OcuLock: Exploring Human Visual System for Authentication in Virtual Reality Head-mounted Display

Shiqing Luo*, Anh Nguyen*, Chen Song†, Feng Lin‡, Wenyao Xu§ and Zhisheng Yan*
*Georgia State University, Email: {sluo10, anguyen139}@student.gsu.edu, zyan@gsu.edu
†San Diego State University, Email: csong@sdsu.edu
‡Zhejiang University, Email: flin@zju.edu.cn
§SUNY Buffalo, Email: wenyaoxu@buffalo.edu

Abstract—**The increasing popularity of virtual reality (VR) in a wide spectrum of applications has generated sensitive personal data such as medical records and credit card information. While protecting such data from unauthorized access is critical, directly applying traditional authentication methods (e.g., PIN) through new VR input modalities such as remote controllers and head navigation would cause security issues. The authentication action can be purposefully observed by attackers to infer the authentication input. Unlike any other mobile devices, VR presents immersive experience via a head-mounted display (HMD) that fully covers users' eye area without public exposure. Leveraging this feature, we explore human visual system (HVS) as a novel biometric authentication tailored for VR platforms. While previous works used eye globe movement (gaze) to authenticate smartphones or PCs, they suffer from a high error rate and low stability since eye gaze is highly dependent on cognitive states. In this paper, we explore the HVS as a whole to consider not just the eye globe movement but also the eyelid, extraocular muscles, cells, and surrounding nerves in the HVS. Exploring HVS biostructure and unique HVS features triggered by immersive VR content can enhance authentication stability. To this end, we present OcuLock, an HVS-based system for reliable and unobservable VR HMD authentication. OcuLock is empowered by an electrooculography (EOG) based HVS sensing framework and a record-comparison driven authentication scheme. Experiments through 70 subjects show that OcuLock is resistant against common types of attacks such as impersonation attack and statistical attack with Equal Error Rates as low as 3.55% and 4.97% respectively. More importantly, OcuLock maintains a stable performance over a 2-month period and is preferred by users when compared to other potential approaches.**

I. Introduction

Virtual reality (VR) technology is boosting exponentially. The market size of VR has witnessed an increase of 178% from 2016 to 2018 [51]. By interacting with head-mounted displays (HMD), a user can enjoy immersive virtual content, making VR become a new personal computing paradigm [33], [55]. Due to the diverse applications of VR in entertainment, healthcare, education, and military, sensitive data can be accessed via HMD. For example, credit card information is often stored in HMD for the convenience of games and add-ons purchase in VR App stores [38]. In medical applications, patients' CT scan models have been viewed in VR HMD to assist the diagnosis of structural abnormalities on human body [24]. Stimuli implying patients' traumatic experience have been displayed in VR during psychological exposure therapies [54]. In military applications, pilots learn the operation of top-secret aircrafts in VR simulator [46]. Protecting HMD from unauthorized access thus becomes critical in guaranteeing users' experience and privacy in VR systems [52].

Unfortunately, VR computing is still at its infancy and state-of-the-art HMD authentication methods suffer fundamental limitations in security. Recent systems have adopted traditional unlock pattern, PIN, and graphical passwords in VR through new input modalities such as remote controllers and head movement navigation [36], [13]. Similarly, common behavior biometrics such as head movement [28], [44] and body motion [40] were also proposed to authenticate users. However, all these systems expose the entire authentication action to the public, making various attacks possible through observation. For example, adversaries have successfully conducted side-channel attacks by observing user input behavior and inferring the virtual input [14], [30]. Since wearing HMD blocks users' real-world visuals and decreases their situation awareness [17], the threat of observation-based attacks in VR is significantly higher than that in traditional computing devices.

We envision an *unobservable* solution that utilizes the distinctive human visual system (HVS) for VR authentication. Since human eyes are fully covered by HMD without public exposure, it is unlikely, if not impossible, for nearby adversaries to observe users' eye activities and execute observation-based attacks. While eye gaze biometrics have been used in PCs and smartphones [41], [45], [47], [56], harnessing HVS for *stable* VR authentication remains challenging. The error rate of previous eye gaze based authentication is still high (e.g., EER of 6.3% [45]) and the performance quickly degrades over time [47]. One likely reason is that eye gaze pattern, as a behavioral biometric, varies when a user attempts the authentication under different cognitive states and such variability becomes more significant as time passes by.

In this paper, we propose to explore the HVS as a whole to build a stable and unobservable VR authentication system.

Network and Distributed Systems Security (NDSS) Symposium 2020
23-26 February 2020, San Diego, CA, USA
ISBN 1-891562-61-4
https://dx.doi.org/10.14722/ndss.2020.24079
www.ndss-symposium.org

We utilize the fact that in addition to the eye globe many other components in the HVS such as the eyelid, extraocular muscles, cells, and surrounding nerves conduct unique activities that can be triggered and sensed in VR environment. In addition to the unobservable nature, comprehensive analysis of the entire HVS in VR also enhances the authentication performance. First, unlike prior eye gaze biometrics that only focused on the movement of eye globe [41], [45], [47], [56], considering the physiological characteristics of various HVS components in the authentication can enhance the performance stability since HVS biostructure is far less dependent on the time-varying cognitive states of users. Second, some less-intuitive features of HVS cells and nerves (besides traditional eye gaze) not presented in physical reality can be triggered by immersive VR content [39] and utilized to increase the distinctiveness among users. These unique and temporally stable features consequently improve the average error rate.

Realizing such a biometric authentication system for VR HMD is non-trivial. While eye globe movement has been captured by previous works using monitor-mounted eye trackers or high-resolution cameras in an illuminated open space, we must trigger and measure low-level activities of HVS components that are not clearly visible but uniquely presented in dark VR environment. To tackle this challenge, we propose an electrooculography (EOG) based HVS sensing framework for VR. EOG measures the electrical signals resulted from biological activities in the HVS and can characterize both behavioral and physiological features of the HVS in VR environment. Since the foam face cover of current VR HMD has direct contact to eye sockets and their surrounding nerves, we attach thin electrodes on the cover to measure low-level HVS signals which are otherwise unavailable via eye trackers or video cameras in the context of PCs and smartphones. We also design a set of visual stimuli to trigger EOG signals that manifest desirable HVS features.

Another challenge of the proposed authentication system is that discriminative features must be extracted from the comprehensive HVS data for efficient model training and reliable authentication. Previous biometric systems [29], [19], [7] trained a two-class classifier to differentiate the owner and others, but a new model had to be trained for every new owner. We propose a different authentication scheme to remove the model training and owner enrollment overhead. Specifically, EOG is first processed to recognize common HVS activities and extract a suite of symbolic features representing the behavior and biostructure of various HVS components. Each feature of the input will be compared with that of the owner's record to generate a matching score. The matching scores for all features are fed to a comparator to indicate if the input matches the owner's record. That way, we only need to train one comparator and can use it for all future owners.

We validate the proposed authentication system, referred as *OcuLock*, through extensive evaluation. During our evaluation that involves 70 subjects and lasts for 2 months, we validate the stability, security and user preference of OcuLock. In the security analysis, the system achieves Equal Error Rates of 3.55% and 4.97% against impersonation attack and statistical attack respectively. The reliability study over a 2-month period shows that the model can maintain a far more stable performance than existing eye gaze behavior based approach. Moreover, the user study demonstrates that OcuLock is preferred over other authentication approaches due to its convenience, security, and social comfort.

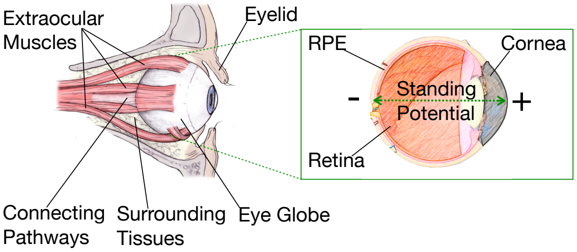


Fig. 1: HVS structure (left) and standing potential (right).

To summarize, the contributions of this paper include:

- We propose an EOG-based framework to measure the HVS as a whole for VR authentication, where visual stimuli are designed to trigger the HVS response and EOG is collected to characterize the HVS.
- We design a record-comparison driven authentication scheme, where distinctive behavioral and physiological features are extracted and accurate authentication decisions are made.
- We perform an extensive evaluation of the proposed OcuLock system including reliability performance of the authentication, security analysis against several attacks, and user study of VR HMD authentication.

