AirAuth: Evaluating In-Air Hand Gestures for Authentication

1st Author Name
Affiliation
Address
e-mail address
Optional phone number

2nd Author Name
Affiliation
Address
e-mail address
Optional phone number

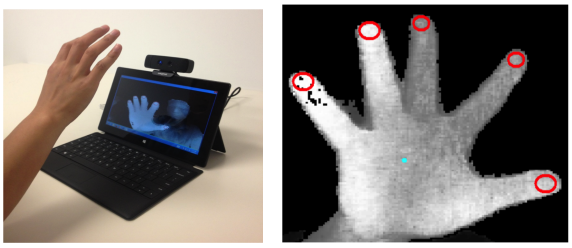
3rd Author Name
Affiliation
Address
e-mail address
Optional phone number

ABSTRACT
Secure authentication with devices or services that store sensitive and personal information is highly important. However, traditional password and pin-based authentication methods compromise between the level of security and user experience. *AirAuth* is a biometric authentication technique that uses in-air gesture input to authenticate users. We evaluated our technique on a predefined (*simple*) gesture set and our classifier achieved an average accuracy of 96.6% in an equal error rate (EER-)based study. We obtained an accuracy of 100% when exclusively using personal (*complex*) user gestures. In a further user study, we found that AirAuth is highly resilient to video-based shoulder surfing attacks, with a measured false acceptance rate of just 2.2%. Furthermore, a longitudinal study demonstrates AirAuth's repeatability and accuracy over time. AirAuth is relatively simple, robust and requires only a low amount of computational power and is hence deployable on embedded or mobile hardware. Unlike traditional authentication methods, our system's security is positively aligned with user-rated pleasure and excitement levels. In addition, AirAuth attained acceptability ratings in personal, office, and public spaces that are comparable to an existing stroke-based on-screen authentication technique. Based on the results presented in this paper, we believe that AirAuth shows great promise as a novel, secure, ubiquitous, and highly usable authentication method.

Author Keywords
In-air gestures; authentication; shoulder surfing; user experience; acceptabilty.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
A growing number of users store and manipulate important and sensitive information online, on their personal computers and mobile devices. As such, finding methods of secure and easy-to-use authentication is of increasing importance, since tradeoffs exist between the users' desire for security and the compromises in user experience they are willing to take [10].



(a) (b)

Figure 1. (a) A user is performing an authentication gesture in front of a short range depth camera. (b) Hand features (3D locations of finger tips and hand center) captured and used by AirAuth.

At present, passwords and PIN numbers are the most widely-used authentication methods for gaining access to PCs, mobile devices and online accounts, and they are well-understood by the users. However, such knowledge-based systems can have disadvantages, such as requiring the user to learn complex passwords (for increased security vs. "trivial" passwords). Also, as the number of accounts or devices the user needs to access grows, the mental burden increases for the user to remember multiple passwords or variants thereof. That is one reason why a number of users resort to using only a few passwords (perhaps with simple variants) used for all of their authentication activity [27], which puts the users at significant risk if one of these passwords is compromised.

Traditional password entry can be prone to *shoulder-surfing attacks* [25]. The growing amount of video surveillance in public spaces and the potential of misusing it compounds this danger. In addition to shoulder-surfing attacks, mobile devices equipped with touch screens are prone to *smudge attacks* [2]. In the case of smudge attacks, the attacker attempts to reconstruct the device pin or login stroke from finger smudges left on the device's screen after authentication.

A possible solution to problems of traditional password entry is to use biometric features during the authentication process. Instead of relying solely on the user's knowledge of a secret, biometric authentication systems can enhance security twofold: (1) the systems can directly use distinct physical features of the legitimate user; (2) the system can make use of behavioral traits of the legitimate user during the authentication process as a further authentication layer.

Regarding the user experience, this means that users are permitted to retain less complex secrets, since the biometric adds extra layers of security. Furthermore, the authentication process itself can be made more engaging. We argue that instead of using a keyboard or pinpad for password entry, gestures can be used for authentication, where the system checks for the knowledge of a secret as well as physical and behavioral properties of the user. In daily life, some gestures, such as writing one's signature, are performed very often and can thus be easier and more engaging to use than traditional password or PIN-based authentication.

In this paper, we present a novel authentication system, *AirAuth*, that uses hand gestures made in the vicinity of a computing device (Figure 1(a)). Using a short-range depth camera, our system tracks hand gestures input by the user. In contrast to longer-range systems such as Kinect, the short-range depth camera allows us to track individual finger tips and the center of the user's hand in 3D space (Figure 1(b)). This data provides us with an abundance of features allowing the decoding the user's authentication secret, classifying biometric properties of the user's hand, and classifying the movement style of the user.

Our system achieves a higher authentication accuracy than previous work [12, 22] that does not track multiple points in space. Furthermore, we do not require a large multitouch screen [22] for gestural authentication to work. Also, the mobile device itself does not need to be moved [12], which might prove impractical in some situations.

We present the results of three user evaluations we conducted of our system. In the first study, we analyze the EER-based authentication accuracy of our system and achieved an accuracy of 96.6% including predefined gestures and 100% for personalized gestures only. In a second study, we analyze the susceptibility of our system to shoulder-surfing attacks, by asking a second and distinct set of users attack user models trained in the first study. The results of this study confirm those of previous work [23], demonstrating the resistance of gesture-based approaches to shoulder surfing attacks. A third longitudinal study was conducted to analyze the robustness of AirAuth over time and the effect of a template update strategy on the authentication accuracy. We also evaluated our system in terms of user experience and location acceptability. The results indicate that gestures that are enjoyable to the users are also more secure and difficult to forge, an important characteristic which is not present in traditional password based authentication systems. In addition, our technique was also highly rated by users for use in personal, office, and public spaces when compared to the presently available stroke-based on-screen authentication technique used in Android phones. Finally, our technique is simple, robust, and easy to implement. We used the Dynamic Time Warping (DTW) algorithm [24] which is fast and easily deployable on embedded hardware. This characteristic makes our system highly ubiquitous; can be used anywhere where secured authentication is crucial, e.g., on door panels, fixed PCs, laptops or mobile devices.

