Continuous Mobile Authentication Using A Novel Graphic Touch Gesture Feature

Xi Zhao, Tao Feng, Weidong Shi
Computer Science Department, University of Houston
4800 Calhoun Rd, 77054, Houston, TX
zhaoxi@ieee.org, {tfeng3, larryshi}@cs.uh.edu

Abstract

Behavioral biometric on mobile devices has begun to gain attention in recent years and the feasibility of touch gestures as a novel biometric modality has been investigated lately. In this paper, we propose a novel Graphic Touch Gesture Feature (GTGF) to extract the identity traits from the touch traces. The traces’ movement and pressure dynamics are represented by intensity values and shapes of the GTGF. To evaluate its usability on the authentication problem, touch gesture datasets have been collected which includes six commonly used touch gestures. A Equal Error Rate of 2.62% has been achieved combining six gestures together, which demonstrated the effectiveness of the proposed methods.

1. Introduction

There will be 640 million tablets and 1.5 billion smartphones in use in 2015 globally according to the market analysis [1]. The increasing ubiquity of such hand-held devices makes the problem of securing the data stored and accessed from such devices of paramount importance. However, current dominated point mobile authentication is deficient. It cannot detect intruder after user passed the point-entry, and is vulnerable to simple attack, such as smudge attacks [2] after loss or theft. Thus, the continuous or implicit authentication approach, which intelligently monitors and analyzes the user-device interaction to ensure the correct identity, can either complement the point based authentication or even substitute it if the approach satisfies particular accuracy requirements.

The feasibility to use touch gestures as a biometric modality has been discussed in [6][15][7]. Touch traces are affected by two biometric features, i.e., the user hand geometry and the muscle behavior. The variations of biometric characteristics have the potential to provide user differentiation. It can be observed in Fig. 1 that most of the touch traces from one user tend to follow a similar pattern while the patterns vary among different users.

In this paper, we investigate the touch based continuous mobile authentication via proposing a novel Graphic Touch Gesture Feature (GTGF). The GTGF converts the touch traces to images so that the touch dynamics can be represented in an explicit manner which combines the discriminative power from touch trajectories and tactile pressures. The touch sequences are first segmented and normalized so that traces have a fixed number of sample points. Then the samples on the normalized traces are converted into shapes and intensity values of the GTGF. We choose to evaluate the proposed method on three sets of touch gestures (flick up/down, flick right/left, zoom in/out) simply because they are among the most used gestures during user-device interaction [3]. Flick up/down are generally used when user reading texts, scroll menus etc. Flick right/left are generally used when users browse photos, translate screens, unlock phones etc, and zoomin/out are generally used whenever users switch between details and overviews of the contents.

Our main contributions are: (i) proposed the Graphic Touch Gesture Feature to represent touch dynamics intuitively and explicitly; (2) compared the discriminative power of the six gestures under three different score metrics; and (3) compared the proposed methods with the literature for continuous authentication.

The rest of paper is organized as follows. First, we dis-
cuss the related work in section 2. In section 3, we briefly overview our touch based authentication system. The extraction of imaged touch gesture feature is depicted in section 4, and the score metrics and normalization methods are presented in section 5. Section 6 presents the experimental results. In section 7, we conclude our study.

2. Related Work

Different from physiological biometrics which relies on static physical attributes (i.e., fingerprint [12], facial features [20]), behavioral biometrics adopts identity-invariant features of the human behavior to identify or verify people during the daily activities [14].

Prior research on behavioral biometrics has shown that biometric traits can be extracted from physiological characteristics of hand geometry [19], signatures [5], [10] using a pen or stylus, and key stroke and mouse dynamics [11] [16]. Studies on exploring the touch gesture as a biometric modality for mobile devices is quite recent [17]. Feng et al. [6] extracted finger moving speed and acceleration of touch gesture as features for identification. Luca et al. [13] directly computed the distance between traces using dynamic time warping algorithm. Sae-Bae et al. [15] designed 22 special touch gestures for authentication, most of which involve all five finger gestures. They computed dynamic time warping distance and Frechet distance between multi-touch traces. However, users have to perform their pre-defined touch gestures for authentication, which thus cannot be transparent and non-obtrusive to users. Frank et al. [7] studied the correlation between 22 analytic features from touch traces and classified these features using k-nearest-neighbors and Support Vector Machine approaches. However, the analytic features did not represent the touch trace dynamics.

