Wave to Me: User Identification Using Body Lengths and Natural Gestures

Eiji Hayashi[1] ehayashi@cs.cmu.edu
Manuel Maas[2] manuel.maas@gmx.de
Jason I. Hong[1] jasonh@cs.cmu.edu

[1]Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Ave. Pittsburgh PA 15213, USA

[2]Institute of Cryptography and Security
Karlsruhe Institute of Technology
Kaiserstraße 12, 76131 Karlsruhe, Germany

ABSTRACT
We introduce a body-based identification system that leverages individual differences in body segment lengths and hand waving gesture patterns. The system identifies users based on a two-second hand waving gesture captured by a Microsoft Kinect. To evaluate our system, we collected 8640 gesture measurements from 75 participants through two lab studies and a field study. In the first lab study, we evaluated the feasibility of our concept and basic properties of features to narrow down the design space. In the second lab study, our system achieved a 1% equal error rate in user identification among seven registered users after two weeks following initial registration. We also found that our system was robust even when lower body segments could not be measured because of occlusions. In the field study, our system achieved 0.5 to 1.6% equal error rates, demonstrating that the system also works well in ecologically valid situations. Lastly, throughout the studies, our participants were positive about the system.

Author Keywords
User Identification; Gesture; Natural User Interface;

ACM Classification Keywords
H.5.2 User Interfaces: Interaction styles.

INTRODUCTION
Many future visions of interactive computing have us moving away from keyboard and mouse interactions, opting instead for natural user interfaces that bridge the physical and virtual worlds. These natural user interfaces might include input modalities such as speech, handwriting, or gestures. Here, we focus just on physical gestures.

While there has been a great deal of exploration into what the user experience might be like for gesture-based interfaces in smart environments, there has been little work on how one might manage user identification. In GUI-based systems, a user specifies his identity by choosing it from a list of users or typing a user ID; the system then typically asks for a password to verify the claimed identity.

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.

http://dx.doi.org/10.1145/2556288.2557043

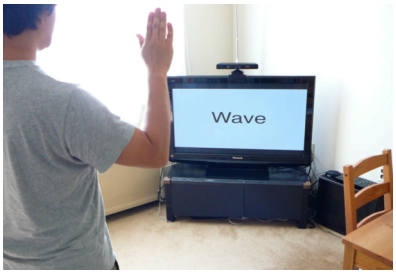

Figure 1. The proposed system can identify users based on 2 seconds of hand waving gesture measured by a Microsoft Kinect.

However, this approach assumes the existence of displays and keyboards (either physical or virtual), which may not hold or make sense, especially if gesture-based interfaces become popular. For example, many electric appliances, such as air conditioners, room lights, thermostats, and stereos, have limited input and output capabilities, but could potentially benefit from gesture-based controls as well as personalization based on user identification. A simple example would be turning on an air conditioner with a gesture and setting a personalized temperature based on user identification. Other examples might include being able to access game consoles, parental controls, personalization for electric appliances like DVRs, or showing some personalized information on wall displays (such as today's schedule or feeds from SNSs). Furthermore, many of these applications will be deployed in places with reasonably good physical security, such as homes, meaning that user identification can be designed in favor of usability over security.

In this paper, we investigate a novel body-based identification scheme that uses a two-second hand wave gesture captured by a Microsoft Kinect to identify users (Figure 1). We evaluated our system through three user studies. In our first study, we evaluated the feasibility of using body segment lengths as well as four natural gestures for user identification. Our analysis showed that body segment lengths and gesture patterns were not reliable enough when used separately; however, when we combine body lengths with a natural gesture, the system achieved reasonable performance. Furthermore, we found that a hand-waving gesture yielded the best performance among the four gestures. In our second user study, we further evaluated hand-waving gesture through a two-week lab

study. The study showed that user identification was stable over this timeframe and performed well when trained with data from two sessions. In our third study, we tested our system in participants' living rooms. Our system still had good performance here, where environmental conditions were not controlled and some partial body occlusions occurred. Finally, throughout the three studies, our participants reported that they were positive about body-based identification system.

In summary, we make the following three research contributions. First, we measured, in the context of user identification, the reliability of body segment lengths and gesture data extractable from Microsoft Kinect's standard APIs. Second, we developed a body-based identification system that works on the basis of body segment lengths and hand-waving gesture patterns, and evaluated it through three user studies consisting of 75 participants in total. Third, we measured the robustness of our body-based identification to partial-body occlusions, the reliability over a two week period, robustness against environmental changes, and participants' responses to the system.

RELATED WORK

Many pieces of past work have leveraged gestures captured by sensors for user identification and authentication. User identification and authentication are similar but distinct concepts. User identification is a process where a system identifies a user among a set of pre-registered users and verifies the identity, whereas user authentication is a process where a system verifies whether a user is actually the pre-registered user that he claims to be [12]. In other words, a system performing user identification does not know a user ID, whereas one performing user authentication does.

Several lines of past work have used gestures with devices for authentication. For example, Chong et al. used discrete gestures consisting of a series of orientations as a password for mobile phones [1]. Jiayang et al. used a tri-axis accelerometer that yielded 2% false rejections and 12% false acceptances for authentication [9]. Patel et al. developed a system that authenticates users to public terminals through gestures captured by a mobile device [18]. Mayrhofer and Gellersen used a shaking gesture for device-device authentication [16].

There have also been authentication systems using Microsoft Kinect. For example, KinWrite authenticates users based on handwriting in mid-air [23]. With BroAuth, a user touched her own and her virtual partners' bodies in a sequence to authenticate [15]. For all of the systems above (mobile and Kinect), users need to first register and memorize a secret gesture. Although this approach has advantages on small devices where typing text is hard, the memorability of these gestures is not clear, especially if there are multiple gestures that need to be memorized. In contrast, our system uses a hand-waving gesture for all users; thus, it does not require any memorization.

