EyeVeri: A Secure and Usable Approach for Smartphone User Authentication

Chen Song, Aosen Wang, Kui Ren, Wenyao Xu
Department of Computer Science and Engineering, SUNY at Buffalo, Buffalo, NY, USA
Email: {csong5, aosenwan, kuiren, wenyaoxu}@buffalo.edu

***Abstract*—As mobile technology grows rapidly, the smartphone has become indispensable for transmitting private user data, storing the sensitive corporate files, and conducting secure payment transactions. However, with mobile security research lagging, smartphones are extremely vulnerable to unauthenticated access. In this paper, we present, *EyeVeri*, a novel eye-movement-based authentication system for smartphone security protection. Specifically, *EyeVeri* tracks human eye movement through the built-in front camera and applies the signal processing and pattern matching techniques to explore volitional and non-volitional gaze patterns for access authentication. Through a comprehensive user study, *EyeVeri* performs well and is a promising approach for smartphone user authentication. We also discuss the evaluation results in-depth and analyze opportunities for future work.**

I. Introduction

Smartphones have overtaken personal computers (PC) and become the most prevalent devices for communications and computing services [1] in daily life. By 2015, there were 1.5 billion smartphones in use globally [2]. In contrast, the mobile security for access control and data privacy has been overlooked when compared to PC security and severely lags behind in spite of the ubiquitous nature of smartphones. It is not surprising that smartphone users experience more unauthenticated access than PC users as reported in [3], because smartphones intrinsically tend to have higher risk of loss or theft [4].

Until recently, biometrics-based authentication has attracted more attentions and becomes an alternative to traditional authentication methods like passwords or PINs. In general, it can be categorized as two types: Physiological and Behavioral Biometrics.

Physiological biometrics include fingerprint [5], facial recognition [6], speech analysis [7], [8], and iris scans [7], [8]. These methods require users to pass the biometric verification in order to obtain the access authority to smartphones. The potential risk is that most of these bio-features can possibly be obtained or replicated by adversaries. For example, fingerprint can be stolen and voice pattern can be counterfeited by professional software. Even some facial recognition systems can be fooled by an appropriately sized photo of a legitimate user. *Behavioral biometrics* are based on the way people do things. In the past few years, researchers have explored various touch-based behavioral cues to provide security protection on smartphones such as keystroke patterns [9] and touch gestures [10]. However, these methods require interaction with the screen pad, which means that the "password" can be threatened by shoulder-surfing. Other methods such as gait patterns [11] and in-air signatures [12] need obtrusive interaction with smartphones and are still possible for adversaries to mimic. In conclusion, a secure biometric can not be explicitly obtainable in public or based on easy-controllable body component.

Human eye movement is driven by the complex interaction network between brainstem control and extraocular muscles via neurological components [13]. Due to the fact that it compiles both physiological and behavioral aspects in nature and has miniature scale, human eye movement is highly immune to the accurate replication by the adversaries. In this paper, we present *EyeVeri*, a novel authentication system for smartphones based on eye movement. It captures human eye movements (i.e., fixations and saccades) and extracts a unique gaze pattern for access authentication. Firstly, eye-movement pattern is related to eye bio-structure and each individual has a unique extraocular muscle condition. Specific eye structure or muscle-related features such as range of view or eye-movement speed are highly individual-dependent. Secondly, eye movements usually take place in response to specific mental processes of human beings. Studies have discovered the close relationship between eye behaviors and human emotions [14] and the *Eye-Mind Hypothesis* states that there is a direct correspondence between a person's gaze and attention. Therefore, individuals hold unique eye-behavioral features in eye movements because of their different experiences and habits.

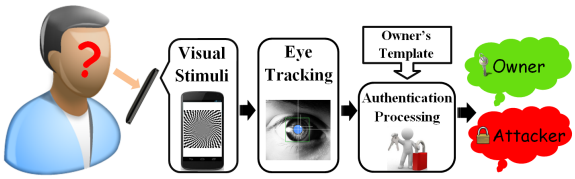


Fig. 1: *EyeVeri* flowchart illustrates the whole process from the time when an unknown user requests to enter the smartphone to the time that the system makes a final decision.

In our work, we develop an authentication system (Fig. 1) and implement the eye-tracking module on an android smartphone. Generally, eye movement can be categorized into volitional and non-volitional movement. Volitional movement is when people scan objects on their initiative and non-volitional one is when people passively observe objects. We design and examine four types of visual stimuli to stimulate various eye-related physiological and behavioral features. We

978-1-4673-9953-1/16/$31.00 ©2016 IEEE

comprehensively evaluate the system performance, mainly from three aspects: accuracy, efficiency as well as long-term stability. Our system can achieve accuracy of around 88% during the experiments, as well as sufficient time efficiency ($5 \sim 10$ seconds) and long-term stability (around 5% fluctuation). Our study proves the concept of smartphone authentication based on eye movement is feasible and promising. Generally, *EyeVeri* has the following main advantages as an authentication approach: 1) *Secure*: it enables the user to avoid conspicuous interaction with the smartphones and against shoulder-surfing or smudge traces; 2) *Unique*: it stimulates the highly individual-dependent features of extraocular bio-structure and eye-movement behaviors; 3) *Constant*: it obtains a fairly stable performance as shown in our long-term study, which enhances its feasibility. To the best of our knowledge, this is the first work to explore both volitional and non-volitional eye movement as a biometrics on the smarphone.

Our main contributions can be summarized into three-fold:

- We propose and implement an end-to-end authentication system based on eye movement on the smartphone without any extra hardware;
- We explore a set of visual stimuli which can stimulate eye-movement features from different aspects and show the promising result of applying non-volitional eye response as a biometrics;
- We perform a complete and intensive user study to understand in-depth the system performance and limitation.