The proposed method is quite different from existing works. Instead of directly comparing the trace trajectories, we convert the traces’ geometry trait to the image space so that the dynamics of traces can be represented intuitively and explicitly. Furthermore, the pressure dynamics is emphasized in our method via fusing it with the movement dynamics. This is different from extracting a single mid-stroke pressure value in [7] or simply being ignored in other studies [15][13].

3. Touch Based Continuous Mobile Authentication

We designed a touch based user authentication system for continuous user verification during the normal user-device interaction. Figure 2 depicts a high level diagram of the system. After user has been authorized by the login point authentication, the system continues to authenticate the mobile user in the background using intercepted touchscreen readings from the touch gesture engine. The authentication module is invoked by the event capture module when touch-related inputs are received. The authentication process is transparent to the smartphone user so that the usability is not affected. Only when sufficient evidences are collected that the current user is not the owner, traditional point user authentication is activated from the explicit user access control module. The authentication algorithm running in the user authentication module is based on the proposed GTGF and detailed in the following section.

4. Graphic Touch Gesture Feature

A touch trace is a series of x-y coordinates of finger touch points with pressure values and time stamp. From each series, we can extract the time duration, the length of touch traces, the directions and speeds of finger movements, and the tactile pressures. All these extracted features imply the user’s hand geometry and muscle behavior. In this section, we introduce a novel feature to convert a touch trace into an image, where all aforementioned features can be represented in an intuitive and explicit manner.

The first step is to extract the touch traces from the touch screen outputs. All captured points with their pressure over a threshold are read. Thus, a single fingertip touch trace is encoded as a series of N point samples \( S_n = (x_n, y_n, t_n, p_n), n \in 1, 2, ..., N \), where \( x_n, y_n \) is the touch coordinate, \( t_n \) is the time stamp and the pressure \( p_n \). Multiple fingertip gestures (i.e., Pinch, Spread) include multiple series of samples. Extracted traces are further filtered into one of the six predefined gestures, as described in Tab. 1. The filtering is based on screen regions where the traces start and end. In case of multiple fingertip gestures, the Euclidean distances between two fingers at the start and end of traces are also used.

In order to normalize and register the traces with different number of points and time intervals, we use cubic interpolation, as shown in Eq.1, to resample the traces in terms

\[
R_{\text{GTGF}}(x_n, y_n, t_n, p_n) = \sum_{i=0}^{N-1} \left( x_n, y_n, t_n, p_n\right) \cdot \text{cubic interpolation} \left( t_n \right)
\]
of their x-y coordinate, time and pressure series so that all traces have a fixed number of sampled points (e.g., 50).

\[
\hat{S}(f_0, f_1, f_2, f_3, t) = \left( -\frac{1}{2} f_0 + \frac{3}{2} f_1 - \frac{3}{2} f_2 + \frac{1}{2} f_3 \right) t^3 \\
\quad + \left( f_0 - \frac{5}{2} f_1 + 2 f_2 - \frac{1}{2} f_3 \right) t^2 \\
\quad + \left( -\frac{1}{2} f_0 + \frac{1}{2} f_2 \right) t + f_1
\]

(1)

where \( t \) represents timestamp where to interpolate, \( f_n \) represents the known samples at predefined time, \( \hat{S} \) is the cubic interpolation function and outputs the interpolated value \( f_t \) (i.e., \( x, y, p \)) at timestamp \( t \). Cubic interpolation is chosen because it is the simplest method that offers true continuity between the samples. After normalization, a single touchtip trace \( \hat{S} \) (i.e., UD, DU, LR, RL) is obtained consisting of 50 samples or a multiple touchtip trace (i.e., ZI, ZO) is obtained consisting of 100 samples, 50 samples of \( \hat{S} \) and 50 samples of \( \hat{S}' \). Each sample includes a pair of x-y coordinates \( \hat{x}, \hat{y} \), a pressure value \( \hat{p} \) and a timestamp \( \hat{t} \). Note that the time intervals between samples may vary since traces may have the different time duration but the same number of samples.

\[
\hat{S}_n = (\hat{x}_n, \hat{y}_n, \hat{t}_n, \hat{p}_n), n \in 1, 2, ..., 50
\]

\[
\hat{S}'_n = (\hat{x}'_n, \hat{y}'_n, \hat{t}_n, \hat{p}'_n), n \in 1, 2, ..., 50
\]

(2)

After obtaining the \( \hat{S} \), we further convert the normalized traces into GTGF \( T \). This conversion needs to be conducted in a way that the discriminative power of the original traces are preserved while the graphic features are easy to compute for a mobile based platform. Thus, we create an zero-valued image template \( T \) with resolution set to \( 100 \times 150 \). This size is a tradeoff between the feature ‘discriminability’ and computational efficiency.

