13.2	17.2	8.8	13.9	11.3	10.0	6.8	11.2
Jogging	7.8	10.8	16.2	15.2	9.7	12.7	9.0	11.2	8.3	11.2
Stairs	13.4	12.5	19.3	23.9	9.3	18.9	8.4	14.1	6.9	14.1
Sitting	10.4	23.7	14.5	32.1	8.8	17.0	10.0	21.1	10.2	16.4
Standing	12.1	22.1	16.7	31.6	10.9	15.2	10.0	21.5	7.7	16.4
Typing	8.3	15.4	13.0	20.7	8.9	14.0	8.6	13.3	8.8	12.3
Teeth	10.1	14.0	13.3	20.0	10.2	14.4	10.8	14.9	8.2	12.9
Soup	7.3	19.2	17.0	22.3	6.1	13.3	7.8	17.5	8.0	13.2
Chips	9.9	21.5	14.7	25.9	10.3	18.1	8.5	17.2	8.0	14.9
Pasta	8.0	23.7	14.3	26.6	8.9	18.5	9.0	19.6	5.4	14.9
Drinking	11.3	19.2	16.6	25.1	10.2	13.9	10.9	19.9	8.1	15.0
Sandwich	9.9	17.9	17.5	25.7	11.4	17.7	8.2	16.2	9.3	14.9
Kicking	10.6	19.4	21.0	24.1	11.0	16.6	10.1	18.8	11.0	15.8
Catch	9.7	19.3	16.3	15.5	10.0	14.9	9.3	13.9	10.0	13.2
Dribbling	10.3	21.0	16.4	16.1	9.7	14.5	10.0	11.8	11.5	13.5
Writing	8.7	15.7	10.7	21.3	9.2	11.6	9.0	16.0	10.1	12.5
Clapping	9.4	13.4	12.9	17.2	10.1	13.2	8.1	14.8	8.5	12.0
Folding	7.9	18.6	17.0	23.4	10.0	17.3	8.1	16.2	7.1	14.0
Avg.	9.7	17.6	15.6	22.4	9.6	15.3	**9.3**	16.0	**9.3**	

This is encouraging since it indicates that good performance is possible with a modest amount of training data. These findings are from models without voting since the benefits of voting would obscure the impact of limited training data. Voting is quite effective at improving identification performance, as shown in Table 5. The results in Table 5 only show the results when voting using 5 examples (n=5). Through experimentation, we found that the benefit of voting using additional examples diminishes after n=5.

C. Authentication Results

As described in Section IIID, authentication is a binary classification task that involves verifying the identity of a specific user. To evaluate our authentication models we compute Equal Error Rate (EER), a metric commonly used to compare different authentication models [8]. This metric is calculated as the point where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR). FAR is the rate at which the model incorrectly accepts an imposter as a legitimate user, while FRR is the rate at which the model incorrectly rejects a legitimate user. Both FAR and FRR can be altered by varying the probability threshold value used for assigning a classification. The lower the EER, the better the performance of the model. The EER for each activity and sensor combination is presented in Table 6. Each result in the table is averaged over the 51 individual user models.

Looking at the average performance over all activities in Table 6 (last row), the best sensor combinations to use are the accelerometer for both the watch and phone ("Accel"), and the combinations of all sensors ("All"), both highlighted in bold. Note that these two sensor combinations are the same as the two best sensor combinations for identification (see Tables 3 and 5). Examining the results in more detail, the accelerometer performs best for 4 of the 18 activities, while the combination of all sensors performs best for 8 of the 18 activities.

In terms of individual activities, walking and jogging have the lowest EER while sitting and standing have the highest EER (lower is better). This makes sense since the walking and jogging activities involve a wide range of motions while sitting and standing do not. Other activities like writing and typing have similarly good performance. However, they require equipment such as a keyboard or pen and paper, making them less practical as a biometric trait. Activities involving eating or drinking also suffer from the same problem. These activities had the worst performance overall. Lastly, the exercise related activities (excluding walking and jogging) had relatively good performance. In terms of practicality, these activities are the least practical as they require equipment as well as open space.

V. RELATED WORK

Sensor-based biometrics have been heavily explored due to the wide range of practical applications. As smartphones continue to access and store increasingly sensitive personal information, traditional passwords are inadequate in terms of security, and there is a desire to move towards multi-factor authentication. Other applications include authenticating a user to keep a computer account safe, or even to unlock the front door to your home. Sensor-based biometrics can be carried out either using specialized sensors [9] or built-in sensors on commercial devices [6]. In addition, smartwatches have also been used to perform biometrics [2], which have the advantage of being in a more consistent location on the wrist (as opposed to in a pocket or bag). Our research utilizes both

smartphones and smartwatches because they are more affordable than specialized sensors and can easily be deployed in a real-world application.