RELATED WORK

We have structured our discussion of related work as follows: first, we discuss *usable security* in general and look at previous works on password-based authentication schemes, to highlight how our proposed method could solve some of the problems of those schemes. Next, we discuss some of the most common *threat models* that are presented in the literature. Following that, we look at *biometric methods* used in previous works. Finally, we discuss previous work done on *gesture-based authentication*.

Usable Security and Password-Based Authentication

Kainda et al. [10] highlighted the tradeoff between usability and security. Security engineers always strive to make systems more secure, but it has been shown that usability suffers as a consequence. New authentication schemes should thus strive to make authentication more usable and engaging, while keeping the authentication process secure.

The vast majority of today's online services, computers and mobile devices use a password or PIN-based scheme for ad-hoc (immediate) user authentication. Although this type of authentication is well-known by both the user base and also developers of secure systems, it does have some drawbacks. Password-based authentication systems become more secure as the length of the password grows. However, longer passwords are more difficult to remember, and, instead of using a distinct password for different services that require authentication, users commonly use variations of a basic password. Also, users tend to change their password infrequently, once it has been set [20]. Von Zezschwitz et al. report that users use simpler password variations for low-sensitivity authentications. If these low-sensitivity accounts are compromised by attackers, highly sensitive authentications that share the same basic password root are put at risk [27].

Threat Models

As already mentioned in the introduction, *shoulder-surfing* and *smudge attacks* are two very important threats to password-based authentication on mobile devices.

There have been numerous approaches proposed to counter the shoulder-surfing threat, for instance, using fake cursors to confuse attackers [5], making use of relative locations of known password icons and distractor icons [29], using set-intersection techniques to extract a correct pin using binary choices made by the user [21], entering the password out of sight from potential attackers [4] or using haptically-cued false entries [3]. We verified empirically that our proposed method is resilient against shoulder-surfing attacks through a user study with video-based attacks.

A further threat model, especially for touch-based mobile device and perhaps to a somewhat lesser degree on keyboards or keypads is the *smudge attack*, which lets attackers reconstruct password, PIN or stroke entry by imaging or otherwise analyzing the oily residue left by the user's fingers on a pre-cleaned surface after a legitimate user has performed an authentication [2]. On touchscreen devices, smudge attacks can be mitigated by randomizing the locations where user needs to touch the screen for authentication [28]. AirAuth is fully

resistant to smudge attacks, since no touch-based interaction whatsoever is required for authentication.

Biometric Techniques

A wide range of input devices and sensors have been used to gather biometric data for authentication. Jorgensen et al. explore mouse dynamics as a biometric for behavioral authentication [9]. Similarly, keystroke dynamics have also been explored [16]. Alpcan et al. propose signature-like strokes on a touchpad as a biometric [1]. This concept has been extended to multi-touch surfaces with good results [22]. Kumar et al. propose gaze-based password entry [11] as method of avoiding shoulder-surfing attacks. Similarly, [14] propose using specific gaze points on still images for authentication. Further biometrics that have been used include fingerprint scanning, retinal scanning, face recognition, voice recognition, and palm vein scanning [13, 30]. However, these techniques are complex and require specialized hardware. In contrast, AirAuth is relatively simple and requires only a depth camera, hardware which may soon be integrated into smartphones [7]. In contrast to pure biometrics, AirAuth also has the additional advantage of requiring the knowledge of a gesture for successful authentication.

Gesture-based Authentication

Motion gestures on mobile phones have been proposed as an authentication biometric. During a gesture input, the user moves the entire phone through the air. The movement is recorded by the mobile phone's motion sensors (typically an accelerometer and a gyroscope). The results presented so far in the literature appear promising [8, 12, 17]. A relatively closely related touch-less biometric has been proposed by Sahami et al. [23], which uses a mobile device's magnetometer to track free-space input via a magnetic widget. The main drawback of their method is that only a single point is tracked in space and the requirement for the magnetic widget. AirAuth tracks multiple points in free space without the requirement for additional user instrumentation. [22] explored multi-touch gestures on on touch screens with positive results. However, that technique may not be suitable for phone-sized devices due to the space required for 5 finger gestures and it also may be susceptible to smudge attacks. Finally, *BroAuth* [15] used full-body gestures tracked by a Kinect for authentication. We believe that using hand gestures as in AirAuth is more practical, because it requires much less interaction space and it also supports small devices such as handsets and tablets.

THE AIRAUTH SYSTEM

In the following, we describe the hardware and software implementation of AirAuth prototype, which we also used during our user studies.

Hardware

Our hardware prototype (Figure 2) consists of a Creative Senz3D[1] short range depth sensor placed on a computing device, in our case a Microsoft Surface Pro Tablet (Figure 1(a)), in such a way that the sensor can image the user's hand while she makes authentication gestures in the vicinity of the device. When the user's hand passes a previously set threshold distance from the device, gesture recording starts. The recording stops when the user's hand moves beyond the threshold distance.

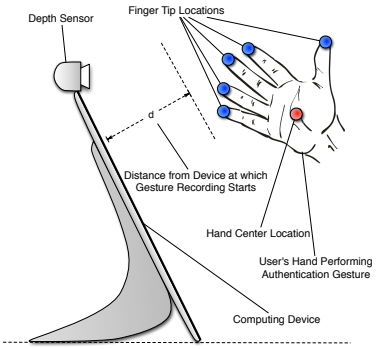


Figure 2. Hardware prototype of our system. A depth sensor is arranged on a computing device so that it can observe the user's hand while it makes in-air gestures for authentication. The depth sensor measures the 3D locations of the user's fingertips and of the hand center.

From the depth image, persistent features can be extracted, i.e., the 3D locations of the users fingertips. The extracted features (tip and hand center locations) are a biometric that our system uses for user authentication. However, our system is not limited to these specific features. Other types of persistent features can also be used given a means to extract them reliably. Since a gesture entry typically takes around 1–2 s to perform, our biometric consists of a time-based stream of the aforementioned features. For our prototype, we used the Intel Perceptual Computing SDK[2] to extract the feature points from the depth image.

Software

In a biometric authentication system, there are usually two operational stages: the *enrollment stage* and the *authentication stage*. In the enrollment stage, the users authentication data is registered with the system by acquiring and storing a biometric template corresponding to a particular user. In the authentication stage, instances of authentication inputs are compared with the registered one(s) in order to authenticate a user.