Kam et al. proposed a hand-gesture-based authentication system for surveillance cameras that requires the same gestures for all users. The system authenticated people based on a series of eight pre-determined gestures [10]. In a more HCI related domain, biometric-rich gestures [17] required users to perform the same multi-finger gesture on touch displays (e.g., touch a display with five fingers and then close the fingers). Their system achieved an equal error rate (EER) ranging from 6.6% to 18.9%, depending on the gestures. Our work differs from these works in that it combines body segment lengths and gestures to improve performance, and uses a two-second hand wave gesture rather than series of gestures or multi-finger gestures.

Biometrics are another class of research using sensor data for user authentication / identification. Biometrics can be classified into two categories [8]. The first category leverages users' *physiological* properties, such as fingerprints (e.g., [1,14]), face, or voice. For example, Heusch et al. developed face authentication with a 2.4% EER under controlled background and light conditions. However the EER became 13.49% under uncontrolled background and light conditions [6]. Wan et al. demonstrated a 4.03% EER using voice [24]. Schmidt et al. developed HandsDown, which identified users based on hand contour [21]. Our paper explores a new design space for user identification in the context of gesture-based interfaces.

Human body sizes, primarily studied in the context of ergonomics design [4,20], can be another physiological property used for biometrics. In 2010, Microsoft released the Microsoft Kinect, which uses a depth camera to identify individuals as well as gestures. Although the details of the Kinect user identification algorithm have not been disclosed, Leyvand et al. as well as Gantenbein reported that the identification was based on a person's height, face, and clothing [2,11]. The user identification is done without explicit user inputs because it is primarily designed for user identification during gaming. As a result, its accuracy is relatively low. Our work investigates ways of improving user identification using a Kinect.

The second category is *behavioral* biometrics, which authenticates / identifies users based on behavioral characteristics, such as key typing pattern [19] and gait patterns [5]. For example, Mäntyjärvi et al. achieved 7% to 19% EER for gait pattern authentication [11]. These approaches are related to our system in a sense that users are identified based on individual differences extracted from common behaviors. However, as far as we know, natural gestures, such as hand-waving gesture, have not been studied in the context of behavioral biometrics.

Finally, there are systems that combined multiple biometrics [7]. Snelick et al. demonstrated that combining fingerprint and face recognition could achieve a 0.64% EER [22]. These pieces of past work focused on achieving high accuracy to improve security. In contrast, our work

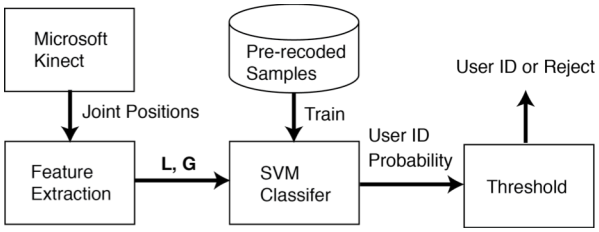

Figure 2. Overall algorithm of body-based identification. A body length feature vector (L) and a gesture feature vector (G) are extracted from joint positions. The vectors are concatenated and used for user identification with a SVM classifier trained with pre-recorded samples.

focuses on achieving high usability with reasonably good security by combining multiple factors.

BODY-BASED USER IDENTIFICATION

Our body-based user identification makes use of both physiological and behavioral properties of users, namely body segment lengths and gestures. We decided to use both of these properties because our preliminary analysis indicated that neither of them was reliable enough for user identification by itself. Our prototype consists of a computer, a display for showing the user instructions, and a Microsoft Kinect.

At a high level, our system works as follows (Figure 2). First, the system is trained by recording users performing a pre-specified gesture (e.g., hand waving). The recordings contain trajectories of three-dimensional positions of users' joints. In the training phase, our system extracts a body length feature vector and a gesture feature vector from the recordings. These two vectors are concatenated and used to train a SVM model that identifies the users. The system trains pair-wise SVM classifiers using a polynomial kernel (exponent=1) with logistic models for all possible pairs of users, and combine these classifiers to build a multi-class classifier that predicts user IDs with probabilities. We also tried Decision Tree, k-Nearest Neighbor, and Naïve Bayes as classifiers. Based on evaluations using a subset of our data, we chose SVM as well as its kernel and parameters.

Once the model is built, the system identifies the users and verifies their identities when they repeat the gesture once. The system extracts the two feature vectors, feeds them into the classifier, and predicts a user ID with a probability. When the probability is higher than a pre-defined threshold, the system accepts the user as the predicted user. Otherwise, the system rejects the user as a stranger. The system needs this threshold because, without the threshold, any person can be identified as one of the registered users. All registered users perform the same gesture, so the system discriminates users based on systematic differences in body segment lengths and gesture patterns. In the following, we describe the process more in detail.

Measurements of Gestures

Our system measures gestures with a Microsoft Kinect using APIs provided in Microsoft Kinect SDK 1.5. When a user stands in front of the Kinect and raises her dominant hand to shoulder level, the system initiates a measuring process. The system then asks the user to perform a hand-waving gesture (or other gestures in the user study #1). While a user is performing the gesture, our system records the trajectories of 21 joints [3] every 20 milliseconds. In total, the system obtains 100 three-dimensional position measurements of the 21 joints over two seconds. In our studies, data was processed off-line using Matlab and Weka. Processing one measurement took less than 0.5 seconds on a Macbook Pro with 2.6GHz Intel Core i7.

