Exploring the Scalability of Behavioral Mid-air Gestures Authentication

Yasmeen Abdrabou
German University in Cairo
Bundeswehr University Munich
yasmeen.abdrabou@gmail.com

Nadeen Mourad
German University in Cairo
nadeen.murad@student.guc.edu.eg

Amr Elmougy
University of Canada in Egypt
amr.elmougy@uofcanada.edu.eg

ACM.
MUM '18,, November 25–28, 2018, Cairo, Egypt
ACM 978-1-4503-6594-9/18/11.
https://doi.org/10.1145/3282894.3289725

Abstract
Gesture-based authentication systems are gaining increasing attention from the research community due to their promising usability. However, the scalability of these systems has not been properly investigated against the number of users and the number of gestures. Accordingly, in this paper, we explore the scalability of mid-air gesture-based systems in both aforementioned dimensions to enhance the already existing systems. We implemented a gesture-based authentication model with 20 gestures and we invited 39 users for data collection. A Support Vector Machine (SVM) classifier with Grid search cross-validation was used for training to prove the concept of the model's prototype. The obtained results proved that with the upscaling of the system from the aspect of the number of users, performance gets worse. On the other hand, as gestures introduced to the system increases, the performance improves.

CCS Concepts
•Human-centered computing → Human computer interaction (HCI);

Author Keywords
User Authentication; User Identification; Gestures; Behavioral biometrics

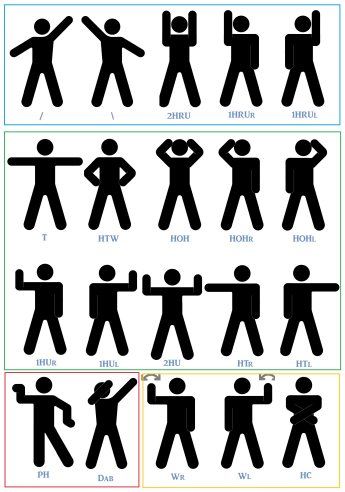

Figure 1: figure

We explore the scalability of mid-air gesture-based authentication systems. Both on the level of the number of users defined and on the level of the number of gestures. The figure shows the 20 used gestures for testing. Each user was assigned to 10 random gestures for authentication. We found a negative correlation between the and the number of users. In addition, we found a positive correlation between the and the number of gestures.

Introduction

Ubiquitous computing technologies have now become widespread in our daily lives. People use these technologies for a limitless number of application domains, where access to private user information is quite often granted. Accordingly, research into authentication mechanisms for ubiquitous computing technologies has propelled over recent years. Modern devices typically come equipped with an arsenal of diverse authentication mechanisms that include biometric mechanisms such as fingerprint, facial, and iris recognition, and non-biometric mechanisms that include numeric and graphical passwords. Research has shown that many of these authentication mechanisms have a number of vulnerabilities. For example, Eiband et al. [6] show that the probability of success of a shoulder surfing attack can reach 91%. In another paper [1], the authors explored thermal attacks on two authentication mechanisms for smartphones "pins and patterns". The study found that thermal attacks are viable and some cases even have a success rate of 100% (non-overlapping patterns). Thus, research has started exploring different biometric authentication systems such as fingerprints [3], voice [12], gait [7] and Iris scanning [10]. These mechanisms have the potential of addressing some of the aforementioned vulnerabilities.

Another biometric authentication mechanism that has been gaining increased attention in the research community is gesture-based authentication. This is because it is an unobtrusive mechanism that comes naturally to humans. Thus, gesture-based systems potentially have a usability edge over other authentication mechanisms. However, none of the already existing gesture-based authentication systems explored the system's scalability, neither against the number of users nor against the number of gestures that the system can recognize. The scalability of these systems has a significant impact on their adoption. Accordingly, this paper presents a comprehensive study of the scalability of gesture-based authentication against the number of users and gestures.

