EyeVeri: A Secure and Usable Approach for Smartphone User Authentication

***Abstract*—As mobile technology grows rapidly, smartphone has become indispensable for transmitting private user data, storing the sensitive corporate file, and conducting secure payment transactions. However, mobile security research remains lagging far behind, leaving our smartphone extremely vulnerable to the threat of 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 characterize users eye gaze pattern for access authentication. Through a comprehensive user study, *EyeVeri* demonstrates the promising performance and is a promising approach for smartphone user authentication. We also discuss the evaluation results in-depth and analyze the opportunities for future work.**

I. Introduction

Nowadays, smartphones have overtook personal computers (PC) and become the most prevailing devices for communications and computing services [1] in daily life. By 2015, there will be 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 the increasing ubiquity of smartphones. It is not surprising that smartphone users experience more unauthenticated access than PC users reported in [3], because smartphones intrinsically tend to have higher risk of loss or theft than PCs [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 into two types, i.e., Physiological Biometrics and Behaviorial 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 from that most of these bio-features can possibly be obtained or replicated by adversaries. For example, fingerprint can be stolen in daily life, and voice pattern can be counterfeited by the 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 screen pad, which means that the "password" could be monitored in public and be duplicated later. Other methods such as gait patterns [11] and in-air signatures [12] need obtrusive interaction with smartphones. Meanwhile, it is still possible for the adversaries to mimic them.

In this paper, we present *EyeVeri*, a novel eye-movement based authentication system for smartphones. It captures human eye-movements (i.e., fixations and saccades) and extracts unique gaze pattern for access authentication. Our approach has distinctive advantages because it compiles both physiological and behavioral aspects in nature. Firstly, eye-movement pattern is related to eye bio-structure and each individual has an unique eye-muscle condition. Specific eye-structural or muscle-related features such as the range of view or the 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 [13] and the so-called *Eye-Mind Hypothesis* states that, there exists the direct correspondence between the people's gaze and attention. Thereby, individuals hold unique eye-behavioral features in eye-movements because of their different experiences and habits, according to certain visual stimuli.

In our work, we develop an eye-tracking module based on the eye spatial model and the related gaze processing algorithms. We design and exam four types of visual stimuli to argument various unique eye-related physiological and behavioral features. We comprehensively evaluate the system performance, mainly from three aspects: accuracy, efficiency as well as long-term stability. Our system can achieve the accuracy of around 88% during the experiments, as well as sufficient time efficiency ($5 \sim 10$ seconds) and the 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 helps the user to avoid conspicuous interaction with the smartphones, which results in its huge advantage against shoulder-surfing or smudge traces; 2) *Unique*: it arguments the highly individual-dependent features of eye bio-structure and eye-movement behaviors; 3) *Constant*: it obtains a fairly stable performance in the evaluation of long-term study, which enhances its feasibility.

Our main contributions can be summarised into three-fold:

- We propose and implement an end-to-end authentication system (Fig. 1) based on eye-movement on the smartphone without any extra hardware;
- We develop a set of visual stimuli which can argument

eye-movement features from different aspects;
- We perform a complete and intensive user study to in-depth understand the system performance and limitation.

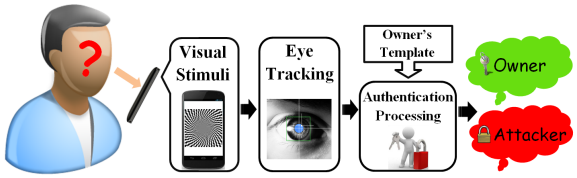


Fig. 1: *EyeVeri* flowchart illustrate the whole process from the time when an unknown user requests to enter the smartphone to the time that the system makes a final decision.

II. Background

A. Related Work of Smartphone Authen.

In recent years, researches have explored lots of novel methods for the next generation smartphone authentication.