Features	Description	# of features
Positions	10, 30, 50, 70, and 90 percentiles of wrist positions for each axis	15
Angles (hand - wrist-elbow)	Max, min and mean of the 3D angles consisting of hand-wrist-elbow	3
Angles (wrist-elbow-shoulder)	Max, min and mean of the 3D angles consisting of wrist-elbow-shoulder	3
Speeds	Mean absolute speed of hand, wrist and elbow calculated in 3D	3
Accelerations	Mean absolute acceleration of hand calculated in 3D	1
# of gesture sequences	Mean cross freq. of wrist positions. 0 for static gestures	1

Table 1. 26 features consisting of a gesture feature vector.

Gestures

Our body-based identification requires all registered users to perform the same gesture. In user study #1, we tested four gestures (hand waving, come-over, one-hand raised, and phone-to-ear) in two postures (standing and sitting), to investigate which gesture yields the best results. Based on the results from the study, we used hand-waving gestures in user studies #2 and #3.

We chose the four gestures to be tested in the study #1 as follows. First, we wanted gestures that people were familiar with performing, so that people could easily repeat the same pattern. Second, we wanted gestures that could be performed by one arm, because gestures performed by other parts of the body (e.g., nodding) could be too simple to discriminate between many users. We also avoided gestures requiring synchronizing movements from multiple body parts as we felt these would be too complicated for most users to reproduce precisely. Furthermore, we hypothesized that there would be a trade-off in reproducibility and discriminatory power. For example, complex gestures that require a lot of arm movements should have very distinct performance profiles across users; however, precisely reproducing these gestures could be difficult for users. In contrast, simple gestures that involve little movement should be easy for users to reproduce but might be performed similarly by other users, as well.

Considering these, we chose two *kinetic* gestures that involved continuous arm movements (hand waving and come-over), and two *static* gestures that involved little movement (one-hand raised and phone-to-ear gestures) to be investigated in the user study #1.

Feature Vectors

For each measurement, the system extracts a *sample* consisting of a body segment length vector and a gesture vector. Our system calculates the Euclidian distances between 17 pairs of adjacent joints. The distances are averaged over the 100 captured frames to mitigate sensor noise. Gesture features are extracted from the trajectory of joints of the user's dominant arm, relative to the user's head. From this trajectory, our system extracts 26 features that comprise a gesture feature vector (see Table 1). Except for the position features, all features were calculated in three-dimensional coordinates. We initially extracted more features, then, chose these features based on evaluation with a subset of our data. While these features worked reasonably well to demonstrate feasibility of our concept, we do not claim these are the best features.

Evaluation Metrics

In our body-based identification, there are two types of errors: false acceptances and false rejections. False acceptances denote the cases where a registered user is identified as a different user. False rejections denote cases where a registered user is rejected. The False Acceptance Rate (FAR) and False Rejection Rate (FRR) were calculated as shown in the Eq. (1) and (2).

$$FAR = \frac{Number\ of\ False\ Acceptances}{Number\ of\ Identifications} \quad (1)$$

$$FRR = \frac{Number\ of\ False\ Rejections}{Number\ of\ Identifications} \quad (2)$$

We used Equal Error Rates (EER) to evaluate performance of the body-based identification rather than the classification accuracy because false acceptances are more serious errors than false rejections, and simple classification accuracies do not make a distinction between the two types of errors. EER is the error rate where FAR and FRR become equal when changing the threshold for rejections. Therefore, smaller EERs indicate better performance.

In this paper, to calculate EER, we simulated having seven registered users to avoid the effects of having a variable number of participants. Our three studies consisted of 36, 25, and 12 participants. If we used all participants as registered users when calculating EERs, we cannot easily compare our results across studies, especially since user identification becomes more difficult when the number of users in a dataset increases. Additionally, for our expected use cases (e.g., home situations), we felt it would not be common to have more than seven registered users. Thus, in the following analyses, we randomly chose 10,000 combinations of seven participants from a dataset, and calculated EER over all combinations.

USER STUDY #1: FEASIBILITY EVALUATION

The overall goal of this lab study was to investigate basic properties of body segment lengths and gesture patterns captured by Microsoft Kinect. Additionally, we chose the most promising gesture among the four gestures based on the result. The chosen gesture was further investigated in the following studies.

Participants

We recruited 36 participants from an existing university recruiting website meant for the general public. Out of the 36 participants, 14 were male and 22 were female. Participants were 32.6 years old on average, ranging from 19 to 64. Twenty-three participants were university students, two were university staff employees and the other eleven were employed. Participants were 168cm tall, on average (SD=10.2), and weighed 78.0kg (SD=22.0). We paid $40 for completing the study.

Data Collection

The lab study consisted of two sessions. All sessions were conducted in the same windowless room, to avoid sunlight interference with the Kinect measurements. Both sessions took about 40 minutes to complete. In the first session, we explained that our system recorded participants' gestures using a Microsoft Kinect and identified them using their gestures as well as their body segment lengths. We then asked participants to perform the four gestures both standing and sitting, 10 times each. Participants stood three meters away from the Kinect. While standing, participants were asked to move half a meter to the left and back between measurements to mitigate habituation. While sitting, we asked participants to stand up and sit back down between measurements for the same reason. Half of the participants started the measuring session standing and the other half started sitting. We counterbalanced the order of the gestures using a Latin square. Three days later, in the second session, we repeated the procedure except explanation of our system. At the end of study, we asked participants to fill out a survey.

We collected 80 (10 samples times eight gesture-posture pairs) for each participant. In total, we collected 5760 samples from the 36 participants. Each sample consisted of a body segment length and a gesture feature vector. All data was processed off-line after the study. Thus, participants were not given feedback about the classification results.

Evaluating Feature Vectors Separately

Before combining the body segment length and gesture feature vectors, we started by analyzing them separately to understand their properties.

Performance Within Session and Posture

As baselines for the following analyses, we first investigated the consistency of the body segment length and gesture feature vectors in the same session and posture through 10-fold cross validation.