Related Work

One of the existing gesture systems was done by George et al. [8], where the authors evaluated a mid-air version of Android patterns for immersive virtual environments. The users wear a Head-Mounted Display (HMD) and authenticate by drawing a pattern on a virtual 3x3 grid using a handheld controller. Another gesture-based system [9] implements identification using behavioral attributes of a few chosen hand gestures. Their work focused specifically on the waving gesture and obtained an EER[1] of 5% by combining behavioral attributes with body segment lengths. The study found that the combination of body lengths and gestures gave more accurate results than using each scheme individually. It was also found that an untrained system suffers increasing EER against the number of registered users. Accordingly, SVM (Support Vector Machine) classifier was used to train the model and caused obvious improvement in the performance. However, the model was not tested on a larger scale of users (25 participants against 7 registered users). Another similar research [4] uses hand geometry and gestures for authentication. A Leap Motion sensor was used to detect gestures and both static and continuous gesture authentication were studied. The model was tested on 16 participants and an EER of 0.8% was reached. This study showed that the Leap Motion combined with a random forest classifier was successful in identifying users with high accuracy rates. A different perspective was taken by Lai et al. [11], where they extracted features from body silhouettes. They used gesture-based authentication by

[1] Equal Error Rate

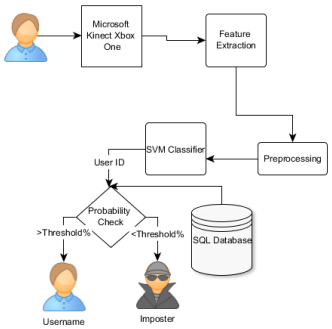

Figure 2: Model Flow Diagram

combining shapes and relative sizes of body parts, with gestures that can be repeated. The study reached an EER of 5-6% for user authentication on single gestures. The model had 20 registered users performing 8 different gestures. Finally, their results were encouraging for a limited number of users having multiple gestures. Aslan et al. [2] also explored mid-air gestures and reached an average of 11.71 % EER. The idea was to use mid-air hand gestures to avoid contamination as a result of dust circulation in medical related clean rooms and fabrication labs.

As a result of the previously mentioned studies, we will focus on exploring the scalability of the mid-air gestures as an authentication method. We will study the relationship between the increased number of gestures versus the increased number of users. We will use the SVM classifier as a machine learning technique.

Model Implementation

We used Kinect v2[2] to capture the gestures. The Depth sensor in the Kinect cannot be fooled by playing a prerecorded video in front of it [11]. Therefore, this feature makes it suitable for authentication applications. The implemented model is made up of two modules, one for signing in and another one for registration. The flow diagram, seen in Figure 2 shows an overview of the model.

When a user performs a gesture in front of the Kinect, predefined features get extracted. Then preprocessing is performed on the whole dataset. If the classifier predicts with a certainty higher than a decided threshold, the user is identified and validated, otherwise, the sign-in attempt is considered to be an attack as seen in equation 1. The gestures we considered in our model consist of 18 static (non-moving) gestures and 2 simple moving gestures. Most of the gestures were simple and relied on different arm orientations as seen in Figure 1.

$$Value = \begin{cases} > \ threshold, & \text{Attacker} \\ < \ threshold, & \text{Logged in User} \end{cases} \quad (1)$$

Following the approach in [9], we decided to extract numerical features instead of graphical features [11]. By using graphical features/image sequence the model does not give a real-time response. So we calculated a total of 55 features and used them in the model. The extracted features can be seen in table 1. We used 36 behavioral features which were extracted from joint positions of the human body which were provided by the Kinect. Also, 19 physical features representing different body segment lengths were used to help in the identification process. Combining both kinds of features was found to yield good results [9].

Feature Extraction

Physiological features are physical characteristics which belong to a user. Though they are not considered unique enough on their own, they still help in the identification of a person among other features. The physiological features taken into account are body lengths. They are extracted by calculating the magnitude of the distance between two points in 3D space, each point representing a joint. There is a total of 19 body part lengths considered in the model.

Behavioral features are more distinct attributes that represent how a user carries out the gesture. Our behaviors tend to be unique in carrying out the same tasks, therefore it was necessary to capture these features to evaluate the behavioral aspect of the model. Figure 1 shows the exact features our model captures when a user performs a

[2] https://support.xbox.com/en-US/Xbox-on-windows/accessories/kinect-for-windows-v2-setup

Table 1: Extracted behavioral features.