II. Background

A. *Human Visual System*

As shown in Figure 1, human visual system (HVS) is primarily comprised of four components: eyelid, eye globe, their surrounding tissues, and extraocular muscles, as well as the bidirectional connecting pathways to the visual cortex and other parts of the brain. The eyelid opens and closes regularly to expose the cornea to the outside, giving vision to humans. The eye globe absorbs incoming light through the cornea (its outermost layer of tissues) and shines it on the retina (its innermost and light-sensitive layer of tissues). After the retina transduces the received images into electric pulses, the connecting pathway delivers the pulses to the brain. Conversely, the brain and the nervous system can also send control signals to extraocular muscles, which contract or relax to motivate the eye globe rotation and the eyelid movement. For example, signals from the sympathetic nervous system are received by eyes to trigger the reaction of alert [26].

Since the size, shape, position, and anatomy of the HVS vary from person to person [21], these HVS components and their daily interaction present unique features that can distinguish people. The opening and close of the eyelid are of different extent and speed due to the distinctive muscle strength of HVS among people [1]. Similarly, the eye globe rotation for each individual is only able to reach a limit determined by the size and shape of the eye globe [53]. In order to transport nourishing substances to the retina and remove wastes via the connecting pathway, a layer of cells in the posterior part of the eye globe called retinal pigment epithelium (RPE) have to conduct metabolism. The metabolic rate of RPE depends on the activeness of cells, which is unique among individuals [49]. Finally, the sympathetic signals transported to the eyes show unique energy patterns dependent on the biostructure of people's sympathetic nerves [26].

Apart from the aforementioned physiological features, HVS also involves voluntary movement that demonstrates discriminative patterns. Eye globe typically has two basic types of movement, where fixations are sessions when eyes maintain a stationary gaze on a single location and saccades are sessions between two fixations when eyes move in the same direction quickly. The trace of fixations and saccades generates the scan-path, which varies among people and is uniquely influenced by individuals' personal emotion and preference [41]. Since immersive VR display triggers different eye globe movement from traditional display [39], comprehensive analysis of eye gaze could benefit the authentication.

Building on these facts, we conclude that HVS contains unique physiological biostructure and voluntary movement to authenticate VR users. While the eye globe movement was individually used as gaze biometrics, we consider the HVS as a whole in this paper to explore low-level visual activities and the interaction among HVS components.

B. Electrooculography

Electrooculugraphy (EOG) measures the electric voltage variance between two *sensing positions* on the skin near human eyes [5]. The voltage variance is resulted from *standing potential* (shown in Figure 1), a steady electric potential field existing between cornea (positive pole) and retina (negative pole). The standing potential is formed by transepithelial potential (TEP), the difference of electric potential between two sides of RPE cell membrane, which is originally caused by the metabolism of RPE in the HVS.

As EOG measures the cornea-retinal standing potential that exists between the front and the back of the eyes, it is closely related to all major components of HVS. If the eye globe moves from the center position toward one of the two EOG sensing positions, this sensing position becomes closer to the positive side (front) of the eye globe and the opposite sensing position becomes closer to the negative side (back). Assuming that the resting potential is constant, the recorded potential between the two sensing positions is a measure of the eye's rotating angle [4]. Similarly, when eyelid moves, the electric pulses generated by extraocular muscles cause a rapid rise and drop of voltage in the eye area which can be detected by EOG [1]. EOG is also able to measure the activeness of RPE cells and sympathetic nerves as the standing potential is highly influenced by the RPE metabolism and sympathetic signal transportation. [49], [26]. Therefore, it is feasible to exploit EOG to measure the characteristics of HVS.

III. Problem Statement

A. System Model

We assume a general VR viewing scenario as shown in Figure 2. The system includes an owner and a VR HMD. The *VR HMD* is equipped with electrodes to collect EOG signals.

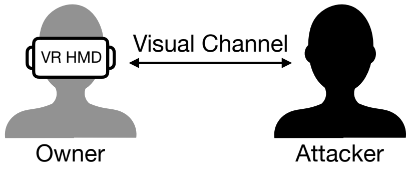


Fig. 2: A general VR viewing scenario.

The VR HMD analyzes the EOG data and compares it with the owner's record to make the authentication decision. The VR HMD that processes EOG data and authenticates users cannot be forced to run unintended code. The *owner* is enrolled to this VR HMD during which she views the visual stimuli by moving her eyes around. Templates of EOG records for all enrolled owners are stored in the VR HMD. The enrollment is a secure process.

The system involves a *visual channel* between the attacker and the owner. The attacker can observe who is using the HMD and thus know the identities of all enrolled owners. The attacker can also visualize the owner's head and body motion if there is any. However, since the owner's eyes and surrounding areas are fully covered by the HMD without public exposure, adversaries cannot observe the displayed content or the owner's eye movement without alerting the owner.

B. Threat Model

We assume a powerful adversary who has enough time and space to freely perform attacks. As HMD is a detachable device that can be unplugged and carried along, attackers can steal the device and conduct attacks in another place. Furthermore, many unauthorized purchases via VR HMD are made by people who are known to the owners without ill intent, e.g., their children [16]. We also assume that the attacker has not installed malware in the HMD to monitor the input. The attacker has not attached additional hardware to the HMD to capture signals. We do not consider the attack that requires nearby complicated device either, e.g., the attack using an antenna near the target HMD to capture electromagnetic emanations and infer user input. However, the attacker may utilize other methods to indirectly obtain information related to user input, e.g., by statistical attack. The objective of the adversaries is to input EOG either directly or indirectly to the VR HMD in order to bypass the authentication. We consider the following types of attacks.

- *Impersonation Attack*: After observing the owners' authentication action, the attacker puts on the HMD and impersonates one of the enrolled users in the system. The attacker then attempts the authentication by providing her own EOG signal.
- *Statistical Attack*: The adversary obtains the statistics about EOG signals for a population similar to the victim owner. The attacker forges new EOG records with the most probable feature values and then attempt the authentication with a higher chance of success. This can be done by observing the enrolled owners and identifying a similar population, e.g., college students. Then the attacker can let a target population attempt the HMD authentication and collect a dataset of EOG

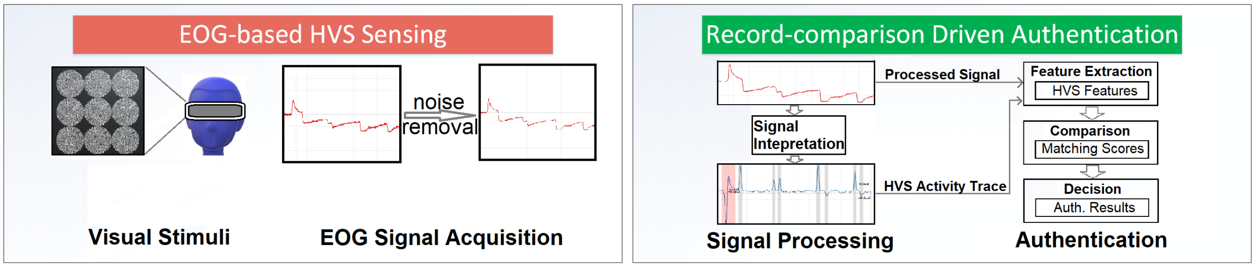


Fig. 3: The architecture of OcuLock.

signals for record forgery. The forged record can be fed to the VR HMD by connecting a voltage generator or injecting the signal to the authentication code.

C. Design Goal

To protect the VR HMD in the above scenario, OcuLock has the following design goals.

- *Secure*: OcuLock must be resistant to all of our proposed attacks from adversaries when the HMD is left unattended or stolen.
- *Reliable*: OcuLock must have small error rates to prevent unauthorized access. The performance should keep stable overtime to avoid the frequent update of authentication biometric.
- *Usable*: The authentication process must be fast and simple. The users should not wear additional sensing devices besides the HMD. No credentials should be memorized.

IV. System Architecture

In this section, we overview the proposed OcuLock system and its authentication protocol. As depicted in Figure 3, OcuLock is comprised of an EOG-based HVS sensing framework and a record-comparison driven authentication scheme.

When a user claims her identity and attempts to access the HMD, OcuLock renders VR visual stimuli on the screen. The visual stimuli are designed to trigger unique HVS activities for VR authentication. While the user is viewing the VR scene, EOG signals are acquired through electrodes embedded in the HMD. Various types of noise are filtered during the acquisition to generate clean EOG signals.

Next, the EOG signal will be analyzed to detect key HVS activities such as eye globe and eyelid movement by using wavelet domain analysis. As a result, a trace of HVS activities manifesting various features is produced. Finally, based on the derived trace of HVS activities, the clean EOG is re-examined to extract the biostructure and behavior features of HVS. The extracted feature vector of the current attempt is compared with that of the claimed owner. The comparison result for each feature is fed to a machine learning model to determine whether the current attempt is from the owner or an attacker.

V. EOG-based HVS Sensing

In this section, we introduce the EOG-based HVS sensing framework proposed by OcuLock. Specifically, we design visual stimuli to trigger EOG response in VR and then collect clean EOG signals manifesting distinctive HVS characteristics.