Figure 3 shows that identification through body segment lengths alone had EERs around 0.6% in the standing conditions and 1.5% in the sitting conditions. This indicates that the body segment lengths measurements were consistent in the same postures in one session. Additionally, the come-over gesture had low EERs relative to other

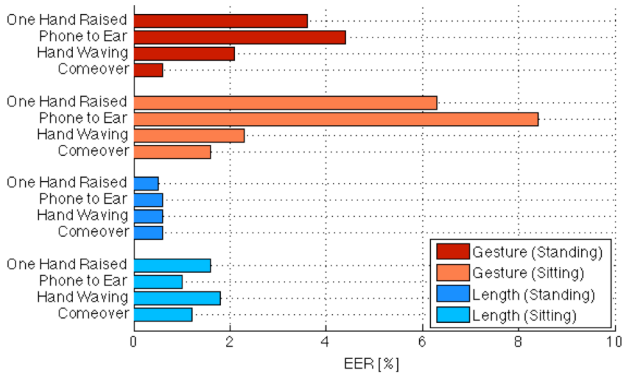

Figure 3. Equal Error Rates (EER) for user identification using one feature vector. Classifiers were evaluated with samples in the same posture as in the first session using 10-fold cross validation. Smaller EERs denote better performance.

gestures. In other words, the hand motions of the come-over gesture varied more between participants relative to other gestures, and thus have greater discriminatory power.

Performance Across Session and Posture

To evaluate robustness of the body segment lengths and gesture feature vector, we tested our system against samples measured in different sessions or postures. The results were similar for all gestures, and so for brevity, we only show the results for the hand waving gesture (Figure 4). We trained samples in the standing condition in the first session. Then, we calculated EERs using the samples of the standing condition in the second session and of the sitting condition in the first session as test sets. We include EERs calculated in Figure 3 as baselines.

Across sessions, the EERs were much higher. Thus, the feature vectors are not very repeatable across sessions. The EERs calculated for body segment length feature vectors and gesture feature vectors increased 20 fold and 5 fold respectively compared to the baselines. This finding indicates that, to further improve the performance, we need to train the classifier on samples acquired from multiple sessions to account for systematic sensor and performance imprecision across sessions.

Furthermore, across postures, the EERs calculated using just the body segment lengths feature vector increased over 80 fold. The EERs calculated from the gesture feature vector for hand waving also increased by a factor of 5. These results suggested that body segment length feature vectors were sensitive to postures and that we needed to train a separate SVM classifier for each posture.

The source of the inconsistencies in the measurement of body segment length remains unclear. We held the environment in which we took measurements constant, so environmental causes were unlikely. Rather, we suspect the inconsistencies across sessions may be caused by clothing changes across sessions. This hypothesis could explain why the samples within sessions are consistent but the samples

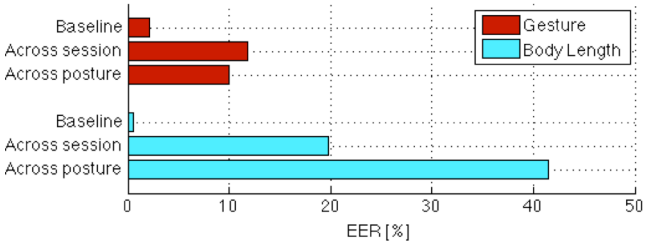

Figure 4. EERs with hand waving across sessions or postures. When classifiers were tested with samples in different sessions or postures, the EERs became higher. Note that the x-axis scale is different from that in Figure 3.

across sessions are inconsistent. However, further study is necessary to confirm the hypothesis.

Evaluation of Body-Based identification

We found that, when used separately, neither the body segment length nor the gesture feature vector was consistent enough to identify users reliably. Next, we evaluated identification performance using both feature vectors.

The previous analysis indicated that we needed to train our classifier with samples collected in difference sessions to achieve better performance. However, in this study, we have only two sessions. Therefore, we decided to train our classifiers only using samples from the first session to separate training and test sets clearly. We evaluate training with multiple sessions in the next study.

Dataset

For all samples, we concatenated body-lengths feature vectors and gesture feature vectors. For each gesture-posture pair, we trained SVM classifiers using samples collected in the first session, and evaluated with samples collected in the second session.

User Identification

Figure 5 shows EERs for each posture-gesture pair. For instance, when the hand waving gesture feature vector was combined with the body segment lengths feature vector, our system obtained 4.3% and 6.2% EERs in the standing and sitting conditions respectively. For other gesture-posture pairs, the EERs range from 2.6% to 7.6%. Thus, across sessions, our system performed much better by combining both body segment lengths and gestures when compared to either one separately as shown in Figure 4. This shows

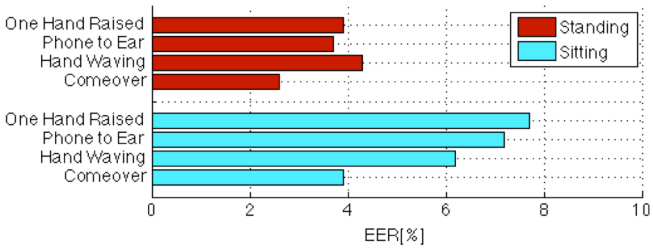

Figure 5. EERs for gesture-posture pairs. The system was trained with samples collected in the first session and tested with samples collected in the second session.

Gesture	Naturalness	Ease of Use	Security
One Hand Raised	4.0	5.0	3.0
Phone to Ear	2.0	4.0	3.0
Hand Waving	4.0	5.0	3.0
Come-over	2.0	4.0	3.0

Table 2. The numbers represent the median of the five-point Likert scale ratings by the participants. Higher numbers denotes positive ratings.

feasibility and effectiveness of the user identification based on both body segment lengths and gesture patterns.