Feature type *Type*	**Feature name** *Name*	**Description** *Desc*	**Features** *Number*
Angles	Hand-Wrist-Elbow (R&L) (Mean, Min, Max)	Angle at the wrist between Hand and Forearm	6
	Wrist-Elbow-Shoulder (R&L)(Mean, Min, Max)	Angle at the Elbow between Wrist and Shoulder	6
	Elbow-Shoulder-Spine (R&L)(Mean, Min, Max)	Angle at the shoulder between Elbow and Spine	6
Relative Positions	Wrist Position (R&L)(X, Y, Z)	Wrist relative to spine	6
	Elbow Position (R&L)(X, Y, Z)	Elbow relative to spine	6
Velocities	Hand Velocity (R&L)	Average speed of hand movements	2
	Wrist Velocity (R&L)	Average speed of wrist movements	2
Accelerations	Wrist Acceleration (R&L)	Average acceleration of wrist movements	2

gesture in front of the Kinect. In this phase, the model receives a stream of input from the Kinect and then the feature accumulators are averaged and stored for registration or evaluated for signing in. The readings are calculated as follows: Body part lengths are calculated using the Euclidean distance. When obtaining angles, three positions were tracked. The two needed vectors are obtained using these three points. The angles were calculated as seen in Equation 2. Relative positions were calculated by subtracting the X, Y and Z coordinates separately from a fixed joint position's coordinates. To calculate the velocity, in every two consecutive frames, using the joint positions, we divided the distance the joints moved between frames by the duration between the capturing of the two frames. To calculate the acceleration, divide the difference in speeds between the two frames by the duration between the capturing of the two frames, to then obtain an average of all these accelerations.

$$angle(v,u) = cos^{-1}(\frac{v.u}{||v|| * ||u||}) \quad (2)$$

Gestures Choice

The gestures were chosen to have a combination of distinct positions and overlapping ones. In addition, they were chosen to cover similar and different gestures (in terms of elbow angle and hand orientation) to test the model's accuracy and ability to differentiate between them.

Support Vector Machine for User Classification

The authentication model trains a support vector classifier with user samples. The SVM was chosen as it is the simplest one for proving the concept. The classifier is trained using 20 samples per user. The samples were taken over two sessions (two consecutive days) to ensure factors such as clothing and user behavioral changed over the day do not affect the performance of the classifier. If participants made different gestures than the 20 known ones, then the classifier maps it to the similar one defined. Finally, we used SciKitLearn library[3] for grid search and cross validation.

[3]SciKitLearn Machine Learning Libraries: http://scikit-learn.org/stable/index.html

Evaluation Methodology

In order to evaluate our model, a user study was conducted. The experiment setup was not facing any doors or windows, to avoid any external movements from interfering with the readings. The room was artificially lit and all window curtains were closed. We made sure that the field of view of the Kinect was not obstructed by any object or piece of furniture. The Kinect sensor was placed at a height of approximately 1.2m. A constant distance of approximately 2.5m was obtained between the Kinect and the user. A triangular space was set up using tape to specify the limits of the field of view of the Kinect in the X-axis. The experiment setup is shown in Figure 3.

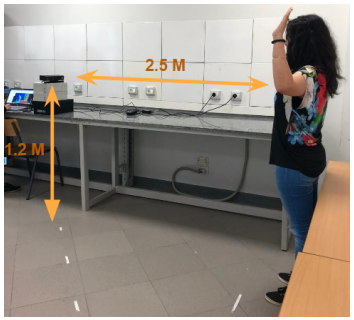

Figure 3: Experiment Setup

Participants

A total of 39 participants took part in the experiment, 25 were male while 14 were female. Their ages were all between the range of 20-27 years (M = 21.72, SD = 2.03). Their heights were in the range from 155cm to 195cm (M = 169.78, SD = 10.98). The participants were recruited by a word-of-mouth.