Figure 3 shows an overview of our software implementation. A new user needs to undergo an enrollment phase (a), where the user performs three authentication gestures. If the gestures are not sufficiently similar to each other, the user has to repeat the task. Once the user passes the enrollment phase, those three gestures are processed and their average is stored by the system as the template. The maximum pairwise distance between the three initial gestures is calculated and stored as a threshold ε. For distance calculation, we used the Dynamic Time Warping (DTW) algorithm [24].

[1]The maximum IR depth resolution of this sensor is 320x240 and its diagonal field of view is 73°. The sensor captures images at a rate of 30 Hz.

[2]`http://software.intel.com/en-us/vcsource/tools/perceptual-computing-sdk`

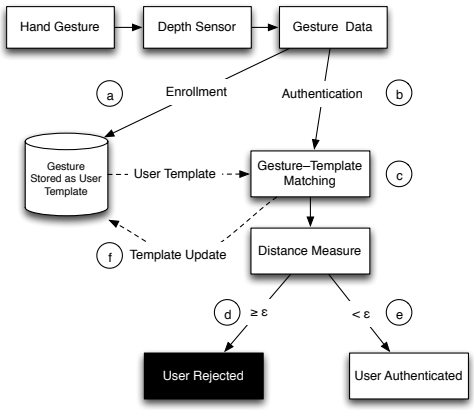


Figure 3. A conceptual overview of the AirAuth software implementation. Note the two main stages of the authentication system—enrollment (a) and authentication (b).

For the authentication phase (b), our system compares the gesture data input by the user with the template and calculates the distance score (c). If the distance score is greater than the threshold ε, the user is rejected (d), since this indicates a fraudulent authentication attempt. If the distance score is below ε, the user is authenticated (e) by our system. To compensate for temporal changes in the biometric, the user template is periodically updated (f) with data from authentication gestures with a low distance to the existing template.

Data Processing

From the depth sensor, we obtain raw 3D coordinates (x, y, and z) of all (visible) fingers of the user and also the user's hand center. However, the location of user's hand with respect to the depth sensor can be anywhere and in any orientation in the air and differ one instance to another. Hence location and orientation of the hand need to be normalized before applying any matching algorithm. Therefore, at first we normalize the data to a $[0, 1]$ interval by scaling the raw 3D points using the following three equations [26].

$$x_1{}^{i^s} = \frac{x_1{}^i - \min_i(x_1{}^i)}{\max_i(x_1{}^i) - \min_i(x_1{}^i)} \tag{1}$$

$$y_1{}^{i^s} = \frac{y_1{}^i - \min_i(y_1{}^i)}{\max_i(y_1{}^i) - \min_i(y_1{}^i)} \tag{2}$$

$$z_1{}^{i^s} = \frac{z_1{}^i - \min_i(z_1{}^i)}{\max_i(z_1{}^i) - \min_i(z_1{}^i)} \tag{3}$$

Here $x_1{}^i$, $y_1{}^i$, and $z_1{}^i$ are raw coordinates of any single finger or the hand center and $x_1{}^{i^s}$, $y_1{}^{i^s}$, and $z_1{}^{i^s}$ are coordinates after performing scaling operation. After scaling, we translate all the points with respect to the origin by subtracting the mean of the point set from all points:

$$(x_1{}^{i^s})_t = x_1{}^{i^s} - \frac{\sum_{i=1}^{n} x_1{}^{i^s}}{n} \tag{4}$$

$$(y_1{}^{i^s})_t = y_1{}^{i^s} - \frac{\sum_{i=1}^{n} y_1{}^{i^s}}{n} \tag{5}$$

$$(z_1{}^{i^s})_t = z_1{}^{i^s} - \frac{\sum_{i=1}^{n} z_1{}^{i^s}}{n} \tag{6}$$

Here $(x_1{}^{i^s})_t$, $(y_1{}^{i^s})_t$, and $(z_1{}^{i^s})_t$ are scaled coordinates after being transferred to the origin.

Matching Gesture Data

While capturing the gesture, the depth sensor collects 3D coordinates of all five fingers and the hand center in each frame. Therefore, data for each input gesture is actually a time-series data of hand locations. So, at first an input gesture is formatted as a time series data of hand locations

$$\text{Input}(t) = [\text{Location}_1, \text{Location}_2, \ldots, \text{Location}_n] \tag{7}$$

where n is the number of frames captured during the entire gesture. Location_i is the ith sequence vector composed of the x,y and z coordinates of the positions of the user's five fingers and his hand center, thus:

$$\text{Location}_i = [x_1^i, y_1^i, z_1^i, x_2^i, y_2^i, z_2^i, \ldots, x_6^i, y_6^i, z_6^i]^\intercal \tag{8}$$

After processing these raw data points using the scaling and translation equations, DTW is used to calculate the distance between two gesture instances. We used the Euclidian distance as the distance measure for DTW.

USER STUDIES

We conducted four user studies to evaluate AirAuth. First, to find a set of simple gestures users would be willing to use, we conducted a pilot study to obtain a set of gesture suggestions from several test subjects. Second, we conducted a user study that comprised enrollment and authentication phases, with the goal of determining AirAuth's accuracy. Third, we conducted a user study to evaluate the resilience of AirAuth towards shoulder surfing attacks. Finally, we performed a longitudinal study to assess authentication similarity and repeatability over time.

User Study 1: Finding a Set of Simple Gestures

One of the goals of our research was to test the accuracy as well as the vulnerability of predefined, *simple*, in-air gestures. Hence, we tried to identify a set of gestures that are simple and easy to reproduce from a user point of view.

Therefore, as a pilot study, we conducted a short interview session with 10 participants from an industrial research lab. At first we explained to them the purpose of the study and asked them to perform up to three different air gestures that they thought were simple and easy to reproduce. Participants could choose any gesture they wanted. We asked them to find gestures that can be performed with both single and multiple fingers. The whole interview session took only 2–3 minutes per participant. As a result of this study, we found a set of 15 gestures. Figure 4 shows the 2D sketches of the gesture suggestions we obtained from the study.

