Smartwatch-Based Transcription Biometrics

Francesco Ciuffo and Gary M. Weiss
Department of Computer and Information Science, WISDM Laboratory
Fordham Univesity
441 East Fordham Road, Bronx NY 10458
{fciuffo, gaweiss}@fordham.edu

***Abstract*--Handwriting analysis and the identification of people based on individual typing patterns are of great interest in the field of biometrics. The increasing need for user authentication in our modern society has underscored the need for smarter and more automated security protocols. The goal of this research is to explore the feasibility of using relatively inexpensive commercial smartwatches as a reliable platform for transcription-based biometrics. This study demonstrates that it is possible to accurately identify individuals performing transcription tasks (i.e. handwriting sentences, typing on a keyboard, etc.) using only a smartwatch.**

Keywords—smartwatch, biometrics, transcription, security, authentication.

I. INTRODUCTION

The need to provide proof of identity is becoming increasingly common. The majority of people living in a technologically advanced society are asked to authenticate their identity on a daily basis. Such situations are so ordinary they are not given a second thought; entering a digit-based PIN to draw cash from a bank, providing a photo ID at security checkpoints in an airport, or even accessing online accounts with simple alpha-numerical passwords. The purpose of each of these procedures is to provide enough evidence to support a claimed identity. The global increase of automated information systems makes this process of authentication and identification more relevant than ever before.

Biometrics is concerned with measuring unique physical or behavioral characteristics of human beings for the purpose of identification. Given the evolution of our security needs, mobile and continuous biometric protocols are becoming increasingly necessary. Mobile biometrics has become very popular—finger print scanners are available on newer iPhones and an Iris scanner is available on the Galaxy Note 7. The relatively recent introduction of smartwatches introduces a new, powerful, and convenient tool for mobile biometrics. In a study by Johnston and Weiss [2], smartwatches were shown to be effective for gait-based identification and authentication. Relatively accurate results were achieved using simple descriptive statistics generated from non-overlapping 10 second 'windows' of data. Using the accelerometer data from gait measurements, the best authentication model achieved an accuracy of 98.3%.

Using smartwatches as a tool for biometric authentication is attractive for a number of reasons. First, they are relatively cheap (especially the Android smartwatches) and may ultimately become very common. Also, their small form and the fact that they are multi-functional tools make them an ideal solution for a non-invasive authentication system. They are also worn in a consistent position and orientation, which makes them much more suitable for biometric purposes than smartphones. Finally, because they have wireless capabilities and can communicate with other devices by using a smartphone as an intermediary, any biometric information they collect can be used to provide access to these other devices.

By utilizing the motion sensors in a smartwatch, it is possible to modernize and greatly improve conventional biometric practices. Signature and handwriting analysis is a well-known practice, especially in forensics. Traditionally, analyzing a person's handwritten signature involves post-transcription visual inspection. The identifiers are based on the physical appearance of the written sample. However, using a smartwatch to authenticate a written sample has a couple of advantages. First, the data needed for analysis can be captured in real time. Second, the process of handwriting analysis can be automated and carried out by a computer instead of a person. The identifiers then are based on an individual's unique wrist movements while performing the writing task. The complexity of human hand-eye coordination makes a person's handwriting, and in particular signatures, highly individualistic. By utilizing the wrist movements as a biometric, it is ultimately much harder for one person to spoof another person's signature.

The study described in this paper uses a smartwatch to collect tri-axial sensor data from people performing various transcription tasks. Joyce and Gupta [1] have suggested that the same variables that make a signature a unique human identifier also are exhibited in keyboard typing patterns, and for this reason we consider typing as well as handwriting. Furthermore, typing is a very common task, which is often performed for hours each day, so it is especially useful as a biometric activity. It also can be very useful since a great deal of sensitive information is handled on computers. Monitoring a person's typing patterns with a smartwatch allows for continuous biometric security. If a person walks away from their workstation and someone

978-1-5386-1104-3/17/$31.00 ©2017 IEEE

else starts typing on the computer, a typing-based biometric system could quickly detect the intruder.