Muhammad Shahzad *et al.* [14] designed a set of sliding gestures performed by single or multiple fingers. They extracted behavioral-related features from user's gestures and trained a model by Support Vector Distribution Estimation (SVDE) classifier. The result showed they can achieve an average equal error rate of 0.5% with 3 gestures using 25 training samples for each gesture. Despite the promising accuracy, they did not conduct long-term experiment to see if the user has the consistency in their own gestures. Many other works [15][10][16] also explored the multi-touch based authentication methods on smartphone. Sauvik Das *et al.* [17] 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. The daily events mainly involved communication (SMS, phone calls), content consumption (photos, articles), transportation (driving, public transit), technology usage (apps, software) and weather condition. Specifically, a 14-day long field study was carried out where the participants were required to answer at least 5 questions per day. The result showed that only around 64% questions were correctly answered. Besides, it took users 22 seconds to answer a question on average.

Stefan Schneegass *et al.* [18] 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. The final login success rate of 74% is quite low, and this method is vulnerable to other threats such as shoulder surfing.

Claudia Nickel *et al.* [19] carried out the 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 whole process of segmentation, feature extraction and classification took approximately 7 seconds when implemented on a Motorola Defy smartphone. However, the authentication was based on 30 seconds walk data, during which the smartphone can not be used. Similar work [11] has been done by Derawi *et al.*.

B. Human Vision and Eye Movement

Human visual system is mainly composed of the oculomotor plant and brainstem control [20]. The oculomotor plant is primarily comprised of the eye globe, surrounding tissue, and the extraocular muscles. The extraocular muscles can be further 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 [21].

Various human eye movements are highly related to the neuronal signals [22]. Among those different types, fixations and saccades are of particular interest by 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 [21]. 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 the existing eye-movement theories [22], it provides a potential way to distinguish individuals with the eye-behavioral pattern.

Eye-movement biometric for human identification was initially investigated in 2004 by Kasprowski and Ober [23]. They performed personal identification based on eye movement characteristics, using specific eye-tracking equipments. The experiments proved that it was possible to identify people with this unique bio-feature. Bednarik *et al.* [24] presented a case study investigating 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 [25][26][27] 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 the certain feature extractions of fixations and saccades are able to accurately distinguish the individuals.

Even though eye-movement characteristics contain so many unique features of individuals, to the best of our knowledge, so far there is no existing work on the smartphone authentication based on eye-movement tracking.

III. *EyeVeri* Approach

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.

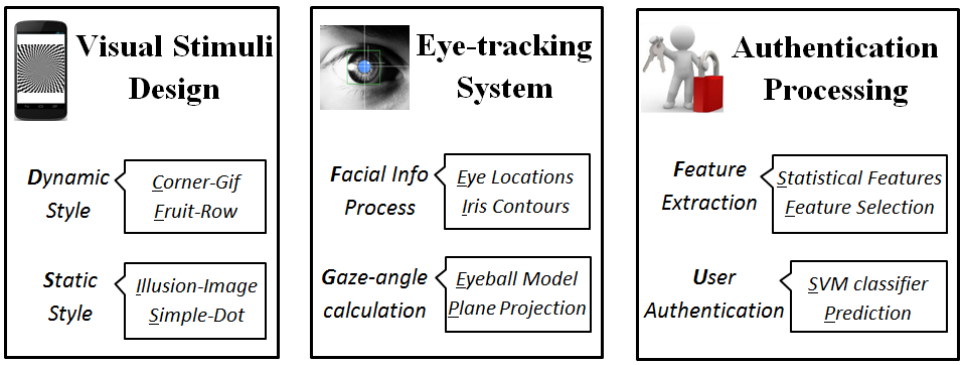
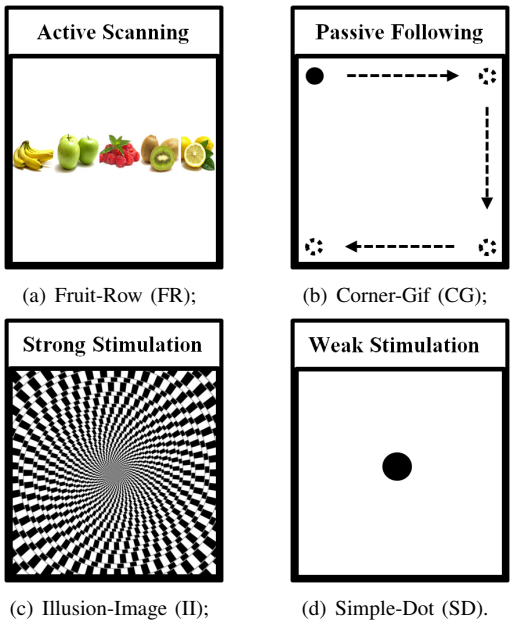