II. Background

A. Related Work of Smartphone Authentication

Authentication methods applying eye-tracking technology have been investigated. Luca *et al.* [15] presented an authentication mechanism based on eye gesture to eliminate the physical contact-type on the public terminals. It enabled users to invoke commands by moving their eyes in a pre-defined pattern. The user's eye movements were categorized into eight strokes. Dachuan *et al.* [16] proposed a new eye-tracking method for smartphone authentication. Objects with different numbers randomly moved on the screen. By matching the user's eye-movement trajectories with the objects', the system determined the actual number the user was looking at. Note that the main principle of these works (including [17] [18]) is to utilize eye movement as an alternative input method to pointing-based interaction, instead of exploring the biometric characteristic it contains.

Single or multiple finger-touch based smartphone authentication methods have been explored [19][10][20]. Shahzad *et al.* [21] designed a set of sliding gestures performed by single or multiple fingers. They trained a model based on the behavioral-related features extracted from user's gestures. The result showed an average equal error rate of 0.5% with 3 gestures using 25 training samples for each gesture.

Other biometrics have also been studied. Das *et al.* [22] proposed an autobiographical authentication method based on people's own daily memories. They built a challenge-response authentication system that queries users about the daily experiences the system recorded. Schneegass *et al.* [23] presented an enhanced authentication system that applies random geometric image transformations to increase the security of cued-recall graphical passwords against smudge trace. They applied transformations such as rotation, scaling shearing and flipping to the image before the user input the password pattern. However, this method is vulnerable to threats such as shoulder surfing. Nickel *et al.* [24] carried out smartphone authentication research based on the way people walk. The gait biometric data was collected through the accelerometer by attaching the smartphone to a specific body part. Based on the feature extraction and k-NN algorithm, they achieved a low EER of around 8%. The authentication was conducted based on 30 seconds of walking data.

B. Human Vision and Eye Movement

The human visual system is mainly composed of the oculomotor plant and brainstem control [25]. The oculomotor plant primarily comprises the eye globe, surrounding tissue, and the extraocular muscles. The extraocular muscles can be categorized into the lateral and medial recti, the superior and inferior recti, as well as the superior and inferior obliques. The human brain controls the extraocular muscles with burst/omnipause neurons, causing muscle contractions and relaxations [26].

Various human eye movements are highly related to the neuronal signals [27]. Among these different types, fixations and saccades are of particular interest to the researches. Fixations occur when the eyes focus on a specific point such that the fovea remains centered on the object to ensure the visual acuity. Saccades are the rapid movements of the eye between two fixations, with very little visual acuity maintained during rotation [26]. The scanpath is the spatial path generated by the fixations and the saccades. Under the same condition, it is almost impossible for two individuals to achieve the same scanpath since no two persons are exactly the same in the bio-structure. Based on existing eye-movement theories [27], it provides a potential way to distinguish individuals with specific eye-behavioral patterns.

Eye-movement biometrics for human identification were initially investigated by Kasprowski and Ober [28]. They performed personal identification based on eye movement characteristics, using specific eye-tracking equipment. The experiments proved the possibility of identifying people with this unique bio-feature. Bednarik *et al.* [29] presented a case study investigating the potentials of eye-movement data for biometric purposes. The results indicated that eye-movements and gaze features contain the discriminatory information between individuals. Later on, Komogortsev *et al.* carried out intensive studies [30][31][32] on eye-movement-based human identification, including the effects of eye stimulus types and movement patterns on the accuracy of biometric verification. The results verified that certain feature extractions of fixations and saccades are able to accurately distinguish individuals.

Even though eye movement contains so many unique features of individuals, to the best of our knowledge, so far there

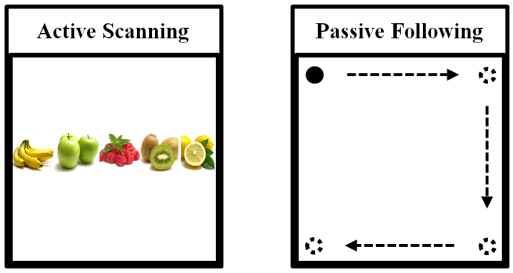
is no work to explore them in smartphone user authentication, especially non-volitional eye movement.

III. *EyeVeri*

A. System Overview

The overall framework of *EyeVeri* is shown in Fig. 2, which comprises three modules: (1) Visual Stimuli Design, (2) Eye-tracking System and (3) Authentication Processing. When someone attempts to gain access to a smartphone, the pre-designed visual stimuli are shown on the screen. The eye-tracking system that runs in the platform background simultaneously captures the eye movement of the subject and records the position information of the focus point. After the user's data are recorded, certain features are extracted from the data and a classifier based on the pre-stored owner template processes the in-coming data to determine whether the user is the owner or not.

B. Visual Stimuli Design



(a) Fruit-Row (FR); (b) Corner-Gif (CG); (c) Illusion-Image (II); (d) Simple-Dot (SD).

Fig. 3: Four different types of visual stimuli.

In Fig. 3, we design *four* visual stimuli for different eye-related features, which are Fruit-Row (active scanning), Corner-Gif (passive following), Illusion-Image (strong stimulation) and Simple-Dot (weak stimulation). For the simplicity of presentation, we use abbreviations FR, CG, II and SD for each of them in the rest of the paper. FR (Fig. 3(a)) contains a fruit sequence in a row and the subject actively scans through the sequence from left to right at his/her own habit. There is no restriction in both time and manner. CG (Fig. 3(b)) includes a gif where a black circle rotates through the four corners in a clockwise way. The subject needs to exactly follow the black circle during the authentication process. II (Fig. 3(d)) is a typical illusion image that can strongly stimulate the unconscious vibration (saccade) of the eyeball. Lastly, SD (Fig. 3(c)) has a simple dot positioned in the middle of the screen and the subject also needs to stare at the dot during the process. When watching these stimuli, the subject should keep head movement to a minimum relative to the position of the face and the smartphone.