This study demonstrates that the accelerometer and gyroscope sensors present in inexpensive commercial smartwatches are capable of identifying a person based on the transcription activities of writing and typing. Our approach involves applying classification algorithms from machine learning to the smartwatch sensor data. The induced biometric models are evaluated and are shown to achieve very high accuracy.

II. METHODOLOGY

This section describes the procedure for collecting data, the data processing steps to extract relevant features, the data set utilized for our study, and the design of the biometric experiments.

A. Data Collection

Five different transcription tasks were designed for the participants to perform while wearing a smartwatch on their wrist: (1) writing their personal signature, (2) writing a short sentence prompt with 39 letters on paper, (3) writing a slightly longer sentence prompt with 58 letters, (4) writing two numerical strings with 10 numbers on each line, and (5) typing a sentence with 105 characters (125 with spaces) on a keyboard. These tasks are described in Table 1.

TABLE 1. DESCRIPTION OF TRANSCRIPTION TASKS

ID	Transcription Task	Prompt
1	Prompt 1*	"One good thing about Fordham is Pugsley's Pizza!"
2	Prompt 2	"You cannot escape the responsibility of tomorrow by evading it today."
3	Prompt 3 (numerical prompt)	23 45 64 83 10 12 64 93 45 33
4	Keyboard (typed)	"Fordham University is composed of ten constituent colleges, four of which are for undergraduates and six of which are for postgraduates."
5	Signature	-Personal Signature-

The sentence prompts and numerical strings remain the same for every participant throughout the data collection phase. This consistency ensures that correct user identification is achieved on the basis of unique neurophysiological variables while performing the experiments, and not differences in the content of what is recorded. Otherwise, differentiation between users would be trivial. Before beginning the study, we obtained approval from our university's Institutional Review Board. This was done to gain approval for "experimenting" on human subjects, regardless of there being virtually no risk of injury while performing the tasks.

A total of 24 people contributed to our data set. Each participant strapped an LG G-Watch to the wrist of the hand that they use for writing. The smartwatch is paired with a Samsung Galaxy S4, which runs a custom application for the data collection process. The participants are first asked to provide examples of the first writing prompt. A researcher initiates data collection on the paired phone and the participant waits 5 seconds—there is a delay between initiating the app and the time when the sensors in the watch actually start sampling data—and then writes the prompt once. The participants are told when to start transcribing by the smartwatch, and the data collection terminates once they stop writing or typing. This process is repeated 9 more times, so that a total of 10 examples are recorded. Every participant provided 10 samples of each prompt. The only exception was for the keyboard task, for which each participant recorded 5 examples of the typed sentence prompt.

Each "example", written by participants, is sampled by the gyroscope and accelerometer sensors at 20Hz. Each sensor generates values for the x, y, and z axes and appends a timestamp to the values.

B. Feature Extraction

The raw time-series sensor data must be transformed into examples before the classifier induction algorithms can be applied. An approach similar to the one carried out in a prior study on smartwatch-based biometrics is utilized [2], except a sliding window is not used; rather the data for an entire activity is used to generate a single example. The reason for this difference is that in this case the activity, transcription, is not repetitive in the same way that the "walking" activity was repetitive—since walking normally is made up of nearly identical strides.

The raw data is pulled off the phone running the data collection app and at this point each task example is represented by a series of time-stamped axis values. These values are then used to generate a set of simple descriptive statistics for each example. The accelerometer and gyroscope features are generated independently, but generate the same features. The sampled sensor values are transformed into 32 features, which are outlined in the list below. All features, except for the average resultant acceleration, are based on the sensor values for a single axis, but 3 versions of each feature are generated corresponding to the 3 axes associated with the sensor data. The bracketed numbers represent how many features of a given type are generated.

- Average[3]: Average sensor value (each axis)
- Standard Deviation[3]: Standard deviation (each axis)

- Average Absolute Difference[3]: Average absolute difference between the values and the mean of these values (each axis)
- The Positive Peak[3]: Peak positive value in one example (each axis)
- The Negative Peak[3]:Peak negative value in one example (each axis)
- Average Resultant Acceleration/Angular Acceleration[1]: For each of the sensor samples in one recorded example, take the square root of the sum of the squares of the x, y, and z axis values, and then average them.

Each example is appended with a numerical ID value that uniquely identified each participant. This ID field serves as the class value for the identification task.