A. Visual Stimuli Design

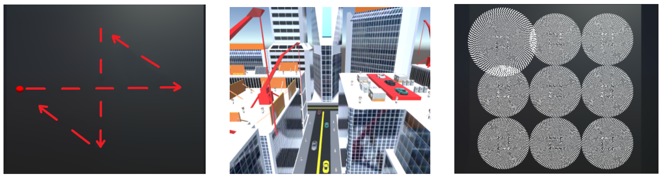
1) Design Principle: Without proper stimulation, users' eye movement may be minimal and some HVS characteristics, e.g., the extent of eye globe rotation, may not be manifested. Hence, it is important to design visual stimuli to trigger desired HVS activities.

In consideration of usability, it is important to keep the stimuli simple and intuitive so that no special efforts are required. The only instruction for OcuLock's users is that they follow moving objects using their eyes. Users need not memorize any types of credentials during the authentication. They also do not need to wear extra hardware or take off the HMD. This level of user efforts is minimal and is consistent with existing authentication methods, such as speaking for voice recognition and typing passwords.

The procedure of stimuli design is as follows. First, three typical types of eye behavior that trigger eye globe/eyelid-level and cell/nerve-level HVS activities were identified [47]: fixed-trajectory movement, free exploration, and involuntary micro-saccades. Second, to elicit the behavior, we designed Fixed-Route, City-Street, and Illusion, respectively, in a 2D image form. Third, to exploit the powerful graphic rendering techniques in VR, the 2D stimuli were converted to 3D.

All visual stimuli are displayed on the main viewport that human eyes face when the HMD is put on. The viewing range of the viewport is 90 degrees from left to right and 60 degrees from top to bottom. Such a setting allows users to view all elements of the stimuli by only rotating their eyes without the need for head navigation [25]. The visual stimuli should also elicit distinctive HVS responses so that OcuLock is able to discover the uniqueness of each user in VR.

2) Visual Stimuli: We design three types of visual stimuli to investigate their impacts on authentication results. In the *Fixed-Route* stimulus (Figure 4a), we follow the principle of simplicity and present a 3D spherical red ball changing positions step by step from left to right and then from top to bottom in a fixed trajectory. The ball stays at each intermediate positions for a given time interval. We aim to study if such a simple stimulus can trigger enough HVS response for authentication.



(a) Fixed-Route; (b) City-Street; (c) Illusion.

Fig. 4: Three visual stimuli. (b) is a static scene while the other two are dynamic.

By responding to this stimulus, all users will have almost the same scanpath. However, the micro-saccades determined by extraocular muscles [8] could be different among users. Since the ball moves from one end of the viewport to the other, it forces the user to rotate her eye globe as much as possible to track the ball. The extent of eye rotation distance depending on the eye globe size and shape is then triggered and collected in EOG signals. As users also blink during the session, it provides an opportunity to observe the extent and strength of eyelid action [1]. Furthermore, the above eye response is triggered along with low-level cell and nerve activities. Therefore, HVS characteristics such as metabolism intensity and sympathetic signal energy would be reflected.

The *City-Street* stimulus (Figure 4b) is a 3D model of a street containing diverse elements such as buildings, vehicles, billboards, and cranes. We aim to investigate how users respond to static objects in a VR scene through this stimulus. Since all objects are static, users will not follow any moving objects. Instead, they will explore the VR scene freely. The scanpath thus reflects the unique viewing interests and habits of a user [41]. Similar to *Fixed-Route*, the low-level HVS signals from cells and muscles can be triggered by the user's movement.

Finally, the *Illusion* stimulus (Figure 4c) contains nine spinning vortexes, among which one special vortex is growing larger and shrunk to its original size within a short time interval. Each vortex takes turns to become the special one in a fixed order (from left to right and top to down). By following the expansion and shrinkage of the spinning vortexes, users present a fixed scanpath but different low-level HVS data as in *Fixed-Route*. The physiological features of eye globe rotation can also be triggered since the vortexes cover the entire viewport. Moreover, since spinning vortexes elicit more micro-saccades and blinks [8], this visual stimulus is designed to characterize fine-grained HVS actions and more cell and muscle activities.

The benefit of the three visual stimuli is that they are designed to trigger a set of physiological and behavioral features of HVS. It is unlikely for adversaries to forge an EOG record containing both similar scanpath and extensive low-level HVS information.

B. EOG Signal Acquisition

1) Hardware Setup: We propose to measure the low-level HVS activities through EOG [5]. By attaching thin electrodes on the face cover of HMD and placing them in the appropriate positions near eye sockets as shown in Figure 5, EOG measures the electric voltage variance around eye areas. As shown in the figure, we attach two electrodes to the outer edges of eyes to collect the horizontal voltage variance and another two electrodes to the upper and lower part of the right eye to collect the vertical voltage variance. One more electrode is attached to the forehead for signal reference.

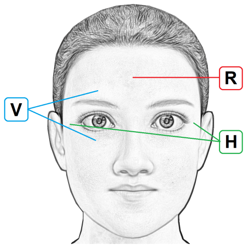


Fig. 5: Electrodes placement for EOG acquisition.

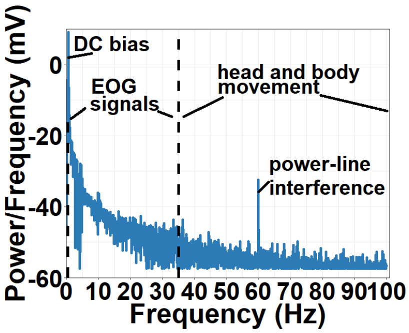
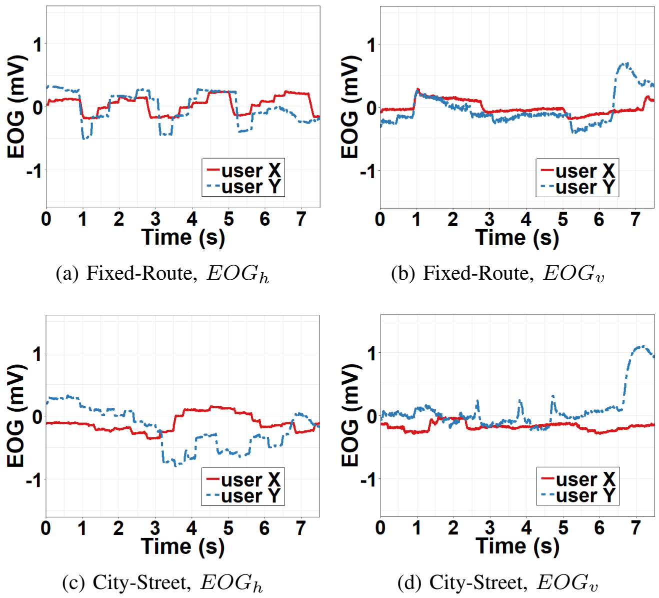


Fig. 6: The frequency domain of raw EOG signals. Various types of noise can be removed by filters.

We measure the variance detected by the two groups of electrodes at a sampling rate of 200 Hz through two channels. Each channel is equipped with an adapter which integrates the voltage variance detected by electrodes into electrical signals. The electrical signals are then digitized and the raw EOG signals representing horizontal and vertical variance are generated.

The proposed EOG-based sensing enables the measurement of HVS signals that are otherwise unavailable in previous systems capturing high-level eye gaze patterns. Due to the miniature nature of electrodes, users will not be burdened by the weight of the extra hardware. In fact, EOG sensors have been embedded in commercial smart glasses, e.g., JINS MEME, to sense eye activities [32]. More importantly, the collected EOG signal is a time series that can be processed and analyzed without high computation overhead.

2) Noise Removal: After the raw signals are collected, OcuLock removes various types of interference from the horizontal and vertical EOG and generate two clean signal components, EOG_h and EOG_v, for further authentication analysis. The measured raw electrical signals contain DC bias, power-line interference, and electricity generated by neurons and muscles when subjects move their head and body during EOG collection. As illustrated in Figure 6, since each type of noise is of a specific frequency, we remove them using



(a) Fixed-Route, EOG_h

(b) Fixed-Route, EOG_v

(c) City-Street, EOG_h

(d) City-Street, EOG_v

Fig. 7: EOG signals of two users are distinctive.

filters. The DC bias has a frequency lower than 0.05 Hz. The frequency of power-line interference is around 60 Hz. As for noise from head and body movement, its frequency is much higher than the frequency bands of EOG signals that is between 0 to 35 Hz. Therefore, OcuLock applies a bandpass filter to pass frequencies between 0.05 and 35 Hz.