For each sample \( \hat{S}_n \), we use a block \( C_n \) which has a width of three columns to represent its information. The block is evenly divided into upper block \( C_n^u \) and lower block \( C_n^d \). In general, the upper subblock describes the \( x \) direction related features exclusively, and the lower subblock describes the \( y \) direction related features exclusively. The height \( H_n^p \) and the intensity of the upper and lower subblock \( I_n^u, I_n^d \) are the three important properties which are used to represent the tactile pressure and the movement dynamics along \( x \) and \( y \) axes at the timestamp \( \hat{t}_n \). They are computed as:

\[
I_n^u = \left[ I_m \ast \frac{U_x - \Delta \hat{x}_n}{U_x} \right] \\
\Delta \hat{x}_n = \hat{x}_{n+1} - \hat{x}_n
\]

(3)

\[
I_n^d = \left[ I_m \ast \frac{U_y - \Delta \hat{y}_n}{U_y} \right] \\
\Delta \hat{y}_n = \hat{y}_{n+1} - \hat{y}_n
\]

(4)

\[
H_n^p = \left[ H_c \ast \frac{\hat{p}_n}{I_p} \right]
\]

(5)

where \( \lceil \rceil \) is the ceiling function fetching the nearest greater integer. \( H_c = 50 \), which is half of the preset image height, \( I_m = 128 \) is chosen because it evenly divides the intensity space \([0, 256]\) so that the sign and the absolute value of \( \Delta x_n \) and \( \Delta y_n \) can be represented by intensity values. The \( \Delta x_n \) values below 128 imply the hand moves to the right, and the values above 128 imply the hand moves to the left. The
$\Delta y_n$ values below 128 imply the hand moves up, and the values above 128 imply the hand moves down. The greater $\text{abs}(H^p_n - 128)$ or $\text{abs}(T^p_n - 128)$ is, the faster the fingertip moves. Meanwhile, the greater $H^p_n$ is, the harder the finger touches the screen. We repeat Eq. 3, 4, 5 for all $N$ samples on a trace so that all the block $C_n$ are computed. For multiple touchtip gestures, we create $T$ and $T'$ for $S$ and $S'$ respectively. The parameters $L_p, U_x, U_y$ are set to 0.35, 20 and 30 respectively via a grid search on a validation dataset. In this way, the direction, the pressure and the dynamics of the traces are directly encoded into the GTGF. As depicted in Fig. 3, the images within the same row are quite similar with each other. But major differences can be observed between two rows. The differences stems from the different user identities.

The extraction of GTGF has multiply benefits. First, original traces have different spacial topology and temporal duration. This causes difficulties to register the traces. The proposed GTGF solves this difficulty via resampling the traces and fitting them into the graphic template $\mathcal{T}$. Second, the dynamics is considered to be an important factors in other pattern recognition problems, i.e., facial expression recognition [18] and speech recognition [8]. However, they are not commonly considered in the touch gesture based authentication literature. The proposed GTGF is able to represent the gesture dynamics in terms of movement and pressure intuitively and explicitly. Third, due to the inhomogeneity between location and pressure data, it is difficult to combine their discriminative power in the feature level. However, extraction of GTGF takes both features into consideration and fuse their discriminative power together.

5. Score Metrics and Normalization

We use image processing techniques to compute the score between two images. Since all the GTGF have the same dimension (the size of $\mathcal{T}$), the score metrics can be computed directly on the images without extra processing and registration steps.

The first score metrics we use is based on the normalized cross correlation:

$$NC(a, b) = \frac{\left< T_a, T_b \right>}{\left< T_a^2 \right>^{\frac{1}{2}} \left< T_b^2 \right>^{\frac{1}{2}}}$$

where $\left< \cdot, \cdot \right>$ is the inner product and $\left< \cdot \right>$ is the $L_2$ norm. The normalized cross correlation described the similarity between two GTGF $T_a$ and $T_b$. We get the distance in the verification via $D_n(a, b) = 1 - NC(a, b)$.