Survey Results

At the end of the second session, we asked participants to complete a post-test survey. We asked participants to rate our body-based identification in terms of naturalness, ease of use, and security on a five-point Likert scale. For instance, with regard to naturalness, we asked how strongly participants agreed or disagreed a statement "using body lengths and hand waving to log into systems is natural" where 5 stands for strongly agree and 1 stands for strongly disagree. We used "logging in" as an example because it was easy to understand for users and it was one of the typical use cases of user identification. For the remainder of this section, we denote the median of the ratings in parentheses. See Table 2 for a summary of results.

The participants rated both the one-hand raised gesture and hand waving as natural (4.0) and very easy to use (5.0). For example, P11 commented, "I feel that the waving gesture is most appropriate for a login process because it has a connotation of welcoming and saying 'hello', which seems to feel right for a login screen." Indeed, the one-hand raised and hand waving gestures are common greetings that initiate human-human interactions. Thus, it makes sense that participants reported that these gestures were a natural form of identification—the common initiation of many human-computer interactions.

Participants rated body-based identification as neutral in terms of security (3.0). One explanation could be that participants were not familiar with the security properties of behavioral biometrics. P15 had concerns in using gestures for identification. She commented, "For the purpose of logging into a system, I would feel more secure if I could select a gesture, or motion, that I feel is somewhat unique to me, […] and therefore less likely to be copyable."

In summary, participants reported that the body-based identification system was natural and easy to use, thought its level of security was unclear. We agree that impersonation is possible by mimicking a user's gestures especially when an attacker has body lengths similar to a target user. However, for environments with reasonable physical security, such as homes, or for non-critical applications, usability may be favored over security. Furthermore, applications requiring more security could still make use of our work, but it should be complemented by additional techniques.

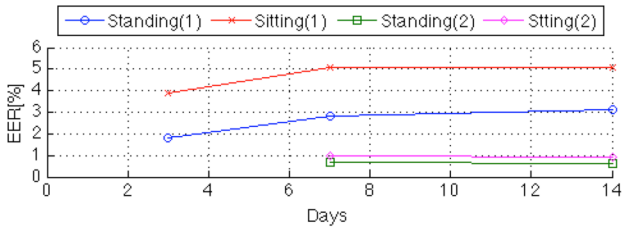

Figure 6. EER for each condition along with the number of days from the first session. The numbers in parentheses denote the numbers of sessions used to train the system. Training the system with samples from two sessions significantly improved performance.

USER STUDY #2: TWO WEEKS EVALUATION

For our second user study, we chose to investigate performance when trained with samples collected in two sessions, as well as longer-term robustness of users' gesture patterns. In this study, we only tested the hand-waving gesture because it was the most promising—it had low EERs and was well liked by users. Gestures were measured in the same manner as the first study (in two postures and 10 samples for each posture in each session). We conducted the measurements in four sessions: on the first day, three days later, one week later, and two weeks later.

We recruited 27 participants using an existing university recruiting website meant for the general public. Out of the 27 participants, 20 participants were male and 7 were female. Their ages ranged from 19 to 62 years old with a mean of 29.7. Sixteen participants were university students, four were university staff employees and seven were otherwise employed. Their average height was 173cm (SD=9.8) and their average weight was 75.1kg (SD=21.1). None of them participated the previous study. We paid $50 for completion of the study.

Performance over a Two-Week Period

Figure 6 shows how the EERs changed over time. In Figure 6, Standing (1) shows the EERs in the standing condition when samples obtained in the first session were used as a training set. According to the figure, the EERs were constant around 3% one week later and two weeks later. Similarly, in the sitting condition (i.e., Sitting (1)), the EERs were around 5.1%. The results were consistent with the results from our first study, indicating that the performance of body-based identification was stable for two weeks.

Training Classifiers with Samples from Two Sessions

The results in our first study indicated that training a SVM classifier with samples collected in different sessions would improve the performance. Standing (2) and Sitting (2) in Figure 6 show EERs obtained when the samples in the first and second session were used as a training set, and the samples in the third and fourth sessions as test sets. As we expected, the performance was improved significantly. In the standing and sitting conditions, the EERs were 0.7% and 0.9%. Furthermore, the EERs were constant in the two-weeks period. These results suggest that samples from two different dates should be enough to represent variations of

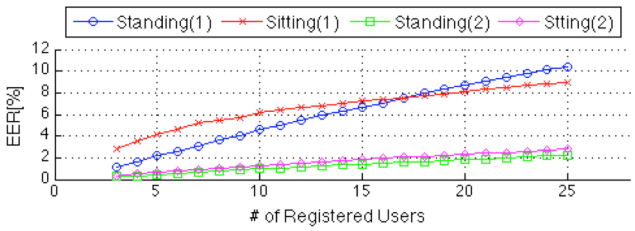

Figure 7. EER for each condition along with the number of registered users. The numbers in parentheses denote the numbers of sessions used for training. When the number of registered users increased, EERs increased almost proportionally

features, and that training SVM classifiers with samples from two dates addresses the variation in body segment length measurements and gestures effectively.

Effect of Numbers of Registered Users

As stated earlier, in this paper, we calculate EERs based on the assumption that the system has seven registered users. However, in practice, the system may have different numbers of registered users. Thus, we investigated how the number of registered users affected EERs. For each N (N=3, 4, …, 25), we randomly chose 10,000 combinations of N participants, and calculated the EERs among the combinations. The classifiers were trained using either the samples from the first session or the samples from both the first and the second sessions as training sets. We used the samples obtained in the fourth session as a test set.