3) Sample EOG Data: In order to validate the feasibility of EOG-based HVS sensing, we conducted a pilot study to investigate the EOG of different VR users. To obtain EOG signals containing enough information, we carried out two experiments with the *Fixed-Route* and *City-Street* stimuli, respectively. The detailed experiment setup is consistent with the main evaluation and will be elaborated in Section VII.

We show the noise-removed horizontal EOG, EOG_h, and vertical EOG, EOG_v, of two users in Figure 7. It can be seen that the EOG signals of user X and user Y are significantly different for both experiments in terms of both horizontal and vertical EOG. Since we set users' right side as the positive pole for EOG_h, a positive EOG_h indicates users are looking to their right. Similarly, a positive EOG_v implies eyes looking up. As shown in Figure 7a and 7b, users' EOG signals achieve a similar fluctuation trend but present distinct details in the first experiment. This is because users followed an identical scanpath while viewing the *Fixed-Route*. However, as EOG is also impacted by other HVS interaction and biostructure, the signals still present an obvious difference. For example, different size of eye globe results in the different magnitude of EOG when gazing at the same location. For the *City-Street* (Figure 7c and 7d), users' scanpath becomes different due to the free exploration and their different areas of interest, which makes the EOG signals more distinct. As we will show in Section VI-B, in addition to the movements of eye globe that are straightforward to visualize using the EOG, there are other important features that can be extracted from the temporal and frequency domain of EOG signals to characterize HVS activities and biostructure.

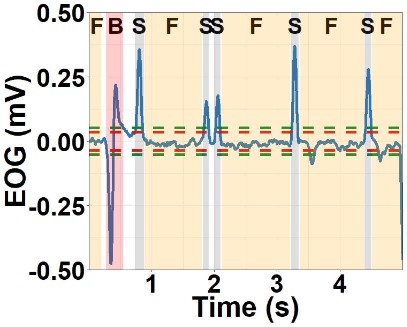


Fig. 8: Wavelet-transformed EOG marked with blinks (B), saccades (S) and fixations (F).

VI. Record-comparison Driven Authentication

In this section, we introduce the record-comparison driven authentication scheme of OcuLock. In particular, OcuLock first analyzes the clean EOG to recognize a trace of HVS activities including saccades, fixations, and blinks. OcuLock then extracts HVS physiological and behavioral features from the clean EOG by utilizing the activity trace and compare the input with the owner's record for authentication decisions.

A. EOG Signal Processing

1) Saccade and Fixation Recognition: As discussed in Section II-A, saccades and fixations are the two basic movements of eye globe that manifest many behavioral and physiological HVS features. In OcuLock, we employ a wavelet transform based algorithm to detect saccades and fixations in order to assist the feature extraction. This algorithm is transplanted from a prior study [5]. It can easily identify a signal segment of specific shape and has been shown to achieve good performance on eye movements recognition. A wavelet-transformed EOG signal is shown in Figure 8. The segments of high EOG changing rate (high eye rotation speed) appear as peaks or valleys in the transformed signal. By applying a threshold th_{sd} (horizontal red dashed lines) and removing segments shorter than 10 ms [11], saccades can be detected (marked with "S"). Similarly, all segments between the thresholds th_{sd} and $-th_{sd}$ and longer than 100 ms are marked as fixations [31] ("F").

To optimize the threshold th_{sd} for our implementation, we collected the ground truth of saccade and fixation for 50 EOG records from 5 users and tested the algorithm performance under varying th_{sd}. By following the methodology in [5], we inspected the EOG visually and identified 698 horizontal saccades and 774 vertical saccades as the ground truth. To evaluate the accuracy of the algorithm, we compare the saccades recognized by the algorithm with the ground truth and calculate F1 Score defined as follows,

$$F1\ Score = 2 * \frac{\frac{TP}{TP+FP} * \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \tag{1}$$

where TP, FP, and FN are the number of true positive, false positive and false negative, respectively. We investigate the accuracy with the threshold varying from 0.01 to 0.05 in

TABLE I: List of HVS features in OcuLock ("V"=Vertical; "H"=Horizontal).

Index	Name	EOG-based Calculation	Category	Component
1	Eyelid Close Speed	Slope of EOG signal during blink close phase.	Physiological	V
2	Eyelid Open Speed	Slope of EOG signal during blink open phase.	Physiological	V
3	Eyelid Stretch Extent	Amplitude of EOG signal during blink close phase.	Physiological	V
4 & 5	Metabolism Intensity	Arden Ratio (AR).	Physiological	H & V
6	Extent of Right Rota. Dist.	Max amplitude of positive EOG/AR.	Physiological	H
7	Extent of Left Rota. Dist.	Max amplitude of negative EOG/AR.	Physiological	H
8	Extent of Up Rota. Dist.	Max amplitude of positive EOG/AR.	Physiological	V
9	Extent of Down Rota. Dist.	Max amplitude of negative EOG/AR.	Physiological	V
10 & 11	Sympathetic Energy	Wavelet transform amplitude from 0.05 to 0.5 Hz.	Physiological	H & V
12 & 13	Fixation Start Time	Start time of fixation.	Behavioral	H & V
14 & 15	Fixation Duration	Duration of fixation.	Behavioral	H & V
16 & 17	Fixation Centroid	Average EOG amplitude during a fixation.	Behavioral	H & V
18 & 19	Saccade Start Time	Start time of saccade.	Behavioral	H & V
20 & 21	Saccade Duration	Duration of saccade.	Behavioral	H & V
22 & 23	Saccade Location	5-point sampling of saccade path.	Behavioral	H & V

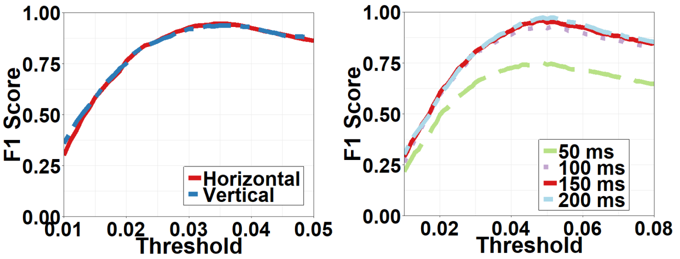


Fig. 9: Saccade/fixation (left) and blink (right) detection algorithms can be optimized by seeking the best thresholds.

50 steps. As shown in Figure 9, the algorithm achieves the highest F1 score when th_{sd} reaches 0.036. We therefore select it for saccade and fixation recognition.

2) Blink Recognition.: As the biostructure features of the eyelid are only presented when the eyelid is moving, i.e., during a blink, we need to recognize all eye blinks before the feature extraction. A blink is the rapid closing of the eyelid accompanied by a rapid eye globe rotation. In the wavelet-transformed domain, the vertical signal component appears as a signal peak followed by a signal valley [5] without a long interval. Hence, blinks can be recognized by first applying two thresholds to the transformed EOG_v to identify segments as peaks or valleys, i.e., th_{bd} and $-th_{bd}$ (two horizontal green dashed lines). An interval threshold th_t is then adopted to drop those segments that are successive saccades. As shown in Figure 8, the segment marked with "B" represents a blink.

To optimize the thresholds for OcuLock, we identified 359 blinks as ground truth from the same set of EOG records mentioned above. We compared the recognized blinks with ground truth across varying th_{bd} from 0.01 to 0.08 (in 70 steps) and different th_t (50, 100, 150, and 200 ms). As shown in Figure 9, the algorithm achieves the optimal F1 score of 0.972 using th_{bd} at 0.052 and th_t at 200 ms, which are selected for blink recognition.

B. Feature Extraction

In contrast to existing eye gaze authentication for smartphones and PCs with a high error rate and variability, OcuLock explores the HVS as a whole and improves the performance reliability by considering low-level HVS biostructure and behavior. Given the trace of HVS activities recognized in Section VI-A, we go back to the EOG signals and extract these features. For example, based on the time interval of a blink, we can derive the eyelid stretch extent by inspecting the EOG amplitude during that interval. We first extracted a long list of features. Then we tested the impact of removing each one feature from the model. If the accuracy of the model remained the same after we removed a feature, then this feature was removed permanently. For example, the median eyelid close speed was removed since it has duplicated effects as eyelid close speed distribution.

The list of features is summarized in Table I. Most features, e.g., eyelid close speed, have multiple samples because an HVS activity, e.g., eye blink, can happen multiple times in an EOG record. In this case, we store the feature as a distribution in the form of probability density function (PDF). This unique design enables OcuLock to capture a comprehensive view of the feature compared to previous eye-based feature extraction that generates a single scalar number for a feature. For features with a single sample, e.g., metabolism intensity, we represent them via a scalar number as well. We extract features from both horizontal and vertical EOG signals except for the eyelid-related features that are only extracted from EOG_v. We indicate this by "H" or "V" in Table I.