From the perspective of the design purpose, each type of visual stimuli contains the disparate information that we are interested in. FR reflects the volitional eye response and highly depends on the reading or scanning pattern and personal favor with regard to the specific fruit type in the sequence. For example, some people have faster scan speed and may spend more time on favored objects. The remaining three types stimulate non-volitional eye response. CG contains a large amount of information about the bio-structure of the eye and the angular size of the subject which is unique in everyone. Both II and SD tend to stimulate the unconscious eye vibration of the individual. Compared with SD, II is prone to excite eyeball vibration and augment personal uniqueness.

C. Eye-tracking System

The eye-tracking system mainly involves two steps: Facial Info Processing and Gaze-angle Calculation. Specifically, Facial Info Processing extracts spatial information of the eyes from the camera preview. Gaze-angle Calculation computes the spatial angles related to the gaze point.

1) Facial Info Processing: This step is to achieve positional information of the eyeball and iris. Based on the camera preview image, the system first detects the face position using the Six-Segmented Rectangular (SSR) filter [33]. After obtaining the face position info, the accurate location of both eyes are extracted based on the eye's regions of interest (ROI) on the image, using a shape-based approach. Then edge detection techniques such as Hough transform are used to effectively figure out the iris contours [34]. The iris contour is nearly a circle which lays on the surface of the eyeball, and the projection of the iris contour on the camera image is an ellipse while the gaze is deviated. Some parts of the iris contour are shadowed by the eyelid. Therefore, the ellipse fitting is implemented to achieve accurate iris contours.

2) Gaze-angle Calculation: The eyeball model [35] is assumed to be a sphere and the iris lays on the surface of the eyeball. The optical/visual axis (we presume they coincide with each other) of the eye is the line passing through the center of the eyeball and the center of the iris. The anatomical axis of the eye is the line passing through the center of the eye ball and is normal to the screen plane. The angle between these two axes is defined as the eye gaze. While changing the eye gaze, the eyeball rotates around its center. The radius of the iris is modeled to be a constant, even though it varies a little bit among users. The two angles that we need to estimate the position of the gaze point are the horizontal and vertical angle, which lay in a gaze-horizontal and gaze-vertical plane, respectively. Gaze-horizontal plane is the spatial plane that passes the horizontal lane of the gaze point and the center of the eyeball. The horizontal angle is between the optical axis and the anatormical axis in the gaze-horizontal plane.

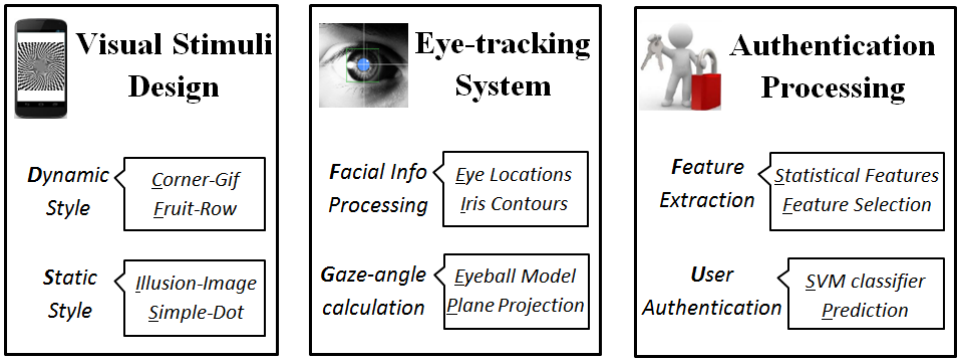


Fig. 2: The proposed smartphone authentication framework based on unobtrusive eye-movement, which comprises three modules: (1) Visual Stimuli Design; (2) Eye-tracking System; (3) Authentication Processing.

Similarly, the vertical angle is between the optical axis and the anatormical axis in the gaze-vertical plane. The two angles are directly related to the gaze point on screen. In this work, we use available techniques [36] to implement our eye-tracking system on the smartphone.

D. Authentication Processing

After the gaze data are collected, the authentication process is conducted to verify if the user is the legitimate owner.

1) Feature Extraction: We propose and develop a set of eye-movement features which contain physiological and behavioral information. The features and corresponding definitions are listed in Table I. Specifically, we categorize them into three groups, based on the main information they contain.

- ***Physiological Info.***: *Max.* and *Min.* of the gaze angles in coordinates depend on the view angle, which contains the typically physiological features of the eye. *Max.* angle value refers to the rightmost point (horizontally) or the topmost point (vertically) a subject can reach. Similarly, *Min.* angle value means the leftmost point (horizontally) or the bottommost point (vertically) of a person's view.
- ***Behavioral Info.***: *STD* as well as *RMS* represent the distribution of the scan area in coordinates, which are related to eye behavior. *Skewness* and *Kurtosis* are behavioral-based features. *Skewness* is a measure of the asymmetry degree and may has limited contribution in a specific direction because of the symmetry layout of visual stimulus, such as CG or FR. *Kurtosis* is a measure of whether the signal is peaked or flat relative to the normal distribution. *Iqr.* describes the signal statistical dispersion. Both *ZC* and *MC* reflect the shift frequency of the eye movement, which are mainly behavior-related. *Corr.* between two coordinates helps differentiate movements that involve translation in single dimension from the ones that involve translation in multi-dimension. CG involves one direction movement in each step. The related movement for FR is also most likely in one direction since the fruit sequence is in a row. However for SD and II, *Corr.* may have limited effect in that the unconscious vibrations are multi-direction.
- ***Physiological & Behavioral Info.***: *Mean* and *Median* can indicate the general focusing area of the eye on the screen, which contain both aspects. *N-order Derivatives* are associated with how quickly the eye moves in coordinates, which are also determined by the muscle feature of the eye. Therefore, both physiological and behavioral information are involved.

Note that all features except *Corr.* are independently applied on horizontal and vertical angles. Therefore we will eventually have 27 features in total.