The second score metrics is $L_1$ distance:

$$L_1(a, b) = \sum_{l \in \mathcal{T}} |T^l_a - T^l_b|$$

where $l$ is the index of image pixels and $\mathcal{T}$ is the image template with the same size as $T_a$ and $T_b$.

The third score metrics is $L_2$ distance:

$$L_2(a, b) = \sqrt{\sum_{l \in \mathcal{T}} (T^l_a - T^l_b)^2}$$

For multiple touch gestures with multiple traces, scores from $T, T'$ are averaged for authentication.

Each query gesture have a set of scores corresponding to query-target pair comparison. However, different queries may have different distributions of their score set. This affects the verification performance when setting the canonical threshold for verification. Thus, we use the Z score normalization [4] to normalize their score distributions into zero mean and unit variance. Their mean $\mu$ and the standard deviation $\sigma$ need to be pre-computed, $D$ is a score value and $Z$ is the normalized score:

$$Z = \frac{D - \mu}{\sigma}$$

6. Experimental Results

6.1. Data Acquisition

To collect user touch gesture data, we developed an Android program which captures touch gestures using a standard API of Android system. When fingers contacted the touchscreen, it started to record the trace by recording raw touch samples from the API. For each sample in a trace, an event flag (e.g., onDown, OnScroll, onFling, Zoom), the absolute event time in $\text{ms}$, and the (x,y) coordinates and the tactile pressure per contacted finger are captured.

We recruited 30 subjects for our study, of which 28 were right-handed, 24 had touchscreen device experience. The data acquisition contains 6 sessions collected over several weeks. In the first session for each subject, we explained the purpose of the study and the usage of our data acquisition program. Then, 5 to 10 minutes were left to the subject for practice. After he/she can use the program in a natural way, the subject provided 20 traces for each gesture in Tab. 1. Then the following 5 sessions occurred with at least three days in-between two consecutive sessions. Subjects provided at least 20 traces per gesture during each session. The touch gestures are performed in the way that the subjects feel natural and comfort. Thus, subjects may vary the manner they held the phone (i.e., left-hand holding vs right-hand holding, whole palm holding vs half palm holding), the hand pose (palm down vs palm up) or the fingers used to perform the single touch gesture (thumb vs index finger) or the finger orientation between sessions. We name the dataset collected in the first session as UH-TOUCHv1 and the dataset collected in the following sessions as UH-TOUCHv2. In our study, the UH-TOUCHv1 is mainly utilized for the feature evaluation and authentication system.
Table 2: Score metric comparison on user verification scheme

<table>
<thead>
<tr>
<th></th>
<th>NS</th>
<th>L1</th>
<th>L2</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>DU</td>
<td>14.18</td>
<td>8.83</td>
<td>9.66</td>
<td>10.89</td>
</tr>
<tr>
<td>UD</td>
<td>16.46</td>
<td>7.07</td>
<td>7.98</td>
<td>10.50</td>
</tr>
<tr>
<td>LR</td>
<td>16.82</td>
<td>12.02</td>
<td>12.61</td>
<td>13.82</td>
</tr>
<tr>
<td>RL</td>
<td>16.83</td>
<td>11.86</td>
<td>12.05</td>
<td>13.58</td>
</tr>
<tr>
<td>ZI</td>
<td>19.31</td>
<td>14.12</td>
<td>15.40</td>
<td>16.28</td>
</tr>
<tr>
<td>ZO</td>
<td>20.16</td>
<td>13.78</td>
<td>16.55</td>
<td>16.83</td>
</tr>
<tr>
<td>Mean</td>
<td>17.29</td>
<td>11.28</td>
<td>12.38</td>
<td>12.38</td>
</tr>
</tbody>
</table>

The Equal Error Rate is presented in each blank as a criterion for user verification. NS stands for the normalized cross correlation, L1 stands for \( L_1 \) distance, and L2 for \( L_2 \) distance.

The UH-TOUCHv2 is mainly utilized in the research on continuous authentication.

Equal Error Rate was mostly used as the evaluation criterion for our verification experiments. It is the common value when the false acceptance rate (FAR) and the false rejection rate (FRR) are equal. We opt to use it simply because it accounts for the trade-off between FAR and FRR. Meanwhile, we also reported ROC curves in some experiments for clarity purpose.