Physiological Features. Eyelid features decided by the unique eyelid biostructure and extraocular muscles are extracted from the original EOG signal and the eyeblink trace. Each blink is presented as a peak in the original EOG signal (before transform), where upward-going signal indicates the eyelid close phase and downward-going signal implies the eyelid open phase. Hence, eyelid close speed can be calculated by the slope of the upward-going EOG segment and eyelid open speed can be computed by the slope of the downward-going EOG segment. Eyelid stretch extent signifies the largest extent the eyelid can move and can be represented by the maximum amplitude of EOG signal during eyelid close phase.

As discussed in Section II-B, the metabolism intensity of RPE, uniquely determined by surrounding cell conditions, can be revealed by the values of standing potential [4] and thus can be measured by EOG. We derive metabolism intensity by calculating the Arden Ratio. Arden Ratio is of positive correlation with the RPE metabolic rate and has been used by doctors to examine the metabolism of RPE cells [43]. To calculate the Arden Ratio, we first search through the entire EOG signals and derive the absolute values of the signal at all peaks and valleys. The ratio between the maximum and minimum absolute values is derived as the result.

The distinctive size and shape of the eye globe result in different rotating and reachable distance of eyes for different users. The EOG signal has a linear relationship to the rotating angle of eyes and the coefficient is determined by the standing potential [4]. We approximate the standing potential by the Arden Ratio [43] and then derive the rotating range in four directions (up, down, left and right) by dividing the EOG amplitude by the Arden Ratio. For example, the extent of right rotating distance represents the angular distance between the central reference point and the rightmost point the eye can reach. We calculate this feature by the maximum amplitude of all peaks in EOG_h divided by the horizontal Arden Ratio.

Sympathetic signals show unique energy patterns which depend on the nature and activeness of individual's sympathetic nerve systems. Such signals concentrate between 0.05 and 0.5 Hz frequency band of EOG signals. We derive its frequency domain information by re-using the wavelet transform results of EOG signals across the above frequency bands. Since this is already computed in the signal processing step, sympathetic signals can be extracted without additional overhead.

Behavioral Features. This category of features characterizes the voluntary movement of the eye globe that signifies users' unique viewing habits and preferences. They are extracted from the trace of fixations and saccades detected from EOG signals.

We extract the start time, duration, and centroid position of all fixation instances. Fixation centroid can be derived by computing the average EOG signal amplitude during a fixation, which represents the horizontal and vertical offset with respect to the resting position. Since these features are stored as a distribution, fixation start time and duration imply the temporal characteristic of fixations while fixation centroid indicates the spatial property of fixations. Similarly, we extract the saccade start time, saccade duration and saccade location. Saccade location represents EOG values at five moments during a saccade: the beginning, the first quarter time of the saccade, the medium moment, the third quarter time, and the ending. Both temporal and spatial characteristics of saccades are extracted.

C. Authentication Decision

To make the authentication decision, previous biometric authentication systems [29], [19], [7] use the extracted features to directly trained a classifier in order to differentiate a given legitimate user from all other users. As a result, a classifier is built for each enrolled owner. Every time a new owner is enrolled, a new classifier has to be trained from scratch to recognize the new owner's feature patterns. Such a methodology requires extra overhead for classifier training during owner enrollment and thus could degrade users' experience in interacting with the authentication system.

To address this issue, we propose a new authentication mechanism for OcuLock to utilize the distinctive features. After features are extracted from the EOG signals, a comparison algorithm is adopted to compare each feature of the input EOG with that of the owner's EOG. A matching score indicating the similarity and ranging from 0 to 1 is generated for each feature. The resulting matching scores for all features are fed to a comparator to determine whether the input EOG matches the owner's EOG, i.e., whether the current user is the owner. If the matching scores of all features are high, the input record is determined to be from the owner.

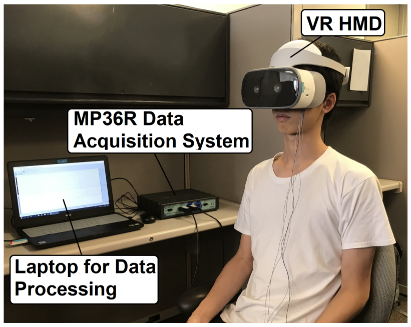


Fig. 10: Experiment setup.

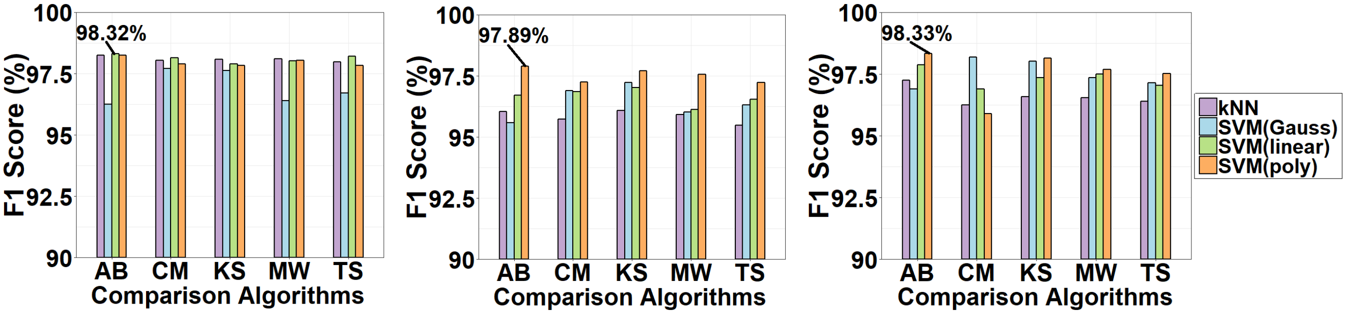
Similar to previous methods, this procedure stores the features, or a template, of the owner. However, it does not require repeated classifier training for each enrolled owner. The comparison algorithm can accept the features of any input user and any owner and gauge the similarity. Therefore, only one comparator that makes the authentication decision based on a set of matching scores needs to be trained. This mechanism significantly reduces enrollment complexity and improves system usability.

As we will show in Section VIII, the choice of the comparison algorithm and the machine learning model to build the comparator affect the authentication performance. Hence, it is important to select the optimal comparison algorithm and comparator model.

VII. Experiment Setting

Apparatus. As shown in Figure 10, the prototype of OcuLock consists of a VR HMD, a EOG acquisition device, and a laptop. The VR HMD is a Lenovo Mirage Solo, the first standalone VR HMD powered by Google Daydream [18]. The EOG acquisition device is a BIOPAC MP36R that measures EOG at a sampling rate of 200 Hz via five Ag-AgCl series lead electrodes. The Dell Inspiron 5577 laptop with a 2.8 GHz CPU is connected with the acquisition device for processing the signal records. The authentication decision is then sent back to VR HMD. Our proof-of-concept prototype adopts a separate data acquisition and processing device for the purpose of software compatablity. However, we point out that integrated device including above three component is already commercially available [32]. Therefore, our prototype design does not decrease the potential of OcuLock.

Subjects. We recruited 70 subjects (27 females and 43 males, age from 19 to 32) through public advertisements and email lists. All participants are university students. Among these subjects, 24 of them wore glasses and they did not remove their glasses during the experiments. 23 subjects have



(a) Fixed-Route; (b) City-Street; (c) Illusion.

Fig. 11: F1 scores for three stimuli using different comparison algorithms and comparator models.

used VR before while 47 never used it. Subjects were told that their EOG will be recorded to extract unique features and differentiate themselves from others. They signed a written consent form in accordance with an existing IRB approval we hold which allows for recording EOG and other responses from human subjects for user authentication and VR system evaluation. A subject sat in a chair in a relaxed posture and wear the VR HMD with electrodes to view the three visual stimuli. Five electrodes were fixed on the HMD cover, the electrode positions on different participants were the same. Each stimulus was viewed for 10 seconds and the corresponding EOG was collected. Each subject viewed the 3-stimuli session for 10 times and a total of 700 EOG records were generated for each stimulus.

Training and Testing Procedure. OcuLock uses a new record-comparison based scheme for authentication decision. To generate training and testing data for the decision-making comparator, the subjects are randomly divided into two halves for training and testing. For the 35 subjects for training, any two records of the 350 records are compared to generate 61,075 samples as the training data. Each sample indicates whether or not the two records are from the same person. A total of 1,575 samples are from comparison between the same subject, i.e., positive samples, while the others are negative samples. Similarly, the testing set also has 61,075 samples and is used for model evaluation. We repeat the above procedure for 10 times and report the average results in the following.

Evaluation Metrics. While accuracy is a popular way to evaluate a machine learning model, the unbalanced composition of our testing data could generate misleading accuracy. A comparator could achieve 97% accuracy even if it predicts all sample as negative. We instead use Equal Error Rate (EER) and F1 score that have been widely used in authentication systems. The EER is the rate when the false acceptance rate (FAR) and the false rejection rate (FRR) are equal.