It should be noted that this gesture set is intended to be simple and reproducible from a user's perspective and also performable with one or multiple fingers. If we examine the gestures closely, we can see that some participants chose their first name initial (gesture IDs 6, 12, 14, and 16), favorite number (gesture IDs 1, 5, and 11), shape, symbol, etc. Participants

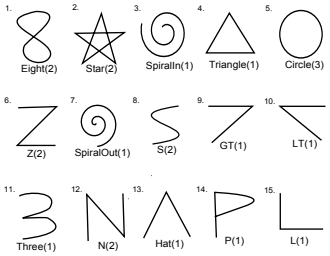


Figure 4. 2D sketches of the user defined *simple* gesture set, consisting of 15 gestures we sourced from 10 participants. The gesture id is denoted on the top left of each gesture symbol. The gesture name and the choice frequency by the study participants are indicated below the gesture symbol.

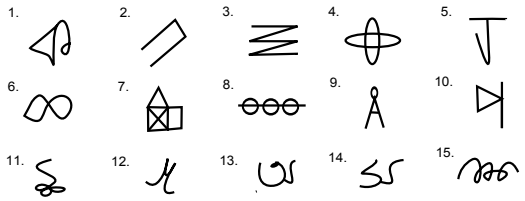


Figure 5. 2D representations of the *complex* gestures the participants of study the 2nd user study chose to enter.

chose certain gestures because, according to them, they were easy for them to perform and also to remember later without any effort. It should also be noted that some of the gestures were chosen more than once by different participants (the choice frequency is indicated after each gesture name in parentheses in Figure 4). This shows that there was at least some consensus regarding the choice of gesture amongst the participants. Some people also chose the existing 2D gesture they use to unlock their touch-based mobile phone since they were already comfortable with the 2D version of the gesture.

User Study 2: Determining AirAuth's Accuracy

In this study, we investigated the accuracy of our system across multiple users. For this study, we recruited 15 participants. 11 were male and 14 of them were right handed. All but one of them was fairly experienced with smartphones and computers. 4 of them were familiar with in-air gestures (mostly using Kinect). The participants' age ranged from 18 to 46 (average: 28.3, SD: 6.5).

Before each user study session, we explained our system briefly to the participants. After that, they were asked to fill out a short questionnaire with demographic information and then proceeded to the gesture trials. To test our system with more complex and personalized gestures, we decided to add the participant's signature (performed as a hand movement in the air) and one custom complex gesture for each participant. So, each participant was asked to enter a total of 17 gestures: the 15 fixed gestures from Figure 4, one signature, and one self-defined complex gesture. Figure 5 portrays 2D sketches of the self-defined *complex* gestures each participant came up with. It should be noted that each of these gestures are complex from the participant's point of view. The participants were asked to come up with gestures that they thought would be more secure and less vulnerable to attackers.

Before entering each gesture, the participants were asked to practice the gesture a few times. Once comfortable, we asked them to perform the gesture 20 times: 10 times with single finger and 10 times with multiple fingers. In case of multiple fingers, participants were free to choose any number of fingers they wanted. They were also free to perform the gestures however they wanted, we just instructed them to be consistent among multiple trials of same gesture. To get rid of the carry-over effect, we counterbalanced the order of gestures to enter between participants. Additionally, we took video footage (see Figure 9) of the participants while performing the gestures and used that footage later for our third user study.

We also wanted to measure the participants' feelings and experience about the gestures qualitatively. Therefore, after performing each gesture, the participants rated that gesture in terms of pleasure, excitement, and easiness of use. To obtain excitement and pleasantness ratings for each gesture in an impartial way, we applied the *EmoCard* technique [6]. In addition, we were curious about the location acceptability of our technique. As mentioned in [19], both location and audience play a significant role in determining which gestures are acceptable. Therefore, after completing all the gestures, participants rated both our technique and the stroke-based on-screen authentication technique from Android phones in different hypothetical locations such as private, office, and public environments. Before leaving at the end of the study, participants were given the opportunity to express written opinions about the usability, applicability, and usefulness of the technique. The whole study took approximately 1 hour per participant.

In the following, we detail our findings on *accuracy*, *qualitative feedback*, and *location acceptability*.

Accuracy Analysis

To determine the accuracy of our system across multiple users, we calculated the Equal Error Rate (EER), a point where (at a certain threshold setting) the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal. FAR and FRR are inversely related. Thus, decreasing one measure will result in an increase in another. Calculating the EER is a standard method in the literature of measuring the accuracy of authentication systems [1]. We measured FAR and FRR using the following formulae:

$$\text{FAR} = \text{number of accepted forgeries / total number of forgery cases} \quad (9)$$

$$\text{FRR} = \text{number of rejected genuine authentications / total number of genuine authentications} \quad (10)$$

We measured the EER of single and multiple finger gestures separately. For each gesture, we had 10 single and 10 multiple finger samples from each user. As we mentioned previously, in our prototype application, the users had to provide 3 samples of their password gesture during the enrollment stage and the average of those were used as our reference template. Therefore, we decided to do the same for EER calculation and used the average of the first 3 samples for each gesture as a reference template for that gesture. Generating the template gesture in this way reflects the state of a hypothetical

Gesture	EER (Single Finger)	EER (Multiple fingers)
01 (Eight)	0.04	0.03
02 (Star)	0.02	0.02
03 (SpiralIn)	0.04	0.04
04 (Triangle)	0.05	0.04
05 (Circle)	0.04	0.03
06 (Z)	0.05	0.04
07 (SpiralOut)	0.04	0.04
08 (S)	0.05	0.05
09 (GT)	0.04	0.01
10 (LT)	0.04	0.02
11 (Three)	0.06	0.03
12 (N)	0.06	0.04
13 (Hat)	0.04	0.04
14 (P)	0.05	0.02
15 (L)	0.04	0.04
16 (Sign)	**0.0**	**0.0**
17 (Complex)	**0.0**	**0.0**
Average (all gestures)	0.038	0.029

Table 1. Equal Error Rate for each gesture, for single finger input and multiple finger input. Note that the *sign* and *complex* gestures have an EER of 0.

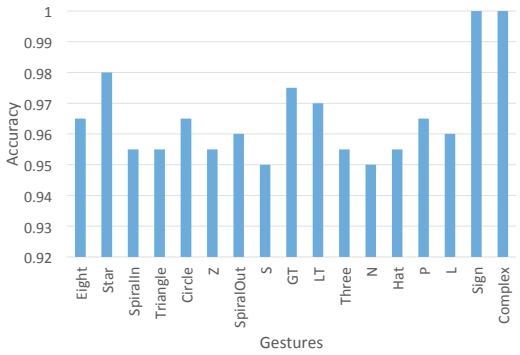


Figure 6. Achieved accuracy of the gestures used in the study. Note that the user-defined gestures (*sign* and *complex*) achieved 100% accuracy.

deployed AirAuth system that has just been enrolled. The last 7 samples were used as a test case. Trials with the same gesture from other participants were used as forgery cases. So, for each gesture performed with single or multiple fingers, we had $7 \times 15 = 105$ gestures to calculate FRR and $10 \times 14 \times 15 = 2100$ gestures to calculate FAR.