C. Data Set

A total of 24 people participated in the experiment. As described in Section 2.1, each person provided 10 examples for each task, with the exception of the keyboard typing task. There are three main categories by which the data is partitioned: Accelerometer, Gyroscope, and Merged features. The "merged" data sets are similar to the other data sets, except they have twice as many features: the features formed from the accelerometer sensor data are concatenated with those generated from the gyroscope data. In each category, there are 5 separate data sets which correspond to the five tasks. Therefore, there is a total of 15 (3 categories × 5 tasks) data sets used to run the identification experiments. Every data set, except the ones containing the keyboard typing examples, have 240 examples (24 users × 10 examples/user). The typing data sets contain 120 examples (24 users × 5 examples/user).

D. Experiments

The WEKA data mining suite was used to implement the classifier induction algorithms. WEKA is freely available and has a large number of tools for preprocessing data, constructing classifier models, and aggregating methods. [11] This study utilizes two of WEKA's algorithms: Multilayer Perceptron (MLP) and Naïve Bayes (NB). These algorithms were chosen because they can be generated and evaluated quickly and thus are applicable to real time biometric identification.

The identification task is to identify a user from the entire pool of participants based on the samples collected from each task. Each participant contributes 10 examples for each writing task and 5 examples for the typing task. Each associated task-specific data set is used to train and evaluate the identification models, using leave-one-out cross validation—which makes the most of our limited data. A single predictive model is generated and evaluated for each task in each category, using both algorithms. The class variable is the User ID. There are 24 distinct class values, corresponding to the 24 participants.

III. RESULTS

The identification experiments involve building a single predictive model to identify a specific user from the pool of users. As mentioned in Section 2.4, leave-one-out cross validation is used to build and evaluate the models. Thus only one example per user per task is tested on at a time, on a model generated from all the remaining examples. The reported generalization error estimate is obtained by repeating this procedure for each of the training examples available and averaging the results.

The results are based on identifying participants based on correctly matching a single example. However, the prediction results can be greatly improved using a simple voting scheme based on the most predicted user. This is a similar approach used by Johnston and Weiss [2]. To better explain the voting scheme, a confusion matrix for one of the identification tasks is provided in Table 2. The rows correspond to the actual users and the columns to the predicted users, so that the values in the diagonals correspond to correct identifications and all other values correspond to errors. This is only a partial matrix taken from the actual 24×24 matrix. The results displayed are based on an identification model generated from Accelerometer and Gyroscope data, and using the Naïve Bayes algorithm.

TABLE 2. Partial Confusion Matrix

USER	101	102	103	104	105
101	**8**	0	0	1	1
102	0	**9**	1	0	0
103	0	0	**10**	0	0
104	0	0	0	**7**	0
105	0	0	1	0	**8**

From the results displayed in Table 2, it's simple to compute the raw accuracies for identifying each user. The accuracy for identifying user 101 would be 80% (8/10). The overall accuracy for a given model is computed by dividing the number of correct predictions by the total number of predictions. From the confusion matrix, it becomes apparent that the diagonal numbers, corresponding to the most predicted user, are the largest. This feature can be utilized to significantly improve the accuracy of the prediction models. All of the predictions for one user are used to assign a single identity. Based on Table 1, participant 101 would be classified as 101 after using all the predicted examples as 'votes.' As such, the accuracy improves from 80% to 100% because incorrect predictions no longer impact the final decision. In the event that a model predicted two users equally, one was chosen at random. However, this voting scheme functions by tallying votes based on multiple provided examples for

a given task. That is, the voting is only feasible if the prompt is repeated and recorded multiple times. This isn't completely unrealistic for biometric security systems.

Tables 3 and 4 show our identification results using a single sensor. Table 3 provides the identification accuracies for the accelerometer sensor, while Table 4 provides it for the gyroscope sensor. The results for both classification algorithms are shown as well as the results without and with the voting optimization. The accuracies are based on all of the predictions over all of the users. That corresponds to 240 predictions (24 users × 10 examples/user) for all but the keyboard task, which is based on 120 predictions (24 users × 5 examples/user). The results clearly show the superiority of the accelerometer sensor over the gyroscope sensor, as well as the effectiveness of using the voting optimization. Finally, MLP generally outperforms Native Bayes.