VIII. Reliability Evaluation Results

In this section, we discuss the reliability of our system under different impact factors such as the authentication duration, the subset of selected features and the performance degradation over time.

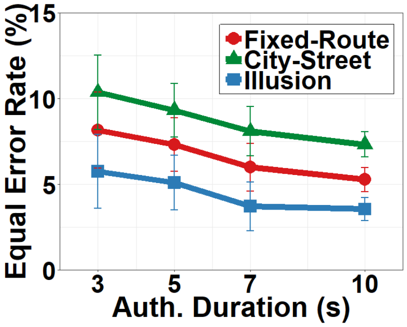


Fig. 12: The EERs using different authentication duration.

A. *Choices of Algorithm and Model*

To ensure that OcuLock achieves its best performance, we test different comparator models including k-nearest neighbors algorithm (kNN), a Support Vector Machine (SVM) using the Gaussian radial basis function as the kernel, an SVM using a linear kernel, and an SVM using a polynomial (poly) kernel. Multiple comparison algorithms including Ansari-Bradley Test (AB), Two-Sample Cramer-von Mises Test (CM), Two-Sample Kolmogorov-Smirnov Test (KS), Mann-Whitney U-Test(MW), and Two-Sample t-test (TS) [20] are also tested. Figure 11 shows the F1 scores for each combination of comparison algorithm and comparator model. The F1 scores reach $\sim 98\%$ due to the unique and comprehensive features considered in OcuLock. We also observe that AB Test achieves better performance. This is because many proposed features are distributions rather than scalar numbers. AB Test can capture the shape information between two distributions and thus characterize each user's EOG more accurately. We herein select the optimal combination for the remaining evaluations.

B. *Time Efficiency*

A practical authentication system should be able to accurately identify the user within an acceptable amount of time. To study the impacts of authentication duration on the performance, we repeat the experiment procedure described in Section VII for all 70 subjects with the viewing duration for stimuli changed to 3, 5 and 7 seconds. The EER results are demonstrated in Figure 12. Using 10-second records, the three stimuli reaches EERs of 5.27%, 7.32% and 3.55% with

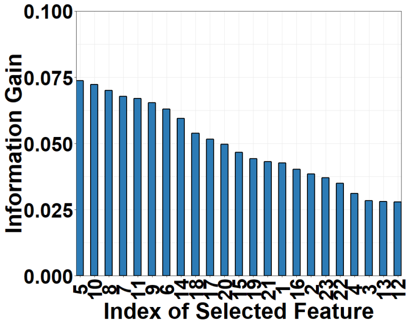


Fig. 13: The information gain of each included feature.

standard deviations of 1.41%, 1.48% and 1.34%, respectively. Decreasing duration slowly increases the EERs. For example, the 3-second authentication achieves an EER of 5.75% for Illusion stimulus. This result suggests a small trade-off between convenience and security.

We also observe that among all three stimuli Illusion achieves the best performance because it elicits more microsaccades and blinks, as well as extraocular cell and muscle activities. To evaluate other impact factors of OcuLock, we use Illusion as an example in the following Section VIII-C and VIII-D.

C. Feature Selection

Feature selection helps identify the important features to reduce the computation complexity and overfiting of the comparator. To verify the impact of feature selection on the comparator performance, we apply minimum redundancy maximum relevance feature selection algorithm (mRMR) [9] to select highly related features while minimizing the interdependence between selected features. At first, the most contributing feature is selected. Then in each round, another feature that enhances the model the most is added to the feature subset. Each time, the authentication is executed on the selected feature set and the corresponding information gain is calculated.

Information Gain. Figure 13 reports the information gain of each feature included for Illusion by mRMR feature selection algorithm. The X axis lists feature indexes as defined in Table I. These features are ranked from the most important on the left to the least important on the right. We observe that the top 5 features are Metabolism Intensity (V), Sympathetic Energy (H), Extent of Up Rotation Distance, Extent of Left Rotation Distance, Sympathetic Energy (V), and Extent of Down Rotation Distance while the best behavioral feature is ranked the 8th. This clearly shows the importance of low-level HVS biostructure in identifying users compared to traditional eye gaze.

Receiver Operating Characteristic (ROC). To investigate the effect of behavioral and physiological features on the performance of the system, we repeat the mRMR algorithm separately on each feature group and report the ROC for FAR and FRR in Figure 14. The area-under-curve (AUC) values for models using behavioral features, physiological features and both categories are 87.59%, 95.43%, and 98.31% respectively. Physiological features perform significantly better (smaller FAR) than behavior features especially when FRR is small. This indicates when the system gives more "accept" decisions (smaller FRR), the model using behavioral features quickly becomes misjudging and accepts wrong users. This is because behavioral features are less distinctive than physiological features and thus more acceptance results in higher FAR.

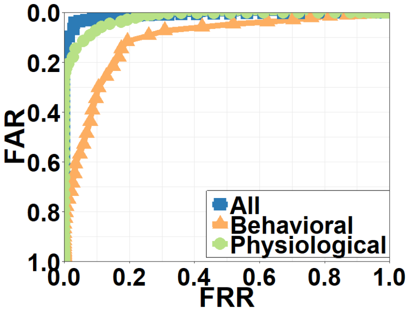


Fig. 14: ROC curves for three types of feature subsets.

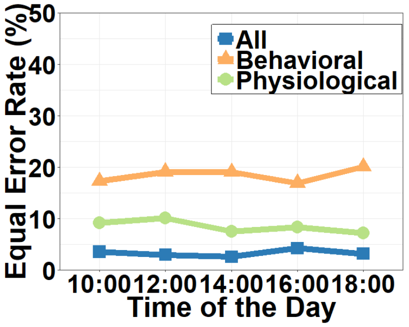


Fig. 15: The EERs are stable in the one-day short-term period.

From the results above, we conclude that the proposed physiological features of HVS play the essential role in differentiating users. Besides, traditional behavioral features based on saccades and fixations is less important but still contributes to the authentication. Therefore, both feature categories should be used in OcuLock.

D. Short-term and Long-term Performance

The physical and mental states of human users change over time. In this section, we investigate the impacts of time on authentication performance. We first conduct a short-term study at different time of a day. We aim to evaluate the impacts of eye fatigue and strains. Five subjects (Two subjects were group members and the rest three were from the general public) were recruited for this experiments. Since they worked on their personal computers extensively during the day, it is expected that their eye fatigue and strain increase with time. We started the first experiment at 10 AM and continued four more experiments until 6 PM. Samples from the 10 AM experiment were set as original. We used the optimal comparator trained in Section VIII-A (AB Test and SVM-poly) to continuously compare samples from later experiments with original ones. Figure 15 illustrates the one-day short-term EER for three models trained using physiological features, behavioral features and both physiological and behavioral

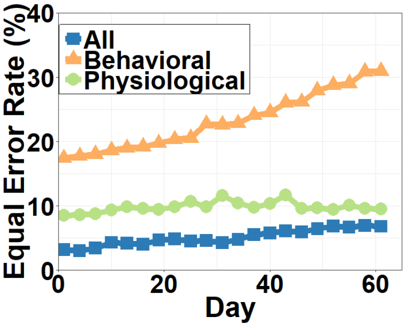


Fig. 16: The EERs over a 2-month long-term period. The models including physiological features is much more stable than the model only using behavioral features.

features (as categorized in Table I). It is interesting to observe that there is no significant fluctuation of EERs. This suggests the negligible impacts of eye fatigue and short-term cognitive states on our HVS-based authentication. The average EER for the model only using physiological features is 47% less than the average EER for the model only using behavioral features. Combining both feature categories further reduces EERs during the day down to 3.71%. This indicates that low-level HVS features, especially HVS biostructure, can be uniquely triggered in VR HMD, and outperform traditional eye gaze behavior features.

One weakness of gaze biometric is that the performance degrades quickly as time passes by because eye movement is highly dependent on the cognitive states and gaze patterns constantly change. We performed a long-term over a two-month period to investigate the stability of OcuLock. The first set of records collected from the recruited 5 users were set as original. Then we kept collecting records from these subjects once every three days and evaluated the consistency of authentication results. The EERs results are demonstrated in Figure 16. EERs for the model using physiological features slowly increase from 8.51% at the first day to 10.64% at the 25th day and remain stable after that. In contrast, EERs for the model using behavioral features quickly and continuously raise from 17.42% all the way up to 30.96%. Two subjects were our group members and their knowledge of the system may bias performance stability. However, we did not observe noticeable difference in performance stability between these group members and others. This is because the biostructural information represented by physiological features are less susceptible to change over time, and it cannot be controlled by subjects. In OcuLock, we utilize this fact to achieve a far more stable EER performance than existing eye gaze biometric by combining both feature categories. Since using both feature categories still slowly increases EER (3.17% to 6.18% over 2 months), VR users can strike a tradeoff between EOG update frequency and authentication accuracy. When users cannot accept the accuracy after 2 or more months, they can record new EOG ground truth samples based on their current physiological and behavioral features.