2) Authentication Algorithm: Authentication is intrinsically a one-class classification problem which differentiates between data that appear normal and abnormal with respect to the distribution of the training data. The main challenge is to find an appropriate distribution for the data and the performance of the model is high related to the locality and completeness of the data. As a result, it is widely reformulated into a two-class classification problem in the biometric authentication field [20], [37]. Initially, the owner's templates are stored in the smartphone. When an access request occurs, the gaze data from the unknown user are collected. For evaluation proposes, we employ support vector machines (SVM) as our classifier. More specifically, we use the Sequential Minimal Optimization implementation of SVM which is provided in the Weka machine learning toolkit [38]. Gaussian radial basis function is selected as the kernel function to map the original data to a higher dimensional space.

IV. System Evaluation

In order to evaluate the system in a comprehensive way, we focus on four main aspects: system accuracy, time efficiency, long-term performance and feature dimension reduction.

We develop the system on Google Nexus 4 with a quad-core CPU of 1.5GHz and a screen size of 4.7 inches. The visual stimuli we use are CG, FR, II and SD. The sampling rate of the eye tracking module is 5Hz, which means there are 50 data samples in 10sec.

A. Evaluation of System Accuracy

A key concern to the feasibility and effectiveness of an authentication system is quantification of the degree to which

No.	Feature List	Feature Definition	Main Contained Info.
1	Maximum (Max.)	The maximum value of the signal over the window	Physiological
2	Minimum (Min.)	The minimum value of the signal over the window	Physiological
3	Standard Deviation (STD)	The measurement of the distribution of the signal	Behavioral
4	Root Mean Square (RMS)	The quadratic mean value of the signal	Behavioral
5	Skewness	The degree of asymmetry of the signal distribution	Behavioral
6	Kurtosis	The degree of peakedness of the signal distribution	Behavioral
7	Interquartile Range (Iqr.)	The difference between 75th & 25th percentiles of the signal over the window	Behavioral
8	Zero Crossing Rate (ZCR)	# of changes between positive & negative	Behavioral
9	Mean Crossing Rate (MCR)	# of changes between below mean & above mean	Behavioral
10	Pairwise Correlation (Corr.)	Correlation between the horizontal & vertical signals	Behavioral
11	Arithmetic Mean (Mean)	The average value of the signal	Physiological & Behavioral
12	Median	The median value of the signal	Physiological & Behavioral
13	Mean 1st Derivatives	The average of 1st order derivatives over the window	Physiological & Behavioral
14	Mean 2nd Derivatives	The average of 2nd order derivatives over the window	Physiological & Behavioral

TABLE I: Extracted features and their definitions, as well as the main information each of them contains.

it can accurately recognize the owner and reject adversaries.

1) Evaluation Descriptions: We recruit a total of 20 participants (3 females and 17 males) in our experiment. Among them, 5 use glasses while the others do not. Their ages are in the range of 25-35. After we train the SVM classifier based on the validation set, the 10-folder cross-validation is conducted to give insight to the overall performance of the classifier.

We refer to one trial as the subject watches four visual stimuli one after each other. Each subject repeats 10 trials in the experiment. Therefore, each subject eventually has 40 collected data (10 for each visual stimulus respectively) and in total we have 800 sample data.

2) Single-user Application Scenario: Considering the smartphone is owned by one individual, we evaluate the system in the single-user, multi-attackers scenario. The attacker scenario is simulated where the owner loses the smartphone and unknown people attempt to enter the smartphone. Each of the 20 subjects is selected exactly once as the owner and remaining subjects' data are used as the adversaries to attack the system. For the two-classification module, we label the owner's data as the positive class, while all other subjects' data (except the one from the attacker) as the negative class. Based on the 10-folder cross-validation, the data set in each class will be randomly divided into the training set and test set. The owner's test set is also included since it will not make sense if the training model rejects all the data.

3) Evaluation Results:

a) Balanced Accuracy Metric: The most straightforward and widely used accuracy metric is defined as:

$$Accuracy(\%) = \frac{TP + TN}{TP + FP + TN + FN} * 100\%, \quad (1)$$

where TP is the true positive, TN is the true negative, FP is the false positive and FN is the false negative. However, this metric can be misleading when the true class distribution is unbalanced. In our case, the sample ratio of the positive class and the negative class is 1:19, which means that we can still have an accuracy as high as 95% even if the model naively predicts all the test data as negative (rejects all the users including the owner and has no practical meaning). In our work, we adopt the balanced accuracy metric (BAC), given its advantage of non-sensitivity to class distribution:

$$BAC(\%) = \frac{0.5 * TP}{TP + FN} + \frac{0.5 * TN}{TN + FP}. \quad (2)$$

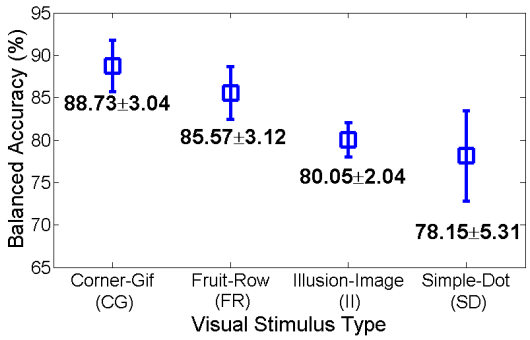


Fig. 4: The average BAC of 20 subjects for four visual stimuli. The error-bars are the STDs of BAC among the subjects.

Fig. 4 shows the average BAC of 20 subjects for four visual stimuli. CG achieves the best accuracy of 88.73%, with the standard deviation (STD) of 3.04%. FR obtains a close accuracy of 85.57%, with the STD of 3.12%. The results prove that the related bio-info (such as angular size) in CG as well as the eye-behavior pattern (such as scan speed) in FR are heavily individual-dependent and are able to distinguish different people. II and SD have the accuracy results of 80.05% and 78.15% respectively. These two images intend to discover the unconscious eye vibration pattern of the individual. However, II more easily stimulates unconscious eye vibration while the corresponding response time varies for SD. In other words, data in II contain more eye vibration information than SD during the experiment. Therefore, it achieves better accuracy.