Many of the studies that use sensor data from smartphones focus on gait as a person's biometric signature, with excellent results. In a study of 14 test subjects, Hoang et al. achieved an EER of 3.5% using the built-in accelerometer [10]. Participants were asked to walk around a track for twelve laps with the smartphone in their trouser front pocket. In a similar study of 36 test subjects, which examined both "normal" speed and "fast" speed gait, Juefei-Xu et al. managed an EER of roughly 5% using a smartphone accelerometer and gyroscope [11]. However, gait-based biometrics is heavily affected by many factors including a person's shoes, physical state (e.g. injury), walking environment, and can lead to inconsistent system performance [2].

As an alternative, some papers have investigated uncommon biometric signatures. For example, Yang et al. achieved an EER of 6% using snapping data (via microphone) from 22 test subjects collected over 7 days [12]. In another study, Buriro et al. achieved a False Acceptance Rate (FAR) / False Rejection Rate (FRR) of 3-4%, using accelerometer, gyroscope and magnetometer sensor data of a person's hand movements as they type [13]. The study also considered the timing of individual keystrokes. However, our work considers a large number of daily activities as a biometric signature.

VI. Conclusions and Further Work

Our study demonstrates that mobile biometrics is feasible using a commercially available smartwatch and/or smartphone. It also provided some more specific, but very important conclusions. First, sensor fusion improves biometric performance and a combination of the phone and watch accelerometer performs best, followed closely by a combination of the accelerometer and gyroscope sensors on both the phone and watch. Second, a majority voting strategy can dramatically improve biometric performance. Voting using 50 seconds of data (5 examples) can greatly outperform using 10 seconds of data (1 example). Third, one can achieve good biometric performance when using a variety of diverse activities even when the activity is not labeled. A two-stage classification process that first involves predicting the activity label can improve performance. These results are a good step toward continuous biometrics in an unstructured environment. Lastly, a variety of activities can generate useful biometric signatures. Walking and jogging are very effective, but so are writing and typing. Other activities perform nearly as well but are impractical because they require specific equipment.

There are several ways to expand our current research. Data from multiple days will be necessary to verify that the biometric models hold up over time and may assist in building more robust models. More sophisticated features and more sophisticated sensor fusion techniques can also be explored. A more ambitious step would be to move closer to the continuous biometrics scenario by allowing additional activities and allowing different users to have different activity profiles. Finally, this research can be applied to build a real-time biometric system, as in the past we have engineered real-time activity recognition systems.

References

[1] Sensors Overview. Android Developers Guide, https://developer.android.com/guide/topics/sensors/sensors_overview.html

[2] A. H. Johnston & G. M. Weiss. Smartwatch-based biometric gait recognition. In *IEEE 7th Int. Conf. on Biometrics Theory, Applications and Systems,* 1-6, IEEE, 2015.

[3] G. M. Weiss, J. W. Lockhart, T. T. Pulickal, P. T. McHugh, I. H. Ronan & J. L. Timko. Actitracker: a smartphone-based activity recognition system for improving health and well-being. In Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on (pp. 682-688). IEEE.

[4] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda & A. J. Schreiber. Smartwatch-based activity recognition: A machine learning approach. In Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on (pp. 426-429). IEEE.

[5] R. Damaševičius, R. Maskeliūnas, A. Venčkauskas, & M. Woźniak. Smartphone user identity verification using gait characteristics. *Symmetry*, *8*(10), 100, 2016.

[6] J. R. Kwapisz, G. M. Weiss & S. A. Moore, Cell phone-based biometric identification. In *4th IEEE International Conference on Biometrics: Theory Applications and Systems,* 1-7. IEEE, 2010.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, ... & J. Vanderplas. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(Oct), 2825-2830, 2011.

[8] D. Gafurov, K. Helkala & T. Søndrol. Biometric gait authentication using accelerometer sensor. *JCP*, *1*(7), 51-59, 2006.

[9] D. Gafurov & E. Snekkenes. EURASIP Journal on Advances in Signal Processing, 7, 2009

[10] T. Hoang, T. D. Nguyen, C. Luong, S. Do & D. Choi. Adaptive cross-device gait recognition using a mobile accelerometer. *JIPS*, *9*(2), 333, 2013.

[11] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad & M. Savvides. Gait-id on the move: Pace independent human identification using cell phone accelerometer dynamics. In *IEEE 5th International Conference on Biometrics: Theory, Applications and Systems,* 8-15. IEEE, 2012.

[12] Y. Yang, F. Hong, Y. Zhang & Z. Guo. Person authentication using finger snapping - A new biometric trait. In *Chinese Conference on Biometric Recognition*, 765-774, Springer, 2016.

[13] A. Buriro, B. Crispo, F. Del Frari & K. Wrona. Touchstroke: Smartphone user authentication based on touch-typing biometrics. In *International Conference on Image Analysis and Processing*, 27-34, Springer, 2015.

[14] Y. Watanabe & S. Sara. Toward an immunity-based gait recognition on smart phone: A study of feature selection and walking state classification. *Procedia Computer Science*, *96*, 1790-1800, 2016.

[15] C. Nickel, H. Brandt & C. Busch. Classification of acceleration data for biometric gait recognition on mobile devices. *Biosig*, *11*, 57-66, 2011.

[16] M. O. Derawi, C. Nickel, P. Bours & C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing,* 306-311. IEEE, 2010.