Figure 7 shows the EERs for different numbers of registered users for each condition. Generally, the EERs increase proportionally to the number of registered users. When the system was trained with samples from the first session, the EERs were 10.4% and 9.0% for 25 registered users in the standing and the sitting conditions respectively. Likewise, when trained with samples from the first and the second session, the EERs were 2.3% and 2.8% respectively (N=25). These results are very encouraging. Considering that our system requires only performing a waving gesture for two seconds to identify a user among 25 registered users, the EERs less than 3% were surprisingly low.

Robustness to Partial-Body Occlusions

In our studies, participants stood or sat three meters away from the Kinect to allow our system to capture the participant's entire body. However, in practice, capturing entire bodies may be difficult—for example, if a user is too close to the Kinect or if there are objects in the room obstructing part of the camera's view.

To estimate the impact of partial-body occlusions on performance, we ran simulations with body segment length feature vectors that only included segment lengths above the spine—using joints from shoulders to hands and heads.

Figure 8 shows the EERs with simulated occlusions along with EERs without occlusions. Classifiers were trained either using samples from the first session (Standing (1) and Sitting (1)) or using samples from the first and second sessions (Standing (2) and Sitting (2)) as training sets. For both cases, samples from the fourth session were used as a test set. Figure 6 shows that the estimated effects of the partial-body occlusions are limited. On average, the EERs increase by just by 0.96% (σ=0.40%). In other words, our body-based identification system worked reliably even if lower body segments were occluded.

USER STUDY #3: FIELD STUDY

In our two lab studies, we demonstrated that the body-based identification performed well in identifying users. However, the lab studies were conducted in a well-controlled procedure (e.g., using a window-less room and controlling a distance between participants and a Kinect). To evaluate the body-based identification system in more ecologically valid situations, we conducted a field study.

We recruited five households (12 participants) by posting flyers in a local neighborhood. Out of the 12 participants, five were male and seven were female. Their age ranged from 18 to 42 years old with a mean of 33 years old. Seven participants were employed, two were students and three were unemployed. Their average height was 159.5cm (SD=11.4) and their average weight was 56.9kg (SD=6.9). None of them participated the previous studies.

This study consisted of three sessions at their homes. The three sessions took place on different days in a week depending on participants' availabilities. The gaps between sessions were not controlled because the results of the study #2 indicated that days from registration had limited impact on the performance. However, successive sessions had at least one-day gap between them to have some distractions between sessions.

For each session, we collected data in their living rooms. We put a Microsoft Kinect on their TVs and asked participants to perform waving gestures in front of the TVs as if they were interacting with the TVs. All living rooms had windows to outside, and room lights were kept as the participants configured prior to our visits. This gave us a variety of infrared and visible light conditions.

We collected 10 hand-waving gestures in two postures, just as in our second study. In the standing posture, participants were asked to stand in positions where they felt confortable,

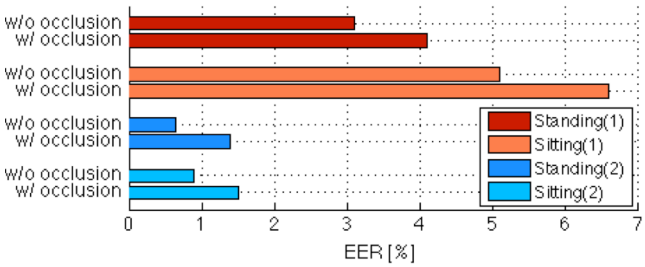

Figure 8. We simulated occlusions by using body part locations only above spines. On average, the EERs became only 0.74% higher.

assuming that they would be interacting with the TVs using gestures. All participants stood two to three meters away from TVs facing to them. In the sitting posture, we asked participants to sit as they usually did in their living room. Eight participants sat on couches, two sat on chairs, and two sat on cushions placed on the floor. After the third session, the participants were asked to fill a post-survey followed by brief interviews based on their responses to the survey. The first and the second session took about 15 minutes, and the third session took about 45 minutes to complete. We paid $40 for completion of the study.

As we anticipated, we observed partial-body occlusions in the subset of samples. The occlusion occurred in the samples from six participants in the sitting posture. For those samples, tables placed between TVs and participants occluded participants' legs. We used classifiers trained with body segment lengths above the spine, as we did in the simulation of partial-body occlusion, to perform user identification with these samples.

In the less controlled study design described above, we collected samples with natural variations in participants' postures, orientations and distances between a Kinect in addition to light condition changes and some occlusions.

User Identification

We trained classifiers for each posture either using samples from the first session, or from the first and the second sessions. Each classifier was tested with samples collected in the third session (Figure 9). When the first session was used as a training set, the EERs of the standing and sitting posture were 3.4% and 5.3% respectively. The EER values were close to the result in the second study (3.8% and 5.1%). When we trained the SVM classifiers with the first and second sessions, the EERs were 0.5% and 1.6% for the standing and the sitting posture respectively, which were also comparable to results (0.9% and 1.0%) in the study #2.

During the study, we noticed that the participants' sitting posture varied more than we observed in our lab studies. We suspect that the participants in this study were more relaxed than ones in the lab study because this study was conducted in their living rooms. However, the EERs remained almost same as ones from the lab studies. This finding offers additional support that we can compensate the variations in the features by training the system with samples obtained in two different sessions. Additionally, the results obtained in this field study show that the body-based identification system also works well in ecologically valid situations with partial body occlusions and environmental changes.

Survey and Interview Results

In the survey, we asked questions about how well body-based identification could be used in practice. First, we asked how many people they would want to register to this system, and who those people would be. They answered that, on average, they wanted to register 5.2 users (SD=1.1).