Moreover, it is worth mentioning that the worst case of CG (85.69%) still performs better than the best cases of II (82.09%) and SD (83.46%). The worst case of FR (82.45%) has close performance with the best cases of II and SD.

b) Receiver operating characteristic (ROC): To take a closer look at the system accuracy under different setups, we investigate ROC curves among four visual stimuli in the study. ROC curve is an effective way to graphically reflect and compare the performance of the classifiers among different

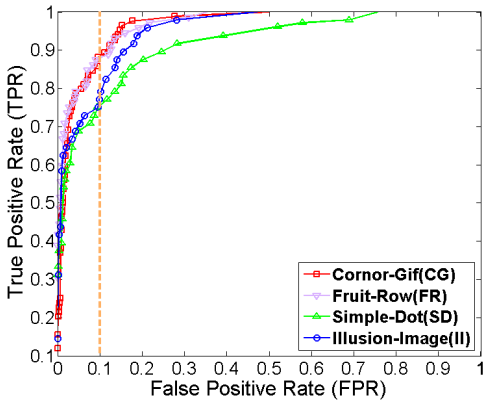


Fig. 5: The average ROC curves of 20 subjects for four visual stimuli. The vertical dashed line in orange implies a 10% threshold of false positive rate.

classification setups. The true positive rate (TPR) and the false positive rate (FPR) are traded off against each other. If the classification is carried out in a strict setup, both TPR and FPR can be extremely low, which means attackers will be rejected as well as the owner. At the cost of wrongly accepting some attackers, the classifier is less sensitive with reasonable TPR.

Fig. 5 displays the ROC curves of four visual stimuli, based on the average result of 20 subjects. The AUCs (the area under the curve) are 96.74%, 95.38%, 93.62% and 90.11%, respectively for CG, FR, II and SD. The performance of CG and FR is quite similar and their curves cross a little bit with each other over different setups. They are all better than II and SD because both curves are completely above the ones of II and SD, which is in accordance with our previous discussion. If we setup a false positive threshold of 10%, TPR of CG and FR achieves around 85% accuracy, while that of II and SD is approximately 80% and 75%, respectively.

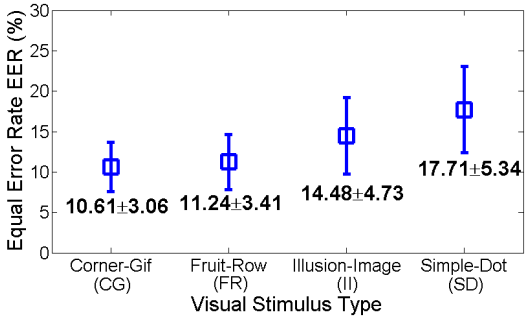


Fig. 6: Equal error rate (EER) of four visual stimuli. The error-bars are the STDs of EER among 20 subjects.

c) Equal Error Rate (EER): In order to account for the usability-security trade-off, we report EER, which is at the sensitivity of the classifier where FPR and FNR are equal. Fig. 6 depicts the outcomes of EER of four visual stimuli. The error-bars represent the STD of EER among 20 subjects. CG and FR have the similar EER of 10.61% and 11.24%. The corresponding small STDs, 3.06% and 3.41%, suggest the universality of these two types of authentication upon the subject group. II has a 14.48% EER with a STD of 4.73%, while SD has a 17.71% EER with a STD of 5.34%. The relatively high EER and STD for SD implies that it is not as distinguishable and stable as the other three visual stimuli.

B. Evaluation of Time Efficiency

Time efficiency is another important metric, especially for resource-constrained smartphones. An effective authentication system is supposed to not only correctly recognize valid access requests, but also efficiently make the authentication decision.

1) Evaluation Descriptions: To evaluate time efficiency, we apply different time restrictions on the visual stimuli. Since FR is designed with the principle of no constraint in manner or time, it will not be included in this evaluation. For the other three visual stimuli, 20 subjects repeat the same experiment process as is described in the preview section, but with four different duration setups: 3sec, 5sec, 7sec and 10sec. Note that for CG, the different authentication durations result in the different stay time of the circle in the corners.

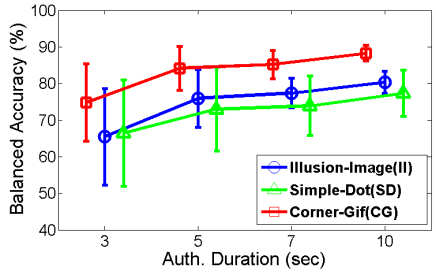


Fig. 7: The average BAC of visual stimuli under different authentication durations. The error bars are the STD of the accuracy results among 20 subjects in the corresponding authentication duration.[1]

Dur.	CG Acc.	Growth	II Acc.	Growth	SD Acc.	Growth
3sec	73.56%	-	65.36%	-	65.89%	-
5sec	84.63%	**15.04%**	75.65%	**15.74%**	72.16%	**9.51%**
7sec	86.02%	1.64%	78.02%	3.13%	74.34%	3.02%
10sec	88.73%	3.15%	80.05%	2.60%	78.15%	5.13%

TABLE II: The BAC and growth rate of 3 visual stimuli under different authentication durations. (Growth: The growth rate is calculated by the accuracy in the current duration and the previous duration.)