Experimental Procedure

After the participant arrive to the lab the purpose of the study was explained and they were introduced to all possible gestures they could perform "see Figure 1". Then, they were asked to fill in a consent form and fill a demographics questionnaire. After that, Each participant was assigned to 10 random gestures of the predetermined 20 gestures. The number of gestures was chosen as for the experiment not to exceed 60 minutes. The user was guided to the exact position to stand on and when to start performing the gesture. The participants showed in 2 different sessions to register and sign in to simulate daily usage. During the registration session, the user was asked to perform every gesture 3 times to feed the classifier and to extract the samples needed. After all registration and training sessions are done, registered participants were called back in a different day to sign-in by having 3 sign-in attempts. Participants were not given any feedback regarding their success or failure to login in order not to affect the way they reproduced their gestures.

Evaluation Measures

In order to evaluate the model, we measured the performance using the FAR[4] and the FRR[5]; the way they are calculated can be seen in the Equations 3 and 4. We also calculated the EER for different scenarios. The EER is calculated by finding the point at which the chosen threshold results in the FAR being equal to the FRR.

$$FAR = \frac{Number\ of\ False\ Acceptances}{Total\ number\ of\ Identifications} \quad (3)$$

$$FRR = \frac{Number\ of\ False\ Rejections}{Total\ number\ of\ Identifications} \quad (4)$$

Gestures Evaluation

Semi structured interviews took place in order to evaluate the user's perception of the gestures. From the answers received they were categorized into four categories highlighted by blue, green, yellow and red as seen in Figure 1. The blue category was not easy for users to perform and they felt unnatural. On the other side, the green category was convenient and easy for users to perform. The red category was not perceived as a serious task (for authentication context). Finally, the yellow category was complex for the kinect sensor to detect skeletal data accurately due to movement and intersections.

[4]False Acceptance Rate
[5]False Rejection Rate

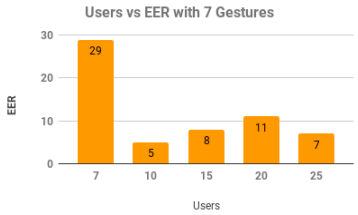

Figure 4: EER for 7 gestures

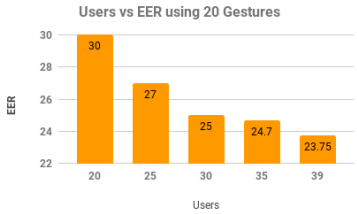

Figure 5: EER for 20 gestures

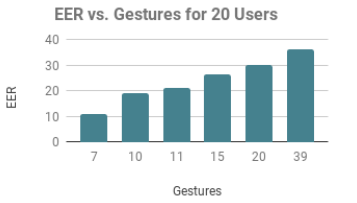

Figure 6: EER Comparison for 20 users with different number of gestures

Results

We explored the performance of the implemented model in multiple scenarios. First, we tested the model for identification accuracy by having one unique gesture for each participant. After adjusting the hyper-parameters using a Grid Search Cross Validation function as suggested by Cuturi [5], we found that the best training and testing scores were yielded by the following values for the hyper-parameters of the SVM: C=10, gamma=0.001 and kernel='rbf'. The final scores for both training and testing classification were 100%. Furthermore, the training datasets' cross-validation score was 95.8% while the testing datasets' cross-validation score was 100% (2 Fold). Also the precision, recall, f1-score scored 100%. These results prove that our model at this scale can perform identification reliably for authentic users.

In order to calculate the FAR, FRR and the EER of the model, we tested multiple scenarios. We first tested the model against a small number of gestures (7 gestures) and a different number of users. We could not find any correlation between the EER and the number of users as the model behavior with a small number of gestures is unstable as seen in Figure 4. While testing the model against a larger number of gestures (20 gesture), a negative Pearson correlation between the EER and the number of users was found, $r = -0.949$, $n = 5$, $p = 0.014$, as seen in Figure 5. On the other hand, when the model was tested with a fixed number of users (20 users) against different numbers of defined gestures, a positive Pearson correlation was found, $r = 0.892$, $n = 6$, $p = 0.017$ as seen in Figure 6.