Fig. 2: The proposed smartphone authentication framework based on unobtrusive eye-movement, which comprises three modules: (1) Visual Stimuli Design; (2) Eye-tracking System; (3) Authentication Processing.

When someone attempts to get access into 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: active scanning type, passive following type, strong stimulation type and weak stimulation type, respectively.

In Fig. 3, we design *four* typical visual stimuli for different eye-related features, which include active scanning type (Fruit-Row), passive following type (Corner-Gif), strong stimulation type (Illusion-Image) as well as weak stimulation type (Simple-Dot). For the simplicity of presentation, we respectively use abbreviations FR, CG, II and SD for each of them in the rest of the paper. *Active scanning* type in Fig. 3(a) (FR) 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. *Passive following* type in Fig. 3(b) (CG) 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. *Strong stimulation* type in Fig. 3(d) (II) is a typical illusion image that can strongly stimulate the unconscious vibration (saccade) of the eyeball. Lastly, *weak stimulation* type in Fig. 3(c) (SD) 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 visual stimuli, the subject is not suggested to move their heads around too much in terms of the relative position between the face and the smartphone.

From the perspective of design purpose, each type of visual stimuli contains the disparate information that we are interested in. FR highly depends on the reading or scanning pattern and personal favor with regard to the specific fruit type in the fruit sequence. For example, some subjects have faster scan speed and may spend more time on some fruits because they like them. CG contains 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 the personal uniqueness.

C. Eye-tracking System Introduction

The eye-tracking system mainly involves two steps: Facial Info Pre-processing and Gaze-angle Calculation. Specifically, Facial Info Pre-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 Pre-processing: This step is to achieve positional information of eyeball and iris. Based on the camera preview image, the system first detects the face position

using the Six-Segmented Rectangular (SSR) filter [28]. After obtaining the face position info, the accurate location of both eyes are extracted based on eye's regions of interest (ROI) on the image, using shape-based approach. Then edge detection techniques such as Hough transform are used to effectively figure out the iris contours [29]. 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. Besides, some parts of the iris contour are shadowed by the eyelid. Therefore, the ellipse fitting is implemented to achieve the accurate iris contours.

2) Gaze-angle Calculation: The eyeball model [30] is assumed to be a sphere and the iris lays on the surface of the eyeball. The optical/visual axis (we suppose 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 the users. Then, the two angles that we need to estimate the position of the gaze point are horizontal angle and vertical angle, which lay in gaze-horizontal plane 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 gaze-horizontal plane. Similarly, the vertical angle is between the optical axis and the anatormical axis in gaze-vertical plane. Since the two angles are directly related to the gaze point on screen, we use them for further process. In this work, we refer publicly available techniques [31] to implement our eye-tracking system on the smartphone.

D. Authentication Processing

After the gaze information is 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 contains the 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 feature 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 the eye behavior. *Skewness* and *Kurtosis* are behavioral-based features. *Skewness* is a measure of the asymmetry degree and may has limited contribution in 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 angels. Therefore we eventually have 27 features in total.

2) Authentication Algorithm: Initially, the owner's templates are stored on the smartphone. When an access request occurs, the gaze data sample from the unknown user is collected. Then for the evaluation propose, we formulate the one-class authentication problem as a two-class classification task. Specifically, we employ support vector machines (SVM) as our classifier. SVM finds a hyperplane in training inputs to separate two different data sets such that the distances from the hyperplane to the support vectors are maximized. More specifically, we use the Sequential Minimal Optimization implementation of SVM which is provided in Weka machine learning toolkit [32]. 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 10s.