2) Evaluation Results: The average BAC results of three visual stimuli with different authentication durations are illustrated in Fig. 7 and the statistical results are summarized in Table II. Collectively, they provide the trade-off information between the accuracy and the duration of the three visual stimuli. We can see that the authentication duration of 3sec is too short for a reliable result for all visual stimuli, with expected low average accuracy, as well as high STDs (10.81% for CG, 12.85% for II and 14.97% for SD). The BAC results

[1]The difference in horizontal is for the purpose of illustration.

of three visual stimuli are all significantly improved when the duration increases from 3sec to 5sec. Specifically, the performances of CG, II and SD are improved by 15.04%, 15.74% and 9.51%. The corresponding STDs are also reduced to 5.82%, 8.12% and 10.53%, respectively. Also, there are no significant improvements when the time increases from 5sec to 7sec. However, when the duration increases from 7sec to 10sec, the performances will gently increase by 3.15%, 2.60% and 5.13%, for CG, II and SD respectively. More importantly, the STDs of three visual stimuli are dropped to 3.04%, 2.04% and 5.31%. Generally speaking, for three visual stimuli, the longer the duration, the better the accuracy result. Regarding the growth rate, the duration of 5sec seems to be a significant turning point. When the STD is concerned, the duration of 10sec is the best choice.

C. Evaluation of Long-term Performance

Long-term performance is a critical aspect in the authentication system. On one hand, in the continuously repetitive experiment, the subjects tend to develop the fix behavior pattern due to short-term memory. The short-term memory may bias the evaluation results, either positively or negatively. On the other hand, some bio-features as well as human behaviors may slightly change over time. Therefore, the evaluation of long-term performance is necessary for practical use.

1) Evaluation Descriptions: In total, three participants (1 female and 2 males) are involved in this two-month evaluation. The average age of the participants is 28 years old. Only 1 male wears glasses. Particularly, the evaluation has two phases for the test.

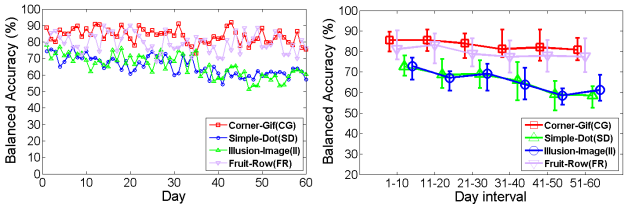
Enrollment Phase: We define each trial as a continuous experiment of four visual stimuli. In order to reduce the effect of short-term memory, four visual stimuli are displayed in a random order. The duration of authentication is set as 10sec in our study. Initially, each subject finishes 10-trial data collection events with a 20-minute break between each two. The collected data are regarded as the templates of the corresponding owners and are used to train the models for the later authentication test.

Authentication Phase: After the classifiers are trained, the long-term authentication phrase is carried out in the following two months. Every day, each subject acts as the owner in turn, while the other two act as the attackers. We define a test round as when a specific subject is selected as the owner. We have 3 test rounds in each day. In each round, 10 trials of the owner and 20 trials of the attackers are performed.

Type	Short-term Acc.	Long-term Acc.	Changes
Corner-Gif (CG)	88.73%	85.18%	-4.00%
Fruit-Row (FR)	85.57%	81.56%	-4.69%
Illusion-Image (II)	80.07%	68.39%	-14.59%
Simple-Dot (SD)	78.15%	67.95%	-13.05%

TABLE III: The system performance comparison between the short-term study and long-term study.

2) Evaluation Results: The average BAC of the subject group for each visual stimulus in each day is illustrated in Fig. 8(a). The system performance comparison between the short-term study and long-term study is summarized in Table III. During the two-month test, the performance of CG and FR is relatively stable, and has no significant descending or ascending tendency. Specifically, CG achieves 85.18% BAC, with a STD of 3.41%. FR obtains 81.56% BAC, with a STD of 5.36%. Compared with the short-term results in the accuracy evaluation, accuracy drops by 4.00% and 4.69%, respectively. This result is expected since the bio-structure of eyeballs and the eye-behavior pattern are hard to change in a short period.



(a) The performance gradually decreases with time.

(b) The performance when averaging 10 days' results.

Fig. 8: The long-term performance results of four visual stimuli.

II has 68.39% BAC, with a STD of 4.89%. SD results in 67.95% BAC, with a STD of 6.13%. After 30 days, the accuracy dramatically drops in both by 14.59% and 13.05%. Note that the large drops in both cases are all due to the decrease in TP, which means that after a certain period, the smartphone may reject the access request from the owner.

We divide 60 days into 6 time intervals, with 10 days in each, to better visualize the trend. As depicted in Fig. 8(b), we calculate the average accuracy in each time interval as well as the best and worst cases for each visual stimulus. The error-bars are related to the best and worst cases during the interval. The performance of CG is overwhelming in all intervals and the worse case of CG is still better than the others in terms of BAC. Both CG and FR can provide stable authentication performance in the 2-month test.

For both SD and II, the best performance occurs in the first interval. And then, the decline appears approximately around 30 days. This observation indicates that the authentication method with II or SD has a shorter lifetime than that with CG or FR, which is about 30 days. To keep the system performance, it is necessary to update monthly the owner template.

D. Evaluation of Feature Dimension Reduction & Sensitivity

Since *EyeVeri* is implemented on resource-constrained smartphones, the demanded resource of the process is important. Feature selection affects both system performance and computational complexity, which is proportional to the demand of CPU and memory. Inadequate features results in bad authentication performance. However, if we extract over-sufficient features, which have no relation with each other, or are heavily dependent on others, those *redundant* features can lower the efficiency of the model, and waste CPU and memory

resources on smartphones, even decrease the battery lifetime. In this section, we examine the effect of feature dimension on classification performance.

We employ Sequential Forward Selection (SFS) to find subsets of features that are most descriptive of the whole feature set [39]. This is a wrapped method in that the feature selection is based on using the classification results themselves and the selection process wraps around the classification. Specifically, the first feature is selected by testing each feature individually in authentication. The feature with the best performance is permanently added to the feature set. In the next round, each of the remaining features combined with the existing feature set is tested and the one with the best performance will be chosen. The process continues until all features are selected. Since authentication always uses previously selected features, redundant features are not selected until the end.