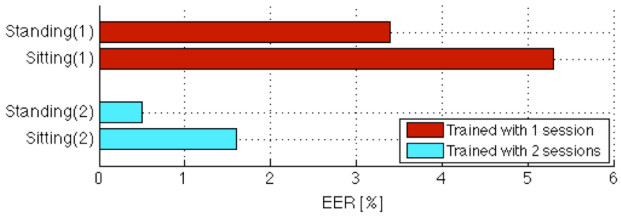

Figure 9. EERs obtained in a field study. Overall, the EERs remained similar to the results in the study #2 showing robustness of the system to environmental changes.

Most of the users were people they were living with (e.g., family members and roommates). Four participants mentioned that they wanted to register one or two friends who visited their home frequently. These responses validate our assumption of having roughly seven registered users, as assumption we made in calculating EERs.

We then asked participants to rate the statement "I will use body-based identification system if it is available on home electric appliances such as TVs" using a five-point Likert scale (1=strongly disagree and 5=strongly agree). Participants responded they agreed with the sentence with a median of 4.0. Although we have to be careful of possible novelty effect, these ratings are encouraging.

We further asked participants for what purposes they wanted to use the body-based identification. Five participants mentioned parental controls. P1 said, "I want control what my child can do with electric appliances such as TVs. They quickly learn how to use remotes. So, it's difficult to stop them using a TV. But, if the TV checks [users'] heights, it will be easy." Two participants mentioned personalization as a potential use case. P5 commented, "If a TV automatically turns on and shows a TV program that I regularly watch when I wave to it, that would be great. I don't need to search a remote." P7 noted, "If I have a wall display, and I can check personalized information such as today's schedule, subjects of received emails, local weather by waving to it in a busy morning, that would be useful." Although these responses were speculative, they indicate that participants saw value for body-based identification in practical situations.

We also asked participants to rate the statement, "It is acceptable that the body-based identification system requires two sessions of data (20 waving gestures for one posture) to register a user" using a 5-point Likert scale. Participants generally agreed with the sentence with a median of 4. Two participants chose strongly agree, eight chose agree, and two chose disagree. P2 commented, "Registration happens only once. So, I don't care. Once I'm registered, it's pretty quick." On the other hand, P9 said, "Doing 10 waving [gestures] would be OK. But, it's troublesome if I need to do more than that." These responses are encouraging because the current un-optimized training process was acceptable for the majority of participants. However, further investigation would be necessary to shorten the process.

LIMITATIONS

While our results are encouraging, there are a few caveats to our findings that represent fruitful opportunities for future work. In our user studies, we did not provide feedback about user identification results to avoid biasing users. However, for real deployments, the system should provide feedback, which would affect how users perform gestures. Some users may consciously reproduce their individual gesture patterns; moreover, some may game the system by registering distinct waving patterns to be easily identified. Investigation of these user behaviors could be interesting follow-up research.

Furthermore, in the second study, two weeks may not be long enough to evaluate the long-term robustness of the system. In the third study, we collected data at participants' living rooms; however, it was not a field deployment (e.g., we still have sessions). As a result, participants' responses to our survey were speculative. A longitudinal deployment of a body-based identification will provide further insight into these issues.

Finally, we do not claim that the features we use and the models we employ are optimal. Our system performed reasonably well using a simple model and feature set, but it would be possible to further optimize performance using other interesting features and models.

DISCUSSION

While our body-based identification was reasonably accurate, there is still a room for improvement. One opportunity is to improve the precision of the body segment length measurements. We found that, within the same session, body segment lengths were reliable, but were unreliable across sessions. We suspected that this would be primarily because participants' clothing changed between sessions. Because we did not control participants' clothing, we think our results reflect variations caused by different clothing. Moreover, in the second study, there were two participants wearing baggy clothing; however, it did not affect our system's performance. While these indicate that real-time joint extraction in the Kinect SDK is reasonably robust, more computationally expensive processing could yields better results in user identification.

Additionally, we found that the performance of body-based identification depends on consistency in postures. However, in practice, it would be ideal for body-based identification to be reliable regardless of the posture variations. Our results offer some hints in this vein. Specifically, we found that removing some body segment lengths from the body segment length feature vector did not significantly affect performance. Consequently, we could omit body segment lengths that have imprecise readings in different postures to stabilize performance across postures.

Similarly, users' orientations to a Kinect could affect performance. A Kinect extracts the three-dimensional position of users' joints. Thus, theoretically, body-based identification should be robust to changes in relative orientations between a user and a Kinect. Moreover, in the use cases where a Kinect is placed near the device that users interact with (e.g., in our third study), users naturally face to the Kinect. Thus, the orientations would be less of an issue. In contrast, in the use cases where we put a Kinect in a room and users interact with devices placed at many different positions in the room, users are unlikely to face to the Kinect. As a result, users' bodies could occlude their gestures and/or body lengths measurements could become less robust. Further investigation would be necessary for these use cases.

Furthermore, the registration process could be improved. In our analyses, we trained the system using 20 samples collected on two different days for each standing and sitting postures. Although, in our field study, majority of participants reported that it was acceptable, requiring data from two sessions could pose challenges in practice.

It is also possible to improve the user identification algorithm. One possible approach is to build models that represent how the feature values measured for the same user could vary. We did build such models and found that they did not perform well (not reported in this paper due to space constraints). However, we still believe that there are opportunities in this direction. Another approach would be to build a layered classifier combining other factors amenable to natural user interfaces, such as face recognition, to improve the overall performance of user identification. An upper layer classifier can use predictions from lower layer classifiers, such as body-based identification, face recognition and voice recognition to perform combined user identification. As we demonstrated, using multiple factors has great potential in improving overall performance in user identification.