TABLE 3. Identification accuracies with accelerometer sensor

EXPERIMENT	NAÏVE BAYES	NB VOTING	MLP	MLP VOTING
KEYBOARD	83.2%	100%	95.8%	100%
P1	89.9%	100%	89.9%	100%
P2	91.7%	100%	89.5%	100%
P3	89.9%	100%	86.6%	100%
SIGNATURE	86.5%	100%	83.1%	100%

TABLE 4. Identification accuracies with gyroscope sensor

EXPERIMENT	NAÏVE BAYES	NB VOTING	MLP	MLP VOTING
KEYBOARD	67.5%	73%	72.5%	90.3%
P1	77.5%	90%	85.4%	100%
P2	83.8%	100%	94.2%	100%
P3	71.9%	90%	77.8%	95.8%
SIGNATURE	70.6%	95.2%	78.5%	100%

Improved results are achieved when using both the gyroscope and accelerometer sensor data and associated features, as demonstrated by the results in Table 5. We see that, as before, the MLP models are still superior to the Naïve Bayes models. With most raw accuracies in the mid to high 90's, it can be concluded that identifying users with a smartwatch, performing writing and typing tasks, is feasible. This is especially true when considering perfect identification is always achieved when utilizing the voting scheme.

TABLE 5. Identification accuracies using merged sensor data

EXPERIMENT	NAÏVE BAYES	NB VOTING	MLP	MLP VOTING
KEYBOARD	84.3%	100%	95.0%	100%
P1	95.4%	100%	98.3%	100%
P2	95.4%	100%	98.3%	100%
P3	92.9%	100%	93.3%	100%
SIGNATURE	90.4%	100%	95.2%	100%

IV. Related Work

Past studies with similar goals are described in this section and are compared to our current study.

A. Handwriting Biometrics

Numerous studies have tested biometric authentication systems based on both static and dynamic signature features. A few studies have used pen-like devices, equipped with accelerometer and gyroscope sensors, to record participant signatures and then accurately identified people based on the collected sensor data. Griechisch, Imran, and Liwicki [3] recorded the acceleration and angular momentum of 300 signatures from 20 participants and were able to achieve an 88.5% overall accuracy using a simple nearest neighbor classifier algorithm. Furthermore, Scheidat, et al., [4] was able to extract dynamic and static statistical features in order to characterize sampled handwritten data. The users were asked to write down a 5-digit numerical PIN, chosen by the researchers. The actual device used to collect the data was not mentioned.

Another study [5] used an accelerometer-based pen device. The goal was to create an automatic, real time authentication protocol for handwritten numerical digits. A total of 1000 digits with a set of 10 numerals from 10 users were used to validate the effectiveness of the proposed pen device and algorithm. The overall user-dependent recognition rate was 90.6%.

In summary, all of these studies used custom pen apparatuses designed for the purpose of handwriting analysis. The advantage of using a smartwatch is that smartwatches are commercially available and can be easily implemented in an already existing authentication protocol. There is no need for an additional device used for data capture. Additionally, the algorithms needed to perform the identification can be implemented directly on the watch. This eliminates the need for more external hardware.

B. Typing/Keystroke Biometrics

Another well-researched area of biometrics is the use of keystroke dynamics for user identification and authentication. Keystroke dynamics refers to the unique features that can be derived by the patterns of people typing on a keyboard, such as the latencies between successive keystrokes, keystroke durations, finger placement, and applied pressure. All of these measurements can be used to construct a unique profile.

Many attempts have been made to utilize these features for user identification. A number of separate studies [1,7,8,9,10] ultimately showed that keystroke rhythms are good unique identifiers. Monrose and Rubin [7] collected keystroke data from 63 people over a period of 11 months. Participants ran experiments from their own machines and the collected data was "free-text," meaning subjects were

allowed to type whatever string of letters they wanted. The features generated were based on data collected directly from keyboards. The accuracy of their classifiers, constructed from the full data set, ranged from 83.22% to 92.14%, depending on which classification algorithm was used to build the model. The authors observed that while recognition based on free-text may be more desirable, free-text recognition was subject to variability under various psychological conditions (i.e. in a state of excitement). Therefore, the use of pre-conceived text prompts for training a biometric authentication model can produce a system with high accuracy.