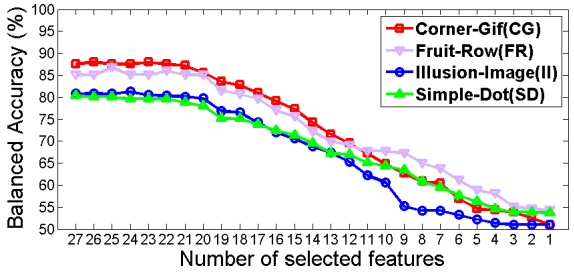


Fig. 9: The impact of feature dimension on authentication accuracy based on Sequential Feature Selection.

Fig. 9 shows the relationship between the feature number and the balanced accuracy for each of the classifiers. In general, the accuracy decreases as less features are included. We can observe sharp accuracy decreases for CG, FR and II, by 37%, 31% and 35% respectively between 21 features and 5 features. SD demonstrates a modest decrease of 33% throughout the whole process. The accuracy results of all visual stimuli are quite stable until the number of features drops below 20.

E. CPU and Memory Footprint

	CG	FR	II	SD
xCPU (Avg.)	45%	31%	36%	30%
Memory (Avg.)	41.4MB	36.9MB	35.4MB	32.4MB

TABLE IV: CPU and memory footprints on Nexus 4.

We investigate the CPU and memory footprints of the *EyeVeri* implementation on Google Nexus 4. Specifically, the CPU and memory usages are measured by the *meminfo* and *cpuinfo* command of the Android Debug Bridge (ADB) shell. The measurement is conducted with four visual stimuli, respectively. As shown in Table IV, the resource consumption in CG is more than the others. Specifically, the average CPU and memory consumptions of CG are higher than those of the other three. This is mainly because CG is a dynamic stimulus implemented in the graphic interchange format (GIF), while the others are static images. The average CPU usage of CG is around 45%, and about 32.3% for the other three. For the average memory consumption, an average of 41.1MB is used for CG, and 34.9MB is used for the other three.

V. User Experience

We conduct two surveys on 20 users to evaluate the usability of the approach. The first questionnaire set in Table V focuses on how they feel about *EyeVeri*. The second one in Table VI is the comparison between our novel method and some other behavioral authentications on smartphones.

	Questions	Score (1-10)	STD
Q1	How comfortable were you when watching CG?	9.0	0.3
Q2	How comfortable were you when watching FR?	9.8	0.2
Q3	How comfortable were you when watching II?	6.1	0.3
Q4	How comfortable were you when watching SD?	6.4	0.2
Q5	Your acceptable auth. duration?	6.2sec	1.3
Q6	Your preferred duration of eye-based auth.?	2.3sec	0.5

TABLE V: Questionnaires and scores about how the users feel about *EyeVeri*. The higher the score, the more comfortable the user feels. The answers of the last 2 questions are in seconds.

	Secure	Reliable	Convenient	Feasible	Average
Eye-behavior	9.3	8.8	9.1	9.1	9.1
Gait-pattern[11]	8.3	7.6	9.5	7.3	8.2
In-air signiture[12]	7.3	7.1	8.1	7.1	7.4
Multi-touch[10]	7.8	9.5	9.1	9.7	9.0

TABLE VI: Questionnaire about how the user feels about the potential behavioral smartphone authentication methods.

Table V shows that all users feel more comfortable with the dynamic interaction type (CG and FR). As for the static type (II and SD), most of the users have negative feedback because staring makes their eyes uncomfortable. Therefore, designing more dynamic and attractive visual stimuli is encouraged. When discussing the acceptable authentication duration, most users feel fine with the duration around 5sec. However, they prefer an average of 2.3sec in the ideal case. To address this concern, we propose to increase the frame sampling rate in our future work.

Next, we ask users to compare the proposed method with some other behavioral methods on smartphones, in terms of security, reliability, convenience and feasibility. We describe gait-pattern [11], in-air signature [12] and multi-touch screen [10] in detail. As shown in Table VI, the users believe that eye behavioral authentications are much more secure than others when smartphones are mostly used in public. When talking about reliability, they regard multi-touch screen as the best option. Meanwhile, most users also have confidence in the eye-behavioral approach that can keep smartphones from unauthorized access. Regarding the convenience, the users feel that the in-air gesture way is too complicated for daily use. For future feasibility, the users agree that both eye-behavioral and multi-touch authentication can eventually be applied on smartphones. Overall, the eye-behavioral method achieves the highest score on average.

VI. Conclusion and Future Work

In this paper, we presented *EyeVeri*, a novel authentication solution for mobile security based on eye movement. We introduced the entire framework of *EyeVeri* and discussed four visual stimuli in the design and the experiment phases. The evaluation shows the promising result of applying non-volitional eye response as a biometrics and indicates that *EyeVeri* is a secure and usable approach for smartphone user authentication. Moreover, *EyeVeri* can combine with other authentication methods, such as face recognition. In the future, we will collect more trials from each subject and recruit more people in the long-term stability study. Also, we will further explore the system performance under different scenarios regarding light sources and body conditions.

Acknowledgment

This work is supported in part by US NSF under grant No. CNS-1423061 and CNS-1421903.

References

[1] "Smart phones overtake client PCs in 2011," http://www.canalys.com/newsroom/smart-phones-overtake-client-pcs-2011/.

[2] "Worldwide Smartphone Markets: 2011 - 2015," http://www.researchandmarkets.com/reports/1871240/worldwide_smartphone_markets_2011_to_2015/.

[3] "Mobile device security threats," http://searchmobilecomputing.techtarget.com/guides/Mobile-device-protection-and-security-threat-measures/.