One interesting implication for gesture-based interfaces would be the usefulness of recognizing both gestures and users' identities. Currently, gesture-based interface systems recognize what gesture users are performing. Then, the systems execute the same commands regardless of users performing the gestures (e.g., when a TV recognizes a user performing hand waving, the TV turns on). However, if a system can recognize both gestures *and* users' identifies, the system can personalize execution of the commands (e.g., turning on a TV *and* showing program that a specific user prefers). The first lab study suggested that users also could be identified when performing other gestures. This could be an interesting direction for future research.

CONCLUSION

We proposed body-based user identification: a new form of user identification based on the body segment lengths and a hand-waving gesture. We evaluate the system through two lab studies and one field study consisting of 8,640 samples from 75 users in total. Our first lab study explored the basic feasibility of the system, in terms of properties of body segment length and four different gestures. Based on the

results, we chose a hand waving gesture as the most promising gesture to be used in subsequent studies.

Our second lab study showed the robustness of our system over a two-week period, as well as robustness to partial body occlusions. Our study also showed that training the system using samples collected on two different days effectively addresses the variations in the body segment length measurements and hand-waving measurements. The EERs improved from 3.0% to 0.7% when trained with samples collected on two different days compared to when trained with samples collected on a day.

Finally, our field study demonstrated that our proposed system worked well in ecologically valid situations, in terms of different environments and partial occlusions. Our system achieved an equal error rate ranging from 0.5 to 1.6% when trained with samples from two sessions.

Given our encouraging results, we believe that this paper makes a first step toward the vision of more natural and usable user identification seamlessly integrated with gesture-based interfaces.

REFERENCES

1. BioLab University of Bologna. Finger Print Verification Competition. Retrieved Sep. 13, 2013. https://biolab.csr.unibo.it/fvcongoing/UI/Form/IJCB2011.aspx
2. Douglas Gantenbein. Helping Kinect Recognize Faces. Retrieved Sep. 13, 2013. http://research.microsoft.com/en-us/news/features/kinectfacereco-103111.aspx
3. Microsoft. Kinect for Windows SDK. Retrieved Sep. 13, 2013. http://msdn.microsoft.com/enus/library/hh855347
4. Corlett N.E., and Clark S.T. 1995. The Ergonomics of Wrokspaces and Machines: A Design Manual. CRC Press. 1995.
5. Gafurov, D. A Survey of Biometric Gait Recognition: Approaches, Security and Challenges, *NIK-2007 Conference*. 2007.
6. Guillaume,H., and Sébastien M. A novel statistical generative model dedicated to face recognition. Image and Vision Computing. vol. 28, Issue 1, 2010, 101-110.
7. Jain, A., Nandakumar, K., and Ross, A. Score normalization in multimodal biometric systems, Pattern Recognition, vol. 38, Issue 12, 2005, 2270-2285.
8. Jain, A., Bolle, M.R., and Pankanti, S. Biometrics: Personal Identification in Networked Society. Springer. 2005.
9. Jiayang, L., Lin, Z., Jehan, W., and Venu V. User evaluation of lightweight user authentication with a single tri-axis accelerometer. In *Proc. of MobileHCI*. 2009.
10. Kam L., Konrad, J., Ishwar, P., Towards Gesture-Based User Authentication, *Advanced Video and Signal-Based Surveillance,* 2012, 282-287.
11. Leyvand, T., Meekhof, C., Wei, Y., Sun, J. and Guo, B. Kinect identity: Technology and experience. IEEE Computer Society. Computer Magaznine, April, 2011, 94–96.
12. Cranor L. F. and Garfinkel S. Security and Usability Designing Secure System That People Can Use. O'Reilly Media. 2005.
13. Mäntyjärvi, J., Lindholm, M., Vildjiounaite, E., Makela, S.M., and Ailisto, H. Identifying users of portable devices from gait pattern with accelerometers. In *Proc. of ICASSP.* vol. 2, 2005, 973-976.
14. Maltoni, D., Maio, D., Jain, K.A., and Prabhakar, S. 2005. Handbook of Fingerprint Recognition. Springer.
15. Maurer, M. E., Waxenberger, R., and Hausen, D. BroAuth: evaluating different levels of visual feedback for 3D gesture-based authentication. In *Proc. of the International Working Conference on Advanced Visual Interfaces.* 2012, 737-740.
16. Mayrhofer, R. and Gellersen, H. Shake well before use: two implementations for implicit context authentication. *IEEE Transactions on Mobile Computing*, vol.8, no.6, 2009, 792-806.
17. Napa, S.B., Kowsar, A., Katherine, I., and Nasir, M. Biometric-rich gestures: a novel approach to authentication on multi-touch devices. In *Proc. of SIGCHI*. 2012, 977-986.
18. Patel, S.N., Pierce, J.S., and Abowd, G.D. A gesture-based authentication scheme for untrusted public terminals. In *Proc. of UIST* , 2004, 157–160.
19. Peacock, A., Ke, X., and Wilkerson, M. Typing patterns: A key to user identification. IEEE Security and Privacy, vol. 2, Issue 5, 2004, 40-47.
20. Pheasant, S., and Haslegrave, M.C. Bodyspace: Authropometry, Ergonomics and the Design of Work. CRC Press. 2005.
21. Schmidt, D., Chong, M.K., and Gellersen, H. HandsDown: hand-contour-based user identification for interactive surfaces. In *Proc. of NordiCHI*, 2010, 432–441.
22. Snelick, R., Uludag, U., Mink, A., Indovina, M., and Jain, A. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.27, no.3, 2005, 450-455.
23. Tian, J., Qu, C., Xu, W., and Wang, S. KinWrite: Handwriting-Based Authentication Using Kinect. In *Proc. of* Annual Network & Distributed System Security Symposium, 2013.
24. Wan, V., and Renals, S. Speaker verification using sequence discriminant support vector machines. *IEEE Transactions on Speech and Audio Processing.* vol.13, no.2, 2005, 203- 210