Discussion

The identification results were successful due to the ability of the SVM to form distinct patterns for each class in its dataset. Results suggest that, as the gestures considered in the system increased, the EER increased (see Figure 6). In cases were the explored gestures were equal to the number of users in the system, we believe that the increase in the EER is justifiable. The dataset, in this case, is made up of classes which are representing both unique gestures and unique users. This is also due to the fact that the features describing the gesture are more than those describing the user, causing features related to the gesture to stand out. Figure 5 confirms that when the number of users increases to a certain limit the EER decreases. At that limit, the classifier's ability to perform user authentication is at its peak. Looking more into the logic of the authentication part of the system. As it is based on probability, the sum of all prediction probabilities will always be 100%. Meaning that when the number of users increases, the threshold of acceptance decreases. In addition, as the threshold of acceptance decreases, the range of possible values of probability between the true users and imposters decrease. This causes the FAR and FRR to increase eventually causing the increase in s. The main limitation is that we didn't compare between different classifiers but the idea was to prove the concept of scalability.

Conclusion and Future Work

This paper explores the scalability of behavioral mid-air gesture authentication. The scalability is studied in two aspects, gestures and users. A model was implemented to collect registration and sign in data. The model used a multi-class SVM. User's data was collected through experiments using a Kinect. The model was tested against different scenarios and a positive correlation was found between the and the number of gestures. In addition, a negative correlation was found between the and the number of users. Further studies can test more complex classifiers for in depth study of the gesture analysis.

REFERENCES

1. Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. 2017. Stay Cool! Understanding Thermal Attacks on Mobile-based User Authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3751–3763. DOI: http://dx.doi.org/10.1145/3025453.3025461
2. Ilhan Aslan, Andreas Uhl, Alexander Meschtscherjakov, and Manfred Tscheligi. 2014. Mid-air Authentication Gestures: An Exploration of Authentication Based on Palm and Finger Motions. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*. ACM, New York, NY, USA, 311–318. DOI: http://dx.doi.org/10.1145/2663204.2663246
3. Raffaele Cappelli, Dario Maio, Davide Maltoni, James L Wayman, and Anil K Jain. 2006. Performance evaluation of fingerprint verification systems. *IEEE transactions on pattern analysis and machine intelligence* 28, 1 (2006), 3–18.
4. Alexander Chan, Tzipora Halevi, and Nasir D. Memon. 2015. Leap Motion Controller for Authentication via Hand Geometry and Gestures. In *HCI*.
5. Marco Cuturi. 2011. Fast global alignment kernels. In *Proceedings of the 28th international conference on machine learning (ICML-11)*. 929–936.
6. Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding Shoulder Surfing in the Wild: Stories from Users and Observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4254–4265. DOI: http://dx.doi.org/10.1145/3025453.3025636
7. Matteo Gadaleta and Michele Rossi. 2018. IDNet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognition* 74 (2018), 25 – 37. DOI:http://dx.doi.org/https://doi.org/10.1016/j.patcog.2017.09.005
8. Ceenu Goerge, Mohamed Khamis, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. In *Proceedings of the Network and Distributed System Security Symposium (USECâĂŹ17). NDSS. DOI: http://dx. doi. org/10.14722/usec*.
9. Eiji Hayashi, Manuel Maas, and Jason I. Hong. 2014. Wave to Me: User Identification Using Body Lengths and Natural Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3453–3462. DOI: http://dx.doi.org/10.1145/2556288.2557043
10. Anil K Jain, Arun Ross, and Salil Prabhakar. 2004. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology* 14, 1 (2004), 4–20.
11. Kam Lai, Janusz Konrad, and Prakash Ishwar. 2012. Towards gesture-based user authentication. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*. IEEE, 282–287.
12. SecureAuth. 2016. UNDERSTANDING BEHAVIORAL BIOMETRICS. (2016). Available at https://www.secureauth.com/resources/blog/understanding-behavioral-biometrics/.