[4] "Lost and Found: The Challenges of Finding Your Lost or Stolen Phone," https://blog.lookout.com/blog/2011/07/12/lost-and-found-the-challenges-of-finding-your-lost-or-stolen-phone/.

[5] T. Stockinger, "Implicit authentication on mobile devices," in *Ubiquitous Computing, Media Informatics Advanced Seminar LMU*, 2011.

[6] "GoldenEye: A Face Recognition Based Authentication," https://thegoldeneye.googlecode.com/files/GoldenEye.pdf?./.

[7] A. C. Morris, S. Jassim, H. Sellahewa, L. Allano, J. Ehlers, D. Wu, J. Koreman, S. Garcia-Salicetti, B. Ly-Van, and B. Dorizzi, "Multimodal person authentication on a smartphone under realistic conditions," in *Proceedings of SPIE*, vol. 6250, 2006, pp. 120–131.

[8] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, no. 10, pp. 955–966, 1995.

[9] S. A. K. M. F. Saira Zahid, Muhammad Shahzad, "Keystroke-based user identification on smart phones," in *RAID '09 Proceedings of the 12th International Symposium on Recent Advances in Intrusion Detection*, pp. 224–243.

[10] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and i know it's you!: implicit authentication based on touch screen patterns," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 987–996.

[11] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*. IEEE, 2010, pp. 306–311.

[12] G. Bailador, C. Sanchez-Avila, J. Guerra-Casanova, and A. de Santos Sierra, "Analysis of pattern recognition techniques for in-air signature biometrics," *Pattern Recognition*, vol. 44, no. 10, pp. 2468–2478, 2011.

[13] R. J. Leigh and D. S. Zee, *The neurology of eye movements*. Oxford University Press, 2015.

[14] M. Porta, S. Ricotti, and C. J. Perez, "Emotional e-learning through eye tracking," in *Global Engineering Education Conference (EDUCON), 2012 IEEE*. IEEE, 2012, pp. 1–6.

[15] A. De Luca, R. Weiss, H. Hußmann, and X. An, "Eyepass-eye-stroke authentication for public terminals," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2008, pp. 3003–3008.

[16] D. Liu, B. Dong, X. Gao, and H. Wang, "Exploiting eye tracking for smartphone authentication."

[17] V. Vaitukaitis and A. Bulling, "Eye gesture recognition on portable devices," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 711–714.

[18] H. Drewes, A. De Luca, and A. Schmidt, "Eye-gaze interaction for mobile phones," in *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*. ACM, 2007, pp. 364–371.

[19] N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon, "Biometric-rich gestures: a novel approach to authentication on multi-touch devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 977–986.

[20] L. Li, X. Zhao, and G. Xue, "Unobservable re-authentication for smartphones." in *NDSS*, 2013.

[21] M. Shahzad, A. X. Liu, and A. Samuel, "Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 39–50.

[22] S. Das, E. Hayashi, and J. I. Hong, "Exploring capturable everyday memory for autobiographical authentication," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 2013, pp. 211–220.

[23] S. Schneegass, F. Steimle, A. Bulling, F. Alt, and A. Schmidt, "SmudgeSafe: geometric image transformations for smudge-resistant user authentication," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 775–786.

[24] C. Nickel, T. Wirtl, and C. Busch, "Authentication of smartphone users based on the way they walk using k-nn algorithm," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*. IEEE, 2012, pp. 16–20.

[25] R. Leigh and D. Zee, *The Neurology of Eye Movement*. Oxford University Press, 2006.

[26] C. D. Holland and O. V. Komogortsev, "Complex eye movement pattern biometrics: Analyzing fixations and saccades," in *Biometrics (ICB), 2013 International Conference on*. IEEE, 2013, pp. 1–8.

[27] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, 2006.

[28] P. Kasprowski and J. Ober, "Eye movements in biometrics," in *European Conference on Computer Vision (ECCV)*, Nov 2004, pp. 248–258.

[29] A. M. R. Bednarik, T. Kinnunen and P. Franti, "Eye-Movements as a Biometric," *Image Analysis*, vol. 3540, pp. 780 – 789, 2005.

[30] C. Holland and O. Komogortsev, "Complex Eye Movement Pattern Biometrics: The Effects of Environment and Stimulus," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 2115 – 2126, 2013.

[31] O. KOMOGORTSEV and C. HOLLAND, "2D Linear Oculomotor Plant Mathematical Model: Verification and Biometric Applications," *ACM Transactions on Applied Perception (TAP)*, vol. 10, no. 4, 2013.

[32] B. Holland and O. Komogortsev, "Biometric identification via eye movement scan paths in reading," in *International Joint Conference on Biometrics Compendium*, Nov 2011, pp. 1–8.

[33] S. Kawato, N. Tetsutani, and K. Hosaka, "Scale-adaptive face detection and tracking in real time with ssr filters and support vector machine," *IEICE transactions on information and systems*, vol. 88, no. 12, pp. 2857–2863, 2005.

[34] Z. Wanzhi, W. Zengcai, and C. J. X. Xiaoyan, "A method of gaze direction estimation considering head posture," *International Journal of Signal Processing, Image Processing & Pattern Recognition*, vol. 6, no. 2, 2013.

[35] J.-G. Wang and E. Sung, "Study on Eye Gaze Estimation," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 32, no. 3, pp. 332–350, 2002.

[36] "Snapdragon SDK," https://developer.qualcomm.com/mobile-development/add-advanced-features/snapdragon-sdk-android.

[37] H. Xu, Y. Zhou, and M. R. Lyu, "Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones," in *Symposium On Usable Privacy and Security, SOUPS*, vol. 14, 2014, pp. 187–198.

[38] "Weka Toolkit," http://www.cs.waikato.ac.nz/ml/weka/.

[39] G. H. John, R. Kohavi, K. Pfleger *et al.*, "Irrelevant features and the subset selection problem." in *ICML*, vol. 94, 1994, pp. 121–129.