None of the current research on keystroke dynamics utilizes accelerometer or gyroscope sensors to extract typing pattern data. Using smartwatches as a method for extracting unique typing patterns is a novel approach in biometric user identification. The keystroke identification experiments, described earlier, do not utilize the classical keystroke dynamics as features (i.e. latencies, applied pressure, etc.). Rather, simple descriptive statistics, derived from sampled accelerometer and gyroscope sensors, are used as features for the examples fed into conventional classifier induction programs. After reviewing the results, it can be concluded that these features are more than sufficient as a typing biometric.

V. CONCLUSIONS

The results of our identification experiments are conclusive. Smartwatches are effective devices for capturing motion data during transcription tasks. This data has successfully been utilized as a biometric for identifying individuals based on their unique wrist movements.

The tasks in this study were devised to be representative of short phrases that may be used as personal security PINS in a broader based authentication protocol. Although using such a system would not necessarily be sufficient as a single means for identification, the results are strong enough for this method to be part of a biometrics system that utilizes multiple identification/authentication mechanisms.

Of particular interest is the keyboard typing identification experiments. Many studies have used keystroke dynamics as a biometric for identification and authentication. However, all of them collected the data using keyboards and extracted complex features to map users to specific typing patterns. The high accuracy resulting from the simple descriptive statistics, generated from the smartwatch sensor data, shows that a very different and simpler approach can be successful. This is true despite the fact that only 5 typing examples were recorded for each user. It is probable that if more examples were taken, the accuracy would significantly improve. This raises another inherent advantage of our type of biometric system: repeated input of a "security phrase" can be automatically integrated into the data set, so that the system can learn and learn and improve over time.

There are many ways we would like to expand this work. This experiment utilized simple statistical features generated from the raw sensor data. The simple features provided the machine learning algorithms sufficient information for accurate identification. However, it may be possible to develop more sophisticated features to describe the profile of a person's handwriting in a more expressive manner. Doing so would further increase the accuracy of our models. One limitation of the study was that the data collected for each user was collected in a single sitting. It would be more realistic for this type of experiment to collect training and testing data on separate days. A reasonable biometric security system should be able to function over an extended period of time and in varying conditions.

References

[1] R. Joyce & G. Gupta, (1990). Identity authentication based on keystroke latencies. *Communications of the ACM*, *33*(2), 168–176. doi:10.1145/75577.75582

[2] H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Arlington, VA, 2015, pp. 1-6. doi: 10.1109/BTAS.2015.7358794

[3] E. Griechisch,M. Imran Malik, & M. Liwicki, (2013). Online Signature Verification using Accelerometer and Gyroscope. *International Graphonomics Society*

[4] T. Scheidat, M. Leich, M. Alexander, & C. Vielhaur, (2009) Support vector machines for dynamic biometric handwriting classification . In *Proceedings of AIAI Workshop*, pp. 118 -125

[5] J.-S. Wang, Y.-L. Hsu, , & Cheng-Ling Chu (2013). Online Handwriting Recognition Using an Accelerometer-Based Pen Device. *International Conference on Advances in Computer Science and Engineering*

[6] M. C. Fairhurst, (1997). Signature verification revisited: Promoting practical exploitation of biometric technology. *Electronics & Communication Engineering Journal*, *9*(6), 273–280. doi:10.1049/ecej:19970606

[7] F. Monrose, & A.D. Rubin, (2000). Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, *16*(4), 351–359. doi:10.1016/s0167-739x(99)00059-x

[8] J. Leggett, G. Williams, M. Usnick, & M. Longnecker, (1991). Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies*, *35*(6), 859–870 doi:10.1016/s0020-7373(05)80165-8

[9] D. Mahar, R. Napier, M. Wagner, W. Laverty, R.D. Henderson, & M. Hiron, (1995). Optimizing digraph-latency based biometric typist verification systems: Inter and intra typist differences in digraph latency distributions. *International Journal of Human-Computer Studies*, *43*(4), 579–592. doi:10.1006/ijhc.1995.1061

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, & I.H. Witten, (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10. doi:10.1145/1656274.1656278