To calculate the EER of a particular gesture, it is necessary to determine the threshold ε at which FAR = FRR. As we described previously in our prototype application, we calculated the maximum distance among the first 3 reference samples and set it as our starting threshold, $\varepsilon_{\text{start}}$. Now if the distance between the reference template and the test case is smaller than $\varepsilon_{\text{start}}$, the gesture is verified. Otherwise it is a reject (see also Figure 3 (d, e)). So for each ε, we calculated the FAR and FRR using the previously defined formulas. By varying ε, we obtained different pairs of FAR and FRR for each ε. We plotted all these and obtained a receiver operating characteristic (ROC) curve for each gesture. The particular point of the curve where FAR = FRR is the value of the EER. Table 1 shows the value of EER for our different gestures. Figure 6 also shows a graphical version of the result where we averaged the single and multiple fingers accuracy of each gesture.

From Table 1, we can see that average EER for single and multi finger gestures are 0.038 and 0.029, respectively. So, we observe that gestures using multiple fingers appear to be more secure than single finger gestures. This trend is also observable in the EER of individual gestures, which is always lesser or equal for the multi finger version of a gesture. Due to the lower EER of multi-finger gestures we can assume that there will be a lower occurence of false positive authentications in comparison to single-finger gestures. A possible explanation for the lower EER we observed for multi-finger gestures is that a larger amount of information usable for authentication can be extracted from the participants during multiple finger gestures.

Qualitative Feedback

As we discussed previously, after each gesture, users rated that gesture in terms of *easiness*, *pleasantness* and *excitement* using the EmoCard approach. Figure 7 (a)–(c) illustrates the user rating for each gesture in terms of excitement, pleasantness, and easiness. We aimed to analyze if there is a relation between these ratings and the authentication accuracy, so we formulated the following hypothesis:

"Gestures that are highly rated by participants achieve a higher accuracy rate for biometric authentication."

To verify this hypothesis, we aggregated the pleasantness, excitement, and easiness rating for each gesture for all the participants and computed an accuracy matrix. We then performed a 3 (pleasantness) $\times$ 3 (excitement) $\times$ 5 (easiness) within subjects ANOVA. The results of this indicate that both pleasantness ($F_{1,13} = 11.09$, $p < 0.01$) and excitement ($F_{1,13} = 7.33$, $p < 0.01$) had a significant effect on accuracy. However, easiness did not exhibit any significant effect on accuracy ($F_{1,13} = 0.23$, $p > 0.05$). Also, no interactions were significant.

To further analyze the effect of these factors on accuracy, we correlated each of the qualitative measures with accuracy (see Figure 7 (d)–(e)). As we can see, both pleasantness and excitement are positively correlated with accuracy, which means that gestures that are more pleasant and exciting are also more secure from a biometric point of view. However, easiness does not have any relation with accuracy. Furthermore, the above ANOVA result also agrees with these correlation tests.

We investigated these results further and analyzed the effect of easiness on accuracy. The fixed set of 15 gestures (Figure 4) that we used in our evaluation is a set of easy gestures, as defined by the test subjects of the pilot study. That is why during the gesture trials, participants also felt comfortable and rated those gestures as easy. The complex and sign gestures were harder to perform compared to the easy gesture set, as the participants rated them as less easy than most other gestures (see Figure 7 (a)). However, the complex gestures were more accurate than the easy gestures. For all the other gestures in the fixed, simple gesture set, accuracy and easiness does not follow any fixed trend. Both ANOVA and correlation tests also agree with this finding. We can therefore conclude that easiness does not significantly influence the accuracy, in contrast to pleasantness and excitement.

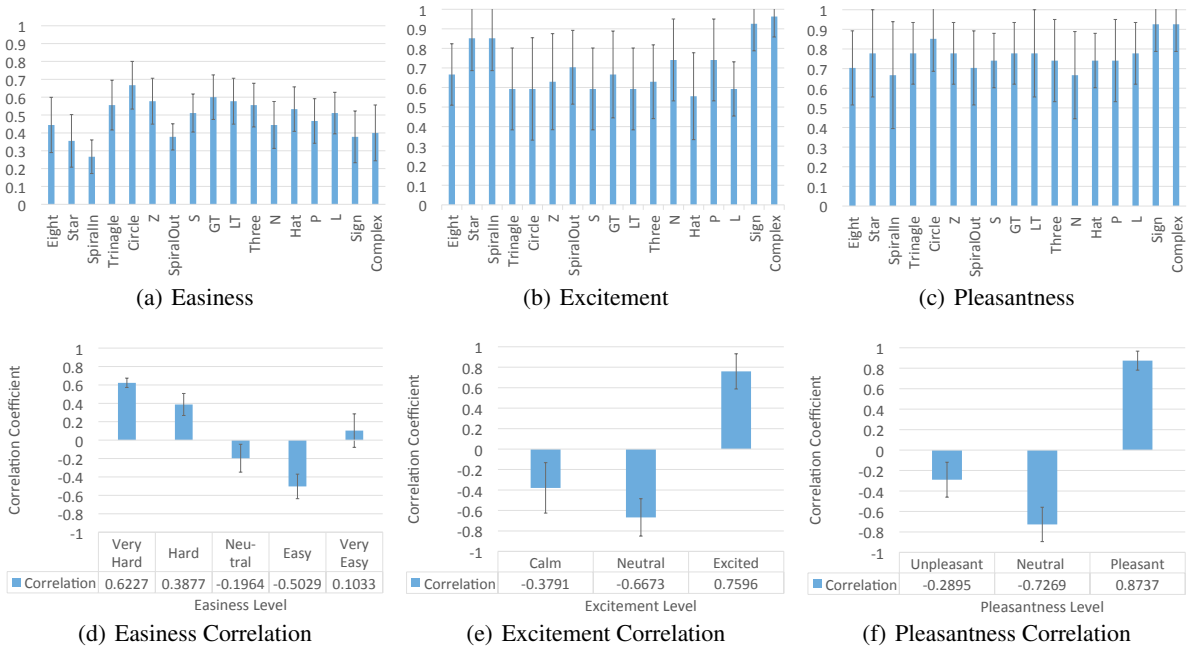


Figure 7. Results of qualitative feedback from the user study. (a)–(c) show the *easiness*, *excitement*, and *pleasantness* rating for different gestures. (d)–(e) show the correlation coefficients for Easiness, Excitement, and Pleasantness associated with gesture accuracy. The error bars show the standard deviation of the mean.