IX. Security Analysis

In this section, we investigate the security of the system against two types of attacks discussed in Section III: impersonation attack and statistical attack.

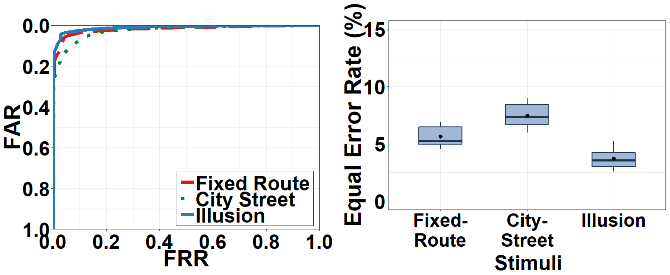


Fig. 17: The ROC curves (left) and EERs (right) under impersonation attack for three stimuli.

A. *Impersonation Attack*

Impersonation attack is the most simple and popular type of attacks in which the adversaries are not required to have prior knowledge about the legitimate user's HVS information. They simply try to be the owner and follow the standard authentication procedure. Since OcuLock is unobservable, adversaries also cannot obtain much information to assist the attacks by observing owner's authentication action. What they can see are only minor head and body motions. To emulate impersonation attacks, one user is assigned as the legitimate user and the others become attackers to perform "attack attempts". This process is repeated for all users in the test set. The system is then evaluated using ROC curves, EER and F1 scores.

ROC. Figure 17 reports how well our system performs against the impersonation attack using ROC curves. During the authentication process, the comparator of OcuLock produces probability values indicating chances that the input signal belongs to the owner or the attacker. An authentication decision is then produced by comparing probability values with a predefined threshold. The ROC could be built by varying this threshold and recording the false acceptance rate and false rejection rate. The ROC gives an overall picture about the system security at every comparator threshold.

In Figure 17, the AUC values for ROC curves corresponding to three stimuli are 97.62%, 96.08% and 98.31%. For all stimuli, the ROC curves of the comparator stay closely to the top-left corner where both FAR and FRR are minimized. The derived AUC values are close to the 1.0 mark even though the curve for City-Street stays a little lower than the others. The shape of the ROC curves and the AUC values indicates OcuLock performs well in terms of false acceptance and rejection for all model thresholds.

EER and F1 score. EER values are depicted in Figure 17, where City-Street shows an EER of 7.32% with STD of 1.48%, while the EER of Fixed-Route is 5.27% with STD of 1.41% and Illusion is 3.55% with STD of 1.34%. Similarly, we calculate the F1 scores for these stimuli and find a strong correlation to EER, i.e., 98.32%, 97.89% and 98.33% for Fixed-Route, City-Street and Illusion respectively.

The resistance to impersonation attacks depends on the stimuli, i.e., it is better in Illusion but a little worse in City-Street. This is due to the fact that the HVS activities when viewing different stimuli are different. The City-Street allows subjects to freely scan through the picture, where subjects tend to conduct "smooth pursuit eye movements" whose speed is voluntarily controlled and not reflecting extraocular muscle

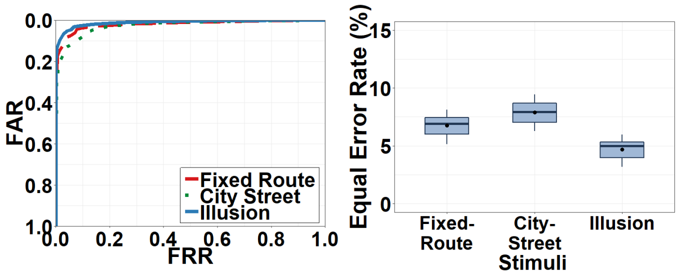


Fig. 18: The ROC curves (left) and EERs (right) under statistical attack for three stimuli.

conditions. Also after scanning the entire picture, subjects would reduce eye movements and stare at one point, which generate more silent segments in their EOG signal records. With less information conveyed by the records, it is harder for the comparator to differentiate subjects. Thus City-Street has the lowest F1 scores and highest EER. On the contrary, the Fixed-Route and Illusion both require subjects to continuously follow a dynamic target jumping at a given speed allowing more HVS activities to be produced. Hence their EOG signal records consistently contain more low-level HVS information and are easier to distinguish.

B. *Statistical Attack*

Statistical attack is a stronger form of attack in which the adversaries are assumed to have some knowledge about the statistics from a group of users. The attacker calculates the probability density function of features from users and then use the most probable feature values to generate the forgery.Statistical attacks have been performed for behavior biometrics [48], [32], [33], [34].

To simulate statistical attack, we assume the attacker gains insights into the statistics of all enrolled users in the system (but not the exact record of the victim). The attackers are also able to forge fake EOG signals with desired feature values. We then follow the procedure suggested by [48] to generate forged features from real feature values. First, the attackers reconstruct a histogram approximating the distribution of the values of one given feature from all users. Each histogram comprises 5 bins of equal size. Fake values of the feature are then created by sampling uniformly from the bin in which the feature value has the highest probability. This procedure is repeated for all features to create one fake sample. We performed the statistical attack using the fake records and records from all 70 subjects. During the test, all the subject's records were compared with each other, which generated 45 positive samples. Meanwhile her records were attacked by the fake records, which generated 1500 negative samples. With 70 subjects, the testing set contained 3,150 positive samples and 10,5000 negative samples. Then the ROC curves, EERs and F1 scores were recorded.

ROC. Figure 18 reports the FARs and FRRs for the model at various thresholds. The AUC for curves from Fixed-Route, City-Street and Illusion are 96.11%, 94.78% and 96.23% respectively. In all stimuli, the curves stay close to the upper-left corner where both FAR and FRR are minimized suggesting the resistance of the model against statistical attack. The AUC score for City-Street is lower than the other two, which is consistent with our previous conclusion about the uncertainty nature of this stimulus. The AUC score for statistical attack is lower than impersonation attack by a small amount suggesting this type of attack is stronger but does not severely affect the model performance.

TABLE II: User Feedback for OcuLock.

	Questions	Score (1-5)	STD
Q1	How comfortable were you when watching Fixed Route?	4.2	1.1
Q2	How comfortable were you when watching City Street?	3.5	1.8
Q3	How comfortable were you when watching Illusion?	4.4	0.8
Q4	Your acceptable duration?	8 sec	2.3
Q5	Your preferred duration?	4.1 sec	2.3

EER and F1 score. Figure 18 reports the EER of the model. As expected, Illusion has the lowest EER at 4.97% while the EER for City-Street is the highest at 7.93% due to the random exploration behavior in this stimulus. The EERs in all stimuli are on average 1.08% higher than EERs from the Impersonation attack. The comparator model attains F1 scores of 97.62%, 96.59%, 97.55% in Fixed-Route, City-Street, and Illusion respectively, which is on average 0.92% lower than F1 scores for models under impersonation attack.

Compared with the reported results in impersonation attack, the performance for the comparator model under statistical attack are worse suggesting statistical attack are more powerful. However, our authentication system still achieve promising scores in both F1 score and EER. We surmise that the resistance of our model against statistical attack is because the feature values among different people spread widely, which makes it harder to predict feature values for a specific person based on the statistical information.

X. User Experience

Gathering subjects' opinions towards system usability is necessary as an authentication system requires extensive interaction with users. To gather the subjects' opinions towards OcuLock, all 70 subjects participating in the experiments were given paper-form questionnaires in person immediately after the experiments. The questionnaire includes several questions on subjects' experience of the authentication system. Each subject was asked to filled out the questionnaire and the answers were scored from 1 (worst) to 5 (best).

Table II quantifies the subjects' feelings towards OcuLock. Most users preferred stimuli with explicit moving targets to follow (Q1, Q3) over complicated static scene where they have to subjectively scan through (Q2). This indicates the desirability to design stimuli with dynamic and attractive patterns. In addition, subjects preferred the duration of authentication to be at around 4 seconds and should not be more than 8 seconds. Our system can provide acceptable performance within this time frame, as discussed in Section VIII-B.