A. Evaluation of System Accuracy

A key concern to the feasibility and effectiveness of an authentication system is quantification of the degree to which it can accurately recognize the owner and reject the 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

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.

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 an insight on the overall performance of the classifier.

We refer 1 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 totally we have 800 sample data.

2) Single-user Application Scence: Considering the smartphone is owned by one individual, we evaluate the system in the single-user, multi-attackers scene. The attacker scenario is simulated that 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 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:

$$Accuracy(\%) = \frac{TPR + TNR}{TPR + FPR + TNR + FNR} * 100\%, \quad (1)$$

where TPR is the true positive rate, TNR is the true negative rate, FPR is the false positive rate and FNR is the false negative rate. 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:

$$BAC(\%) = \frac{0.5 * TPR}{TPR + FNR} + \frac{0.5 * TNR}{TNR + FPR}. \quad (2)$$

Fig. 4 shows the average BAC of 20 subjects for four visual stimuli. CG achieves the best accuracy of 88.73%,

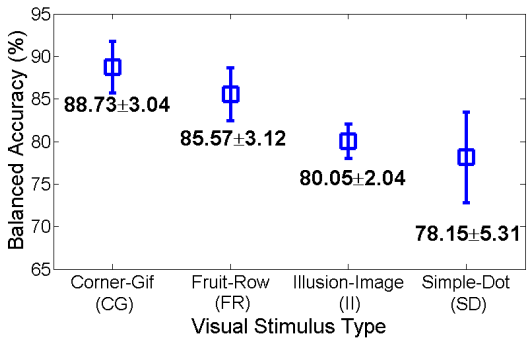


Fig. 4: The average BAC of 20 subjects for four visual stimuli. The error-bars are the standard deviation of BAC among the subjects.

with the standard deviation of 3.04%. FR obtains a close accuracy of 85.57%, with the standard deviation 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 is more easily to stimulate the unconscious eye vibration while the corresponding response time various for SD. In other words, data in II contain more eye vibration information than SD during the experiment. Therefore it achieves a better accuracy.

Moreover, it is worth to mention 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 a close performance with the best cases of II and SD.

b) Receiver operating characteristic (ROC): To take a close 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 classification setups. The two error rates, FNR and FPR, are traded off against each other. If the classification is carried out in a strict setup, both FNR and FPR can be extremely low, which means attackers will be rejected as well as the

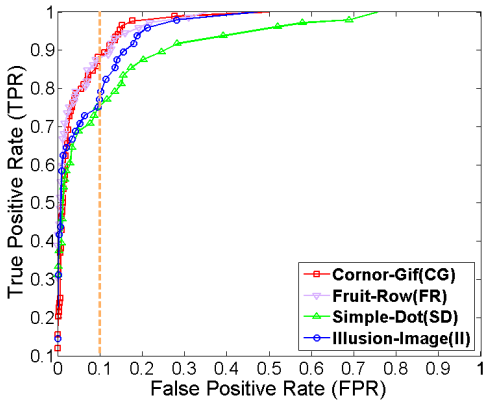


Fig. 5: The average ROC curves of 20 subjects for four visual stimuli. The vertical dashed line in orange implies a 10% threshold of false positive rate.

owner. At the cost of wrongly accepting some attackers, FNR can be reduced and the classifier is less sensitive.

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%, while that of II and SD is approximately 80% and 75%, respectively.

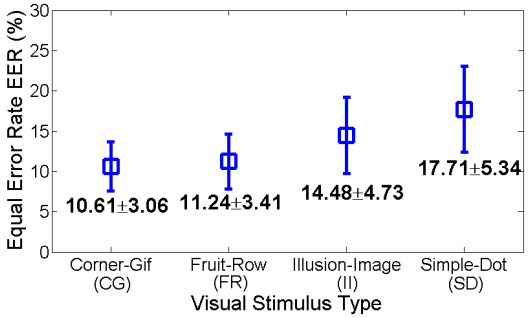


Fig. 6: Equal error rate (EER) of four visual stimuli. The error-bars are the standard deviation of EER among 20 subjects.