Sae et al. performed a similar analysis [22] and, in contrast to the present study found a positive correlation of accuracy with all three ratings. A possible reason for this might be because the gestures in their study were all exclusively predefined, and required the use of all five fingers. Thus, the participants of that study may have found those gestures harder to perform than their own personalized gestures. In our case, the fixed gestures were explicitly selected for their ease of entry and participants were asked explicitly to design the *complex* gestures to be more difficult than the simple gesture set. This may be an explanation why easiness did not have an effect on accuracy in our study.

Location Acceptability

To assess the location acceptability of our technique, we asked participants to imagine AirAuth being used to unlock their phone. Then we gave them three hypothetical usage situations: private (e.g., home), office, and public (e.g., pub, street). After that, we told them to rate both our technique and on-screen gestural password entry as found on Android devices on a Likert scale from 1 to 5 (1 being the least preferred). We only considered feedback from 14 participants because one of the participants was not familiar with the on-screen gesture technique. Figure 8 shows the normalized results we obtained from the participants. As we can see, participants preferred our technique in all three situations. In the private setting, 10 out of 14 participants chose our technique. 4 participants thought that they would be at less risk in a private setting and hence they chose the on screen gesture. In office and public settings, 12 out of 14 participants chose our technique. 2 participants remarked that it would be slightly socially awkward to perform gestures in front of other people.

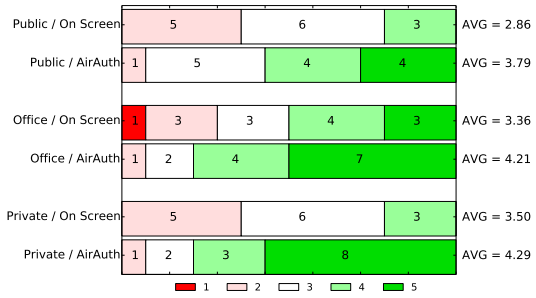


Figure 8. Location acceptability comparison between AirAuth and on screen gestures. The colors represent the given ratings, i.e., 1 (red) = *least preferred* to 5 (green) = *most preferred*. The number within the colored bars show how often the participants rated the technique that way.

Also, they felt more comfortable using a standard password or on-screen authentication method in public settings.

We also collected verbal user comments about why they would prefer our technique which included: *"simple and faster"*, *"more secure than traditional approaches"*, *"simply cool"*. One participant commented: *"It's great because I do not have to touch my phone when I am cooking. I would like to use similar techniques to do more stuff like receiving calls and reading text messages"*.

User Study 3: Shoulder Surfing Attack

To evaluate the vulnerability of our system against shoulder surfing attacks, we conducted a further user study. In this study, we recruited 15 participants. 12 were male and 14 of them were right handed ranging in age from 21 to 36 (average: 27.8, SD: 4.6). All participants were fairly proficient

Figure 9. Setup for shoulder surfing study. Participants were videotaped while doing the gesture trials. A standard (2D) digital SLR camera was used to capture the videos.

with computers and smartphones and one participant was familiar with in-air gestures.

As we mentioned earlier, while participants were doing the gesture trials, we recorded video footage of certain gesture entries from each of the participants. Figure 9 shows the setup we used to capture the videos. In particular, we also videotaped the sign and complex gestures from all users. To record video, we used a standard (2D) digital SLR camera instead of a 3D depth sensor (e.g., Kinect), as standard cameras are portable. They are also readily available to potential attackers [23]. In the previous study, we selected a random gesture from each participant and made a set of 15 videos of fixed (simple) set gesture trials from Figure 4. So, in total we had 45 video clips: 15 videos of fixed gestures, 15 of sign gestures, and 15 of complex gestures. Since we observed (see Table 1) that multiple finger gestures were more secure, we decided to only use those gestures for this study. Therefore, all the videos we selected were videos of multiple finger trials. For each of the 45 gestures, we took first 3 trials in terms of DTW distance: T_1, T_2, and T_3. We then calculated the average of the three trials and used that as our template: $T_{\text{reference}}$. The threshold ε was again chosen as the maximum distance between T_1, T_2, and T_3.

Experiment

After explaining the study to all 15 participants, we asked them to complete a short demographic survey. We then explained the shoulder surfing attack procedure to them. Afterwards, we conducted the experiment. At first they were shown the video clips of the gestures. They could watch the clip as many times as they wanted before they proceeded to forge the gesture. They could also slow down the video. The order of the videos was counterbalanced. We decided to assign each of the participants 15 fixed gesture videos, 3 sign videos, and 3 complex gesture videos. For gestures from fixed the set, participants had 3 attempts to forge the gestures. For sign and complex gestures, they had 5 authentication attempts. So, each participant had $(3\times15)+(5\times3)+(5\times3) = 75$ attempts to forge the video gestures, which resulted in total 1125 attacks. The whole study took approximately 90 minutes per participant. Just like our previous study, for each of the forge trials T_{forge}, we calculate the DTW distance (d) between $T_{\text{reference}}$ and T_{forge}. Now if d is less than ε, the trial T_{forge} is authenticated. Otherwise it is rejected.

Results and Analysis

Out of the total 1125 attacks, only 24 were successful. That means that in 97.8% of times, our system was able to resist a forgery attack. For the sign and complex gestures, the distances between the real and forgery gestures were relatively high. That means that even if we choose a high value of threshold ε, forgery attempts can still be recognized. We investigated this effect on sign gestures and calculated the number of successful attacks by sweeping the threshold value from 20 to 50 [23]. Figure 10 shows the result of the threshold analysis for study 3. As we see, for a threshold value of 20 to 24, our system can reject all the forgery attempts while authenticating all the real attempts (FAR = FRR = 0). Therefore we can conclude that for complex gestures such as a user's signature, user can select a higher threshold value which permits low false acceptance and false rejection rates.