We further conducted another survey of the subject's opinions towards several potential authentication methods for VR HMD. After describing all methods to subjects in detail, opinion scores were recorded on four criteria: security, reliability,

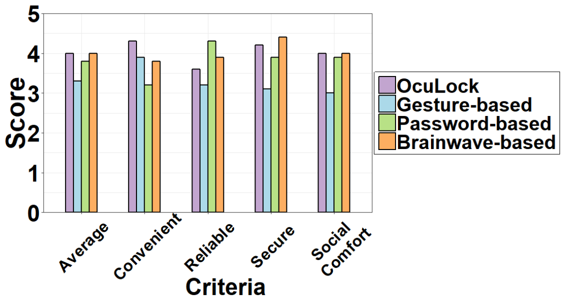


Fig. 19: User feedback about preferred authentication method.

convenience and social comfort (comfortable to use in the public). Figure 19 reports the results among four potential approaches including HVS-based (OcuLock), gesture-based [34], password-based [17] and brainwave-based [29]. Most subjects agreed that HVS-based authentication is more convenient and socially acceptable in the public. HVS-based authentication is also more secure as it is more resistant to observation-based attacks than password-based and gesture-based approaches. In regard to reliability, most subjects consider the password-based authentication as the best choice. However, we conjecture that the high reliability score for password-based methods is from its proven practical usage. Considering the high stability results reported in Section VIII, the subjects' reliability score for OcuLock can be boosted if they are given first hand experiences of our system in a long term.

XI. Related Work

HMD Authentication. Early works of HMD authentication have been focusing on smart glasses for augmented reality. Chauhan et al. [6] proposed a gesture-based authentication system for Google Glass that collects authentication input on the touchpad. Li et al. [28] developed an head movement biometric system in response to auditory stimuli. Similarly, unique head movement was also triggered by a set of pictures to authenticate smart glasses users [44]. Recently, efforts have been made towards VR HMD authentication by migrating traditional authentication methods. Oculus Quest is the first commercial VR HMD equipped with virtual PIN code [36]. George et al. [17] studied the security and usability of authentication methods such as PIN and unlock pattern in VR HMD through remote controllers. Graphical password [13] and body motion biometric [40] were also proposed to authenticate VR users. Although these traditional methods can achieve acceptable error rate, the entire authentication action is exposed to the public and can be observed by adversaries to execute attacks. For example, adversaries managed to observe the authentication action and either mimic the owner's behavior for impersonation attacks or analyze the behavior for side-channel attacks [30], [14]. Although extracting brain signals for VR HMD authentication is relatively secure [29], collecting brainwaves requires a device covering the majority of the scalp via a large number of electrodes. Such a setup is too cumbersome for practical use and is not compatible with the form factor of today's VR HMDs that only has one holding strap [37], [27]. In this paper, we utilize the fact that VR HMD fully covers users' eye area to propose an HVS-based biometric for unobservable authentication. Since the foam face cover has direct contact to skin around eye sockets, we design a usable system only using 5 embedded electrodes to collect the HVS signals through EOG.

Eye-based Authentication. Existing eye-based authentication systems were designed for smartphones and PCs and focused on the uniqueness of eye globe movement. Most of them collected the gaze pattern stimulated by visual stimuli via video cameras or other types of eye trackers and then crafted a set of features for authentication [3], [23], [19], [47], [15]. These features include scanpath and more detailed statistics such as acceleration and duration of saccades [12]. A system using reflexive eye movements was developed to enable fast authentication [45]. Saccades and fixations triggered by implicit stimuli during nornal viewing were utilized to support continuous eye-based authentication [56]. EOG was also used to identify the saccades and fixations to differentiate users [22], [1]. While these systems build the foundation for OcuLock, they only focused on the gaze pattern in using smartphones and PCs and the support of long-term performance stability has not be validated. In this paper, we explore HVS as a whole to utilize low-level physiological and behavioral features of HVS triggered by immersive VR content and leverage the stability of HVS biostructure to achieve a low variability over a two-month evaluation.

EOG Applications. EOG has been widely used for human computer interaction. Barea et al. [2] designed a wheelchair system controlled by eye movements collected from EOG signals. Chen et al. [7] developed EOG-based interfaces to control robots. Qvarfordt et al. [42] proposed a system that explores users' interest based on eye-gaze patterns. Ding et al. [10] implemented a human-computer interface using EOG signals as input. These works validate the promise of adopting EOG in personal computing systems. However, they do not harness EOG as a source of information that presents unique features of HVS. Instead, we take the first step to utilize EOG to explore the whole HVS for VR HMD authentication.

XII. Discussion

A. Advanced Attacks

Replay attack is a common attack for biometric authentication system. In OcuLock, the authentication action is fully covered by HMD. It is unlikely, if not impossible, for an attacker to directly record EOG signals and replay the record as what have been done in replay attacks for voice or face biometric. In some cases, stronger adversaries might obtain the owner's EOG templates, e.g., by measuring electromagnetic emanations while the system is processing critical information. To spoof the system, the attackers can use two ways to feed the stolen template into the EOG sensing system. First, the attackers leverage the voltage generators to produce the exact same EOG signals of the owner according to the template. The generated signals are directly sent to the EOG electrodes through wire connections. However, considering the ever-growing lightweight sensors on modern VR headsets, we can prevent it by adopting existing sensing-based or learning-based liveness detection methods [50], [35]. Second, if the attackers are even aware of the liveness detection, the upgraded attacking method can be building the artificial eyes that contain all HVS functionalities. However, such an artificial eye is currently unavailable in the market and building it from scratch is indeed non-trivial. Hence we do not consider it as a typical attack.

B. Computation Time

To guarantee the usability of an authentication system, computation time is one of the major concerns. The total computation time of OcuLock consists of three parts: the EOG recording time when users view the stimuli, the signal processing time when the EOG is transformed, and the authentication time when features are extracted, the comparison algorithm is run and the comparator model is executed. According to our measurement, the signal processing takes less than 1 ms and the authentication takes an average of 39 ms. The EOG recording time ranges from 3 seconds to 10 seconds as shown in the experiment results. We can see that the total computation time is dominated by the EOG recording time while other time components are negligible because we use efficient algorithms to design our system. We even reuse several intermediate results, e.g., reusing wavelet transform results for sympathetic energy.

It is true that authentication systems in a physical world usually take less time (1-2 seconds) than OcuLock. However, VR interaction is generally slower since it relies on head and/or eye navigation in a virtual world, which is harder than physical-world interaction. User studies showed that the simplest authentication such as PIN code or unlock patterns takes around 3 seconds in VR environment [17]. Therefore, users generally have lower expectation on VR authentication and thus we believe the 3-second authentication time of OcuLock is acceptable. Depending on the tradeoff between authentication error and computation time, users can select a proper EOG recording time for OcuLock.

The memory consumption for the entire authentication is on average 54 MB, which is acceptable in modern VR computing devices.

C. Electrode Placement

In our prototype, we applied conductive gel inside each electrode to help measure EOG signals. However, the gel does not need to be replaced frequently (once every 30 minutes). For future real-world systems, dry electrodes can be used for EOG collection to enhance system usability. This technique has been used by JINS MEME, a commercial smart glasses device [32].

After finishing each session in our experiment, the electrodes were taken off from one subject and attached to another. We point out that it is not this replacement of electrodes that results in the different EOG samples between subjects. In the majority of the evaluation, electrodes were fixed at the HMD cover and thus their positions for different participants were the same (see Figure 5). Even though there is minor placement difference in some cases, we found that EOG measurement is not sensitive to that. In our temporal study (Section VIII-D), the electrodes were detached and attached to the HMD repeatedly with around 1cm of position change. However, the system could still recognize users although electrode positions are not the same, which proves the negligible effects of electrodes position.

XIII. Conclusion

In this paper, we present OcuLock, a stable and unobservable system to authenticate users for VR HMD. Compared with eye gaze based systems, we explore HVS as a whole and extract low-level physiological and behavioral features for biometric authentication. OcuLock is resistant to common and anticipated types of attacks such as impersonation and statistical attacks with EERs of 3.55% and 4.97% respectively. Thanks to the stable physiological features, OcuLock is less variable over time and reduces the frequency of updating EOG template. Our user study suggests promising potential for HVS-based authentication in which the requirement of convenience, security and social comfort can simultaneously be satisfied. Future work should focus on integrate the devices in our prototype into a unified VR HMD for more practical and larger-scale user study.

Acknowledgment

This work was supported in part by National Science Foundation Grant CCF-1852516, CNS-1718375, and National Natural Science Foundation of China Grant 61972348.

References

[1] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, “A novel biometric approach for human identification and verification using eye blinking signal,” *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 876–880, 2014.

[2] R. Barea, L. Boquete, M. Mazo, and E. López, “System for assisted mobility using eye movements based on electrooculography,” *IEEE transactions on neural systems and rehabilitation engineering*, vol. 10, no. 4, pp. 209–218, 2002.

[3] R. Bednarik, T. Kinnunen, A. Mihaila, and P. Fränti, “Eye-movements as a biometric,” in *Scandinavian conference on image analysis*. Springer, