A Gaze Gesture-Based User Authentication System to Counter Shoulder-Surfing Attacks

Vijay Rajanna  
Sketch Recognition Lab  
Texas A&M University  
www.vijayrajanna.com  
vijay.drajanna@gmail.com

Paul Taele  
Sketch Recognition Lab  
Texas A&M University  
www.paultaele.com  
petaele@gmail.com

Seth Polsley  
Sketch Recognition Lab  
Texas A&M University  
spolsley@tamu.edu

Tracy Hammond  
Sketch Recognition Lab  
Texas A&M University  
faculty.cse.tamu.edu/hammond  
thammond@gmail.com

Abstract  
Shoulder-surfing is the act of spying on an authorized user of a computer system with the malicious intent of gaining unauthorized access. Current solutions to address shoulder-surfing such as graphical passwords, gaze input, tactile interfaces, and so on are limited by low accuracy, lack of precise gaze-input, and susceptibility to video analysis attack. We present an intelligent gaze gesture-based system that authenticates users from their unique gaze patterns onto moving geometric shapes. The system authenticates the user by comparing their scan-path with each shapes’ paths and recognizing the closest path. In a study with 15 users, authentication accuracy was found to be 99% with true calibration and 96% with disturbed calibration. Also, our system is 40% less susceptible and nearly nine times more time-consuming to video analysis attacks compared to a gaze- and PIN-based authentication system.

Author Keywords  
Gaze authentication; Gaze gestures; Pattern matching.

ACM Classification Keywords  
K.6.5 [Management of computing and information systems]: Security and protection; H.5.2 [Information interfaces and presentation]: User interfaces
Introduction

Shoulder-surfing is a significant issue for user authentication, due to its nature of attackers looking over a victim's shoulder to extract confidential information, and continues to be a growing problem [1, 10, 27]. These attacks are not only prevalent in crowded places [16, 18, 25], but can be further exploited with vision-enhancing devices (e.g., long-range binoculars). Previous works have tried to address shoulder-surfing with approaches like graphical passwords [12, 28], PIN entry methods through cognitive trapdoor games [23], PIN entry method based on vibration and visual information [17], and gaze-assisted authentication [2, 4, 6, 16]. However, most of these solutions rely on gaze input for PIN entry, fixation on certain points on an image, or making specific gestures; also, they need decent calibration for precise gaze input [3, 7, 11, 16].

Figure 1: Gaze Gesture-Based Authentication System: A user is authenticating by following the three shape gaze password. [1 - Camera, 2 - Authentication interface, 3 - Eye tracker].

Figure 2: Gaze gesture-based authentication interface with 36 shapes. Each shape has a fixed starting and ending points, and traverses along a predefined path.

We present a gaze-based user authentication system that combines gaze with gesture recognition. The interface comprises of 36 moving shapes (Figure 2), and to authenticate, the user has to follow three shapes, one on each frame, on three consecutive frames. A frame is a five-second duration where all the shapes simultaneously move from their source location to destination location. Three secretly selected shapes constitute a user's password. For successful authentication, the scan-paths of the user's gaze should match with the traversed paths of the correct shapes in the three frames. Also, of the 36 shapes only 12 shapes can be selected for a password, and the remaining 24 are fake shapes. Our approach is similar to the idea of pursuit-based authentication presented by Vidal et al. [24]. However, we use gesture recognition principles for scan-path matching, and this method supports high accuracy even with a large number of shapes and complex traversal paths. In addition, fake shapes introduce randomness to frustrate potential attackers from unauthorized access through guess work.

Prior Work

Kumar et al. [16], presented “EyePassword,” a system that can mitigate shoulder-surfing by using gaze-based input methods. The results show that gaze-based password entry performs as efficiently as keyboard-based input. Bulling et al. [3], presented a novel gaze-based authentication system that makes use of cued-recall graphical passwords on a single image. Luca et al. [6], presented an authentication system that is used in public terminals. The authors use an authentication method “Eye-Pass-Shapes,” that uses eye gestures to significantly increase security while being easy to use. Best et al. [2], presented a rotary dial for gaze-based PIN entry that eliminates dwell time. The solution relies on a weighted voting scheme of numerals whose boundaries are crossed by the streaming gaze points. Khamis et al. [15], presented “GazeTouchPass,” a multimodal gaze-and touch-based authentication system for mobiles devices. Cymek et al. [4], presented an authentication method, similar to [24], where the user follows the digits moving in vertical and horizontal directions to authenticate.

The previous systems are limited by low accuracy even with true calibration. In [8] that evaluates 3 well-known gaze-authentication methods, including [6], the least error was 9.5% and the highest error was 23.8%. Also in [2], the authentication accuracy was 71.16% (PIN interface) and 64.20% (Rotatory interface). Furthermore, gaze-authentication is susceptible to video analysis attacks [3, 6]. Though the work presented in [4] recognized 97.57% of the digits entered, the authors did not study password recovery through video analysis attacks, this is important since the digits move in vertical and horizontal trajectories only. Lastly, solutions like [3, 6] need the user to remember the gestures or multiple locations on an image, which may become difficult with a large number of passwords. We will discuss our goals in the hypotheses section.
Design Motivation
Prior research by Wendy et al. [19], Hoanca et al. [13], Davis et al. [5], has shown that graphical passwords such as static images or user-drawn gestures are easier to remember than PIN passwords. User-drawn passwords involve two processes: 1) visual recall of the drawn password, and 2) recall of the temporal order [20]. Primarily, we wanted to liberate the user from remembering complex gestures and the order of strokes that constitute a gesture. Thus, we made the interface such that the user is only required to remember the shapes that constitute the password but not required to remember the gestures and their constituent strokes.

Hypotheses
Considering the limitations with prior research and design motivations for our system, we form the following hypotheses: a) the gaze gesture-based authentication system achieves high accuracy and is robust to calibration errors, b) users commit fewer or no errors when entering passwords with successively repeated shapes (like pie, pie, circle) on our system, and c) our system is less susceptible and more time consuming to video analysis attacks than gaze- and PIN-based password entry systems.

System Architecture and Implementation
The gaze gesture-based authentication system (Figure 1) consists of two main modules: 1) Gaze Tracking Module, and 2) Authentication Engine.

Gaze Tracking Module: This module uses a table-mounted “The Eye Tribe” eye tracker that provides (X,Y) gaze coordinates. We position users at 45-75 cm in front of the monitor and the eye tracker error is $0.5^\circ$-$1^\circ$ of visual angle.

Authentication Engine: This is the central module that runs on the computer and receives the eye tracker’s gaze coordinates. Primarily, it implements the scan-path matching algorithm to authenticate the user.

Authentication Procedure
Password Selection
To choose a password, a user selects three shapes from a password selection interface that lists the 12 true shapes. The first shape selected is followed on the first frame, the second on the next frame, and so on.

Authentication Interface
The authentication interface is shown in Figure 2. The interface is a canvas with 36 shapes placed at different locations on the screen: 12 are true shapes available for password selection, and the remaining 24 are fake shapes not considered during password selection. Each shape is assigned a predefined starting and ending points, and a path along which it traverses. Hence, the user is not required to search for password shapes once their initial locations are known.

For each true shape, there are two fake shapes placed at different quadrants on the screen which perform similar transitions as the true shapes. We introduced fake shapes for two reasons: 1) in brute force attacks, an attacker without knowledge of the fake shapes must assume a password complexity of $36 \times 36 \times 36 = 46,656$, whereas the true complexity is $12 \times 12 \times 12 = 1,728$, and 2) in video analysis attacks, fake shapes introduce enough randomness in the system that it becomes hard or time-consuming to recognize the exact shape through guesswork.

Authentication in Action
To control the interface, the user presses a set of hot-keys: ‘A’ to initiate movement of shapes and record gaze data, ‘Z’ to recover from user mistakes (blink, sneeze, losing the path) and discard recorded gaze data, and ‘M’ to submit the password after following 3 shapes. We minimize authen-
tication failures since users have direct control over each frame. For example, if a user selected Square-Star-Pie as a password, then the user is authenticated by following each shapes’ paths in their respective frames, as shown in the sequence of Figures 3, 4, 5. The user does not receive any feedback, since the gaze point and scan-path are hidden.

**Recognition System**

We match the user’s scan-path against a shape’s traversed path through "Template Matching" algorithm, where we compute the root-mean-square distance of the candidate path (user’s scan-path) from all the template paths (shapes’ traversed paths). The template path of a shape that is at least distance from the candidate path is chosen as the shape followed by the user. Our template matching algorithm is similar to $1 \ [26]$, but we perform only sampling, and calculate the average distance between the two paths.

**Scan-Path Matching and Authentication**

The template matching algorithm first samples the input scan-path to $N = 64$ points as depicted in Figure 6. We chose $N=64$, empirically derived considering the eye tracking frequency of 60Hz we used. To compute the average distance between a candidate path and a template path, as shown in Figure 7, we use equation 1, where $P$ is a $(X,Y)$ point on a path, $C$ - candidate path, $T$ - template path, and $\Delta DT$ - average distance to template.

$$\Delta DT = \frac{\sum_{p=1}^{N} \sqrt{(C[p]_x - T[p]_x)^2 + (C[p]_y - T[p]_y)^2}}{N}$$

**Template Construction**

Our system was trained from traversed paths generated by seven users. First, each user generated paths for 12 shapes that are used as templates in the recognition phase, where the user again followed each of the shapes and the system recognizes the shape followed. For users who achieved more than 90% accuracy, their templates were retained. We repeatedly added and tested new paths until our final system achieved 100% accuracy from paths created by four of those users. Since eye movements involve fixations, saccades, and regressions $[9, 22]$, we generate template paths against which the user’s scan-path is matched, instead of using line paths of the shapes.

**Experiment Design and Results**

We tested the system in two phases.

**PHASE 1: System Accuracy and Robustness**

We recruited 15 participants (12 males and 3 females), some used vision correction devices like glasses and contact lens. All were either graduate or undergraduate students, with ages varying between 20 and 26 ($\mu_{age} = 22.53$). Before each study, the participant was briefed about the idea of gaze- and sketch-based authentication, given a small demo of the working system, and calibrated with the eye tracker on a $1900 \times 1200$ monitor.

**Part 1: Scan-Path Recognition Accuracy**

The goal of this study was to determine the recognition accuracy of the user’s scan-path against the actual path of the shape. Hence, the user follows all 12 true shapes, one on each frame. After the completion of each frame, the system recognizes the shape followed by the user, and the shape’s name is displayed through a pop-up message. In this phase, a small circle moves on the screen reflecting the user’s gaze-point on the screen, and the user’s scan-path is drawn as the gaze moves. This feedback (scan-path) was enabled to verify true positives, i.e., the path followed by the user for a given shape. However, no trial was repeated if
the user didn’t follow the true path resulting in recognition failure, as such errors may occur in real-world scenarios. Table 1 shows the confusion matrix for all the true shapes. We achieved a scan-path recognition accuracy of 99.44% at an F-measure of 0.99.

**Table 1**: Scan-Path Recognition - Confusion Matrix. Key: A - Circle, B - Open Hexagon, C - Triangle, D - Pie, E - Square, F - eye, G - Open Square, H - Ring, I - Star, J - Open pentagon, K - Pentagon, L - Hexagon

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.0</td>
<td>0.93</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.93</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0.07</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Part 2: Authentication Accuracy With True Calibration**

In this part, the user was allowed to choose a password, by selecting three shapes from the password selection window. After selection, the user follows those shapes, one on each frame, but no feedback (scan-path) was shown. Providing no feedback simulates the real-world scenario, as feedback would enable shoulder-surfing attacks. To authenticate, the user should get all the three shapes correct. The user repeated this authentication procedure for three different passwords. We also recorded a video of the user’s eye movements, while entering a password, to use in the comparative study. Lastly, to test the system’s ability to invalidate wrong passwords, the experiment facilitator sets a different password (unknown to user), and the participant attempts to access the system by guessing the password; this was also repeated for three different passwords. This is similar to testing the system with true negatives. We achieved an authentication accuracy of 99%, and the confusion matrix is shown in Table 2.

**Table 2**: True Calibration: Confusion Matrix, Authentication Accuracy, and F-Measure

<table>
<thead>
<tr>
<th></th>
<th>True Password</th>
<th>False Password</th>
<th>Accuracy</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Password</td>
<td>97%</td>
<td>3%</td>
<td>99%</td>
<td>0.99</td>
</tr>
</tbody>
</table>

**Part 3: Authentication Accuracy with Disturbed Calibration**

To test robustness to calibration errors, the user was asked to get up and walk around for a few minutes. Upon return, the eye tracker was not re-calibrated, leaving the authentication system susceptible to calibration errors. Similar to part 2 of the study, the participant chooses three new passwords and enters them on three different trials. Again, the facilitator sets three new passwords, and the participant tries to access the system by guessing the passwords on three different trials, to test true negatives. We achieved an authentication accuracy of 96%, and the confusion matrix is shown in Table 3.

**Table 3**: Disturbed Calibration: Confusion Matrix, Authentication Accuracy, and F-Measure

<table>
<thead>
<tr>
<th></th>
<th>True Password</th>
<th>False Password</th>
<th>Accuracy</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Password</td>
<td>92%</td>
<td>8%</td>
<td>96%</td>
<td>0.96</td>
</tr>
</tbody>
</table>

**PHASE 2: Robustness against Hacking**

Through a preliminary study, similar to previous studies [3, 6], we tested the susceptibility of our system to video analysis attacks in comparison to a gaze- and PIN-based password system. During phase 1, we recorded the videos of participants entering passwords (Figure 8) on both our system and a gaze- and PIN-based system (Figure 10) that
used dwell-based selection. Four users, as shown in Figure 9, analyzed videos, chosen randomly, of the participants entering passwords. We found that gaze- and sketch-based authentication system was 40% less susceptible to video analysis attacks, and it took significantly more time—nearly 9 times longer—to guess the password on our system compared to gaze- and PIN-based authentication system. Users cracked 3/5 shape and 5/5 pin passwords in 4183 and 478 seconds respectively.

Qualitative Evaluation

Positives: P02: “The nature of the actual password entry is really good... I can’t imagine a better way.”
P05: “I liked that it’s simple using shapes and that it’s very secure that no one can really track your eye movements... It’s very innovative.”
P12: “I like that you can use this for authentication and other people can’t really tell that what you selected as your password, because they can’t really follow your eyes.”
P09: “This is definitely something new... It addresses the problem statement pretty nicely.”
P07: “It was pretty impressive, it was able to distinguish between all those shapes when they were colliding with each other... I would use it for everyday jobs.”

Suggestions: P09: “Someone sneezing could be an issue... they have to do the password all over again.”
P05: “The problem would be if people don’t have concentrated eyes.”
P15: “Adjust the eye tracker for different postures.”

Discussion

In testing our hypotheses from our user studies, we first correctly hypothesized high accuracy for scan-path matching and the authentication system with true calibration. Also, the accuracy remained high even when the calibration was disturbed. We attribute high accuracy to relaxed precision on gaze input and unique paths for each shape traversal. However, we anticipate that multiple shapes with similar paths would reduce accuracy. Next, since the user follows a single shape in each frame, we found that the participants had no difficulty in entering a password with repeated shapes. Finally, results from video analysis attacks showed the advantage of fake shapes: although an attacker can guess the direction of a shape’s movement from the user’s eyes, they cannot pick the right shape from numerous options before the system locks out from failed attempts. From the interviews (side-table), we found that the users consider this solution innovative, secure, and simple. However, some expressed that sneezing, lack of attention during password entry, and so on would lead to incorrect gaze input.

While we expected 100% accuracy, we encountered two sources of scan-path distortion that affected accuracy. First, although our system authenticates with five-second shape movements compared to other gaze authentication systems that take from 7.5 seconds [2] to 54.0 seconds [8], users may blink during the shape’s five-second movement and suggested reducing movement to 3 seconds. Hence, we hypothesize that reducing the overall authentication time to less than 10 seconds avoids authentication failure due to erroneous gaze input. Second, the use of vision correction devices lead to imprecise gaze input [14, 21].

Future Work

From our current work, we have identified several potential next steps for improvements and extensions. One next step is to supplement our strong authentication accuracy with further reducing authentication time as elaborated in similar works [2, 8, 16]. Another next step is to investigate users’ requests, from our conducted interviews, for reduced attention time through shorter authentication time, similar to [6, 16]. We also realize that employing fake shapes does not always prevent but instead delays brute force attacks, so we will be investigating additional solutions such as randomizing shape traversal paths, and providing each user with the ability to configure their choice of true and fake shapes. Furthermore, we will be extending this work as an accessible authentication system for users with physical impairments, who cannot use input devices like a keyboard. Lastly, we would like to further scale our system to accommodate smaller form-factor devices than the current desktop setting.

Conclusion

We presented a gaze gesture-based authentication system to counter shoulder-surfing attacks. The interface consists of 36 shapes that move simultaneously on the screen and the user follows three shapes to authenticate, while an eye tracker tracks the user’s gaze. We found that our system can authenticate with over 99% accuracy, and 40% less susceptible and nearly nine times more time-consuming to video analysis attacks compared to existing systems.
References