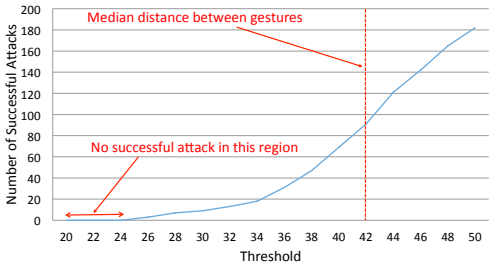


Figure 10. The number of successful attacks is dependent on the threshold setting. For a wide region of the threshold (up to a setting of 24), no successful attacks are possible. The vertical dashed line shows the median distance between gestures, which shows that our system has good overall robustness, since the useful threshold settings are relatively far away from this mark.

User Study 4: Similarity and Repeatability Analysis

To evaluate the similarity and also the repeatability of authentication over time, we performed a longitudal study. We recruited 5 participants ranging in age from 22 to 29 (AVG: 24.8, SD: 2.5). All participants were right handed male. For each participant, gestures were collected in 3 different sessions (S_1, S_2 and S_3) within a period of 10 days. The difference between two consecutive sessions was at least 4 days per participant. In each session, participants performed each of our 17 gestures 10 times. Again, we chose multiple finger gestures as they have a higher accuracy.

As in study 2, we used the average of first three S_1 gestures for each participant as enrollment gesture. Then for each gesture of S_2 and S_3, we calculated the average distance of the gesture to the enrollment gesture. To assess the accuracy under a simulated attack, we also calculated the average *attack distance* of each gesture to video-based attacks performed in Study 3. Figure 11 shows the result of this study. As we see, the average distance from S_1 grows from S_2 to S_3. The average attack distance of each gesture remains considerably larger than the difference between sessions. This indicates that AirAuth is robust enough to accomodate variation of gesture inputs at least for the time span we analyzed. However, the growth of the matching distance between time-separated sessions suggests that periodically updating templates as described previously is important.

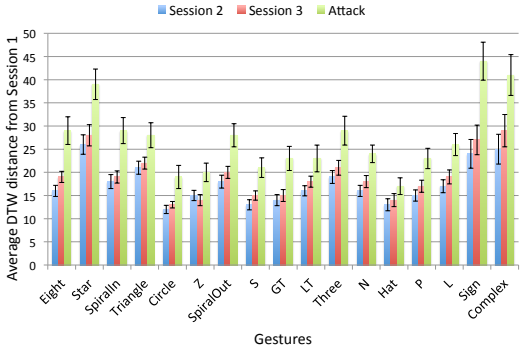


Figure 11. Average distance for each gesture of session 2 (blue) and 3 (red) from session 1. Average distance from study 3 (green) is also shown. The error bars show the standard deviation of the mean.

To explore the accuracy of our system with template updates, we performed another EER analysis (as in Study 2). Again, we considered the average of first three S_1 gestures for each participant as enrollment gestures. As we described earlier, we propose updating the template upon the first valid authentication attempt if the template is older than 24 h. So, when using template updates for this study, the template gesture is updated after the first successful authentication in S_2. Consequently S_3 gestures will be compared to the updated template. Conversely, without template updates, S_3 gestures are compared to the original reference gesture from S_1. We considered both of these cases during EER analysis. For accuracy measurement purposes, all S_3 gestures were considered genuine authentication attempts and all gestures from study 3 were treated as video-based forgery attempts. Figure 12 shows the result.

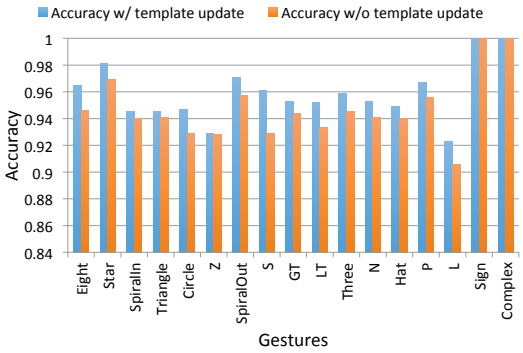


Figure 12. Achieved average accuracy for S_3 gestures with and without template update. User-defined gestures (*sign* and *complex*) achieved an accuracy of 100% in both cases.

We calculated the average across all the gestures and achieved an accuracy of 95.9% and 94.7% with and without template update, respectively. Therefore we can conclude that AirAuth system with template update is more robust against day to day variation of gestures.

CONCLUSION AND FUTURE WORK

In this paper we have presented a novel approach to gesture-based authentication, which makes use of biometric features extracted from in-air gestures tracked by short range depth camera. AirAuth is simple, robust and easy to implement. The DTW algorithm we used does not require large amounts of CPU time, so AirAuth is easily deployable on mobile devices, as well as on desktop systems. We conducted a short survey and found out a set gestures that are simple and easy to perform from user's point of view. Not only did AirAuth show a high EER-bases accuracy of 96.6% for simple and 100 % for complex gestures, in a realistically-designed video-based forgery scenario, AirAuth was able to resist attack for 97.8% of all attack attempts. Our longitudinal repeatability study shows that with a template update mechanism, AirAuth maintains consistent accuracy levels over time.

We obtained qualitative feedback from our participants regarding their experience about each gesture. Based on their rating, we found a positive correlation between the security of the authentication gestures and user experience (expressed as excitement and pleasantness), a characteristic not present in traditional password based systems [27]. Our users also gave our system a higher usage acceptability rating in three different usage situations compared to on screen gesture authentication technique found on Android devices.

Our system allows users to register and use their personalized gestures for authentication. Moreover, the results we obtain indicate that personalized gestures (e.g., signature, complex) are also more secure and rated higher by users in terms of user experience. Therefore, we envision our system to be used in application level security. Whereas in the present work we have studied the immediate validity of AirAuth's authentication technique, in the future, we intend to also put it in more direct comparison with traditional, password-based authentication mechanisms. To accomplish this, we plan to deploy AirAuth to several employees of our laboratory, replacing the standard password-based login for their workstations. We hope that this longitudinal study with a larger user based and vastly more entered gestures will give us further insights on the long-term robustness of the biometric we use, and also how AirAuth will be received in a real-life scenario. Furthermore, we hope to obtain a meaningful comparison to existing authentication methods. On an application level, we also envision using the gesture recognition component of AirAuth to enable controlling device functions, which is useful in situations where the users are unable or reluctant to touch their devices' touch screens, e.g., when cooking. Since many depth cameras are equipped with a combined RGB camera, our biometric could be further strengthened by, for instance, adding face recognition or further 2D visual features of the user's hands (e.g., as in [18]).