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 standard deviation of EER among 20 subjects. CG and FR have the similar EER of 10.61% and 11.24%. The corresponding small standard deviations, 3.06% and 3.41%, suggest the good universality of these two types of authentication upon the subject group. II has a 14.48% EER with a standard deviation of 4.73%, while SD has a 17.71% EER with a standard deviation of 5.34%. The relatively high EER and standard deviation 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 the valid access requests, but also efficiently make the authentication decision.

1) Evaluation Descriptions: To evaluate the 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.

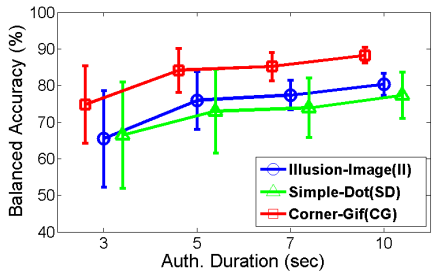


Fig. 7: The average BAC of visual stimuli under different authentication durations. The error-bars are the standard deviation of the accuracy results among 20 subjects in the corresponding authentication duration.[1]

Dur.	CG Acc.	Growth	II Acc.	Growth	SD Acc.	Growth
3s	73.56%	-	65.36%	-	65.89%	-
5s	84.63%	**15.04%**	75.65%	**15.74%**	72.16%	**9.51%**
7s	86.02%	1.64%	78.02%	3.13%	74.34%	3.02%
10s	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 duration. (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 standard deviations (10.81% for CG, 12.85% for II and 14.97% for SD). The BAC results of three visual stimuli are all significantly improved when the duration increases from 3sec to 5sec.

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

Specifically, the performances of CG, II and SD are improved by 15.04%, 15.74% and 9.51%. The corresponding standard deviations are also reduced to 5.82%, 8.12% and 10.53%, respectively. Also, there are no significant improvement when the time increases from 5sec to 7sec. However, when the duration increases from 7sec to 10sec, the performances will gently increased by 3.15%, 2.60% and 5.13%, for CG, II and SD respectively. More importantly, the standard deviations 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 standard deviation is concerned, perhaps the duration of 10s 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 the 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: Totally three participants (1 female and 2 male) are involved in this two-month evaluation. The average age of the participants is 28 years old. 1 male uses glasses and the others do not. Particularly, the evaluation has two phases for model learning and authentication test.

Enrollment Phrase: 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 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 Phrase: 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 a specific subject is selected as the owner. Therefore, we have 3 test rounds in each day. In each round, 10-trial of the owner and 20-trial of the attachers are performed. We use BAC as the accuracy metric.

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. The system performance comparison between the short-term study and long-term study is summarized in Table III.

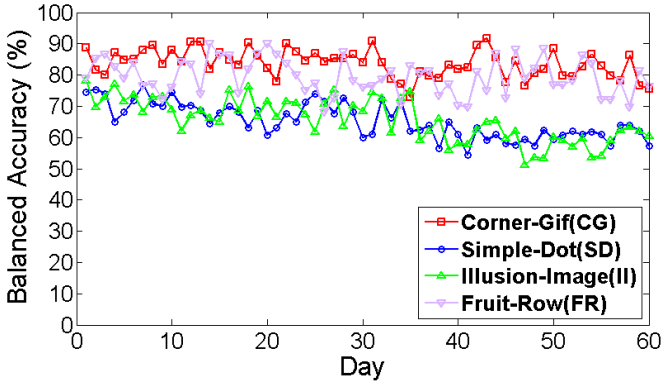


Fig. 8: The long-term performance result of four visual stimuli.