6.2. Experimental Results

First we evaluated the performance of GTGF with different score metrics, including the \( L_1 \) distance, \( L_2 \) distance and the normalized cross correlation. For the experiment setup, we randomly selected half portion of the traces per subject per gesture from UH-TOUCHv1 dataset as gallery (target) set and used the other half portion per subject per gesture from UH-TOUCHv1 dataset as probe (query) set. Table 2 depicts the EER of authentication using each gesture respectively and compares the different distance metrics. The gestures DU and UD achieve average EER below 11% but ZI and ZO achieve average EER above 16%. For different distance metrics, the mean EERs achieved by the \( L_1 \) distance, the \( L_2 \) distance and the normalized correlation are 11.28%, 12.38% and 17.29%. In general \( L_1 \) distance performs better than the two distance metrics. This implies that DU and UD gestures could have more discriminative power than the other touch gestures and the \( L_1 \) distance is more discriminable than the other two metrics among subjects. In the following tests, we opt to adopt the \( L_1 \) distance as the distance metrics for our authentication scheme.

To assess the fusion strategy of gestures for continuous authentication, we then tested it on the UH-TOUCHv2. For the experiment setup, we randomly selected half portion of the traces in each session per gesture in the UH-TOUCHv2 dataset as a gallery and used the left portion from UH-TOUCHv2 dataset as a probe. After score computation and Z-normalization, we can obtain one score matrix for each gesture. Then we fused the multiple gestures using the sum rule, which is proved by Kittler et. al., [9] to be superior in comparison of other rules, i.e., product, min, max, median rule. Figure 4 depicts a comparison of the Receiver Operating Characteristic (ROC) curves for different fusion schemes and methods: a) GTGF-A: fusion of the score matrices of the six gestures computed from the proposed method; b) GTGF-S: fusion of the score matrices of four single touchtip gestures computed from the proposed method; c) TA-A: fusion of the score matrices of the six gestures computed from the method in [7]; d) DM-A: fusion of the score matrices of the six gestures computed from the method in [15]; e) GTGF-M: fusion of the score matrices of the two multiple touchtip gestures computed from the proposed method. A Receiver operating characteristic (ROC) curve was created by plotting the true acceptance rate (TAR) vs. false acceptance rate (FAR), at various threshold settings. Since the true acceptance rate is equal to 100% subtracting false rejection rate, the EER is the intersection between a ROC curve and the black straight line connecting \([0\%,100\%]\) with \([100\%,0\%]\) in the figure. From Fig. 4, the EER are 2.62% (GTGF-A), 4.31% (GTGF-S), 6.07% (TA-A), 7.06% (DM-A) and 7.81% (GTGF-M). Comparing different fusion schemes, the best performance has been achieved by fusing all six gestures while the worst performance is obtained by fusing just two multiple fingertip gestures, namely ZO, ZI. On one hand, as most of score-level fusion scheme, combining more scores from different channels would probably increase the overall performance. GTGF-A scheme includes six gestures while GTGF-M includes only two. On the other hand, gestures ZO,ZI themselves have a relative low discriminative power compared to other gestures referred to Tab. 2. Furthermore, the performance of the proposed method demonstrated a clear improvement over the other methods in the scheme of com-
bining all the six gestures. There are two reasons for the improvement. First, compared with the method in [15], we adopted the tactile pressure in our method, which contains extra clues on subject’s muscle behavior. However, their method just ignored this information. Second, compared to the work in [7] which extracted 22 static analytic feature values, the proposed method includes movement dynamic and pressure dynamic of the touch gesture during feature extraction.

7. Conclusion

This paper presents a novel mobile authentication approach based on touch gestures. The identity traits in touch traces are extracted using the proposed graphic touch gesture feature. We first collected a multi-session touch gesture dataset which contains six mostly common used gestures. Then, three kinds of score metrics are compared and the $L_1$ distance is chosen. Finally, the experimental results on fusing all six gestures demonstrated the effectiveness of the proposed methods.

8. Acknowledgements

This research is supported by the Department of Homeland Security (DHS) under Award Number N66001-13-C-3002, and partially supported by the National Science Foundation under Award Number CNS 1205708. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of DHS or NSF.

References