REFERENCES

1. Alpcan, T., Kesici, S., Bicher, D., Mihçak, M. K., Bauckhage, C., and Çamtepe, S. A. A lightweight biometric signature scheme for user authentication over networks. In *Proc. 4th international conference on Security and privacy in communication netowrks*, ACM (2008), 33.

2. Aviv, A. J., Gibson, K., Mossop, E., Blaze, M., and Smith, J. M. Smudge attacks on smartphone touch screens. In *Proc. 4th USENIX conference on Offensive technologies*, USENIX Association (2010), 1–7.

3. De Luca, A., Von Zezschwitz, E., and Hußmann, H. Vibrapass: secure authentication based on shared lies. In *Proc. CHI*, ACM (2009), 913–916.
4. De Luca, A., von Zezschwitz, E., Nguyen, N. D. H., Maurer, M.-E., Rubegni, E., Scipioni, M. P., and Langheinrich, M. Back-of-device authentication on smartphones. In *Proc. CHI*, ACM (2013), 2389–2398.
5. De Luca, A., von Zezschwitz, E., Pichler, L., and Hussmann, H. Using fake cursors to secure on-screen password entry. In *Proc. CHI*, ACM (2013), 2399–2402.
6. Desmet, P., Overbeeke, K., and Tax, S. Designing products with added emotional value: Development and appllcation of an approach for research through design. *The design journal 4*, 1 (2001), 32–47.
7. Ehrenberg, R. The digital camera revolution: Instead of imitating film counterparts, new technologies work with light in creative ways. *Science News 181*, 2 (2012), 22–25.
8. Farella, E., OModhrain, S., Benini, L., and Riccó, B. Gesture signature for ambient intelligence applications: a feasibility study. In *Pervasive Computing*. Springer, 2006, 288–304.
9. Jorgensen, Z., and Yu, T. On mouse dynamics as a behavioral biometric for authentication. In *Proc. 6th ACM Symposium on Information, Computer and Communications Security*, ACM (2011), 476–482.
10. Kainda, R., Flechais, I., and Roscoe, W. A. Security and usability: Analysis and evaluation. In *Availability, Reliability, and Security, 2010. ARES'10 International Conference on*, IEEE (2010), 275–282.
11. Kumar, M., Garfinkel, T., Boneh, D., and Winograd, T. Reducing shoulder-surfing by using gaze-based password entry. In *Proc. 3rd symposium on Usable privacy and security*, ACM (2007), 13–19.
12. Liu, J., Zhong, L., Wickramasuriya, J., and Vasudevan, V. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing 5*, 6 (2009), 657–675.
13. Liu, S., and Silverman, M. A practical guide to biometric security technology. *IT Professional 3*, 1 (2001), 27–32.
14. Maeder, A., Fookes, C., and Sridharan, S. Gaze based user authentication for personal computer applications. In *Proc. 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing*, IEEE (2004), 727–730.
15. Maurer, M.-E., Waxenberger, R., and Hausen, D. Broauth: evaluating different levels of visual feedback for 3d gesture-based authentication. In *Proc. AVI*, ACM (2012), 737–740.
16. Monrose, F., and Rubin, A. D. Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems 16*, 4 (2000), 351–359.
17. Patel, S. N., Pierce, J. S., and Abowd, G. D. A gesture-based authentication scheme for untrusted public terminals. In *Proc. UIST*, ACM (2004), 157–160.
18. Ramakers, R., Vanacken, D., Luyten, K., Coninx, K., and Schöning, J. Carpus: a non-intrusive user identification technique for interactive surfaces. In *Proc. UIST*, ACM (2012), 35–44.
19. Rico, J., Crossan, A., and Brewster, S. Gesture-based interfaces: Practical applications of gestures in real world mobile settings. In *Whole Body Interaction*. Springer, 2011, 173–186.
20. Riley, S. Password security: What users know and what they actually do. *Usability News 8*, 1 (2006).
21. Roth, V., Richter, K., and Freidinger, R. A pin-entry method resilient against shoulder surfing. In *Proc. 11th ACM conference on Computer and communications security*, ACM (2004), 236–245.
22. Sae-Bae, N., Ahmed, K., Isbister, K., and Memon, N. Biometric-rich gestures: a novel approach to authentication on multi-touch devices. In *Proc. CHI*, ACM (2012), 977–986.
23. Sahami Shirazi, A., Moghadam, P., Ketabdar, H., and Schmidt, A. Assessing the vulnerability of magnetic gestural authentication to video-based shoulder surfing attacks. In *Proc. CHI*, ACM (New York, NY, USA, 2012), 2045–2048.
24. Sakoe, H., and Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing 26*, 1 (1978), 43–49.
25. Tari, F., Ozok, A., and Holden, S. H. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proc. 2nd symposium on Usable privacy and security*, ACM (2006), 56–66.
26. Vatavu, R.-D., Anthony, L., and Wobbrock, J. O. Gestures as point clouds: a $p recognizer for user interface prototypes. In *Proc. 14th ACM international conference on Multimodal interaction*, ACM (2012), 273–280.
27. von Zezschwitz, E., De Luca, A., and Hussmann, H. Survival of the shortest: A retrospective analysis of influencing factors on password composition. In *Proc. INTERACT 2013*. Springer, 2013, 460–467.
28. von Zezschwitz, E., Koslow, A., De Luca, A., and Hussmann, H. Making graphic-based authentication secure against smudge attacks. In *Proc. IUI*, ACM (New York, NY, USA, 2013), 277–286.
29. Wiedenbeck, S., Waters, J., Sobrado, L., and Birget, J.-C. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In *Proc. AVI*, ACM (2006), 177–184.
30. Zhang, Y.-B., Li, Q., You, J., and Bhattacharya, P. Palm vein extraction and matching for personal authentication. In *Advances in Visual Information Systems*. Springer, 2007, 154–164.