During the two-month test, the performance of CG and FR is relatively stable, and has no significant tendency in descending or ascending. Specifically, CG achieves 85.18% BAC, with a standard deviation of 3.41%. FR obtains 81.56% BAC, with a standard deviation of 5.36%. Compared with the short-term results in the accuracy evaluation, the 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 be changed in the short period.

II has 68.39% BAC, with a standard deviation of 4.89%. SD results in 67.95% BAC, with a standard deviation of 6.13%. After 30 days, dramatic drops in accuracy occur for these two visual stimuli by 14.59% and 13.05%. It is important to point out that the large drops in both cases are all due to the decrease in TPR, which means that after a certain period of time, the smartphone may rejects the access request from the owner.

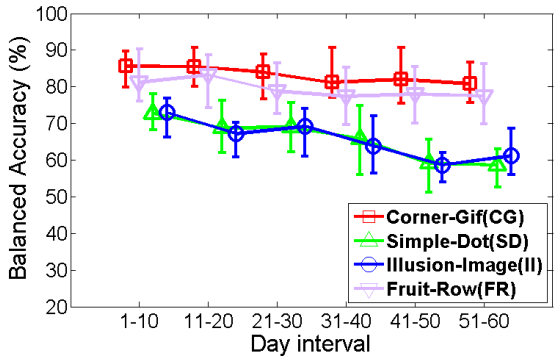


Fig. 9: We divide 60 days into 6 intervals, with 10 days in each. The average BAC of 20 subjects for different visual stimuli in each interval is displayed. The error-bars are the best and worst cases during the interval.

To illustrate the long-term performance in a more concrete way, we divide 60 days into 6 time intervals, with 10 days in each. For each visual stimulus, we calculate the average accuracy in each time interval as well as the best and worst cases. Fig. 9 depicts the summarized results. 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 two-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 monthly update the user template.

D. Evaluation of Feature Dimension Reduction & Sensitivity

Since *EyeVeri* is implemented on resource-constrained smartphones, the demanded resource of authentication 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 the bad performance of authentication. 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 [33]. 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 added into the feature set permanently. 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 the authentication always uses previously selected features, redundant features are not selected until the end.

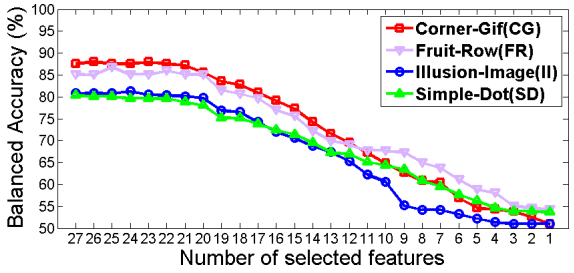


Fig. 10: The impact of feature dimension on authentication accuracy based on Sequential Feature Selection.

Fig. 10 shows the relationship between the feature number and the balanced accuracy for each of the classifiers. In generally, 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 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 the one of the other three is about 32.3%. 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

To further understand the usability, 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?	9.2sec	1.1
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 second.

	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 the staring, especially at II, makes their eyes uncomfortable. This confirms to design more dynamic and attractive visual stimuli to improve user experience. When discussing the acceptable authentication duration, most users feel fine with

the duration within 10sec, and the average acceptable time is 9.2sec. 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 the users compare the proposed gaze-based authentication 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 based on the previous study. As is 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 the reliability, despite that they regard multi-touch screen as the best option, most users also have confidence in eye-behavioral approach to provide smartphones from unauthorized access. Regarding the convenience, the users feel that the in-air gesture way is too complicated for daily use. For the future feasibility, our users agree that both eye-behavioral and multi-touch authentication can eventually be applied on smartphones. Moreover, eye-behavioral method achieves the highest score in average.

VI. Conclusion

In this paper, we presented *EyeVeri*, a novel eye-movement based authentication solution for mobile security. We introduced the entire framework of *EyeVeri* and discussed four visual stimuli in the design and the experiment. The evaluation results indicate *EyeVeri* is a secure and usable approach for smartphone user authentication. Moreover, *EyeVeri* can combine with other authentication methods, such as face recognition.

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] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

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

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

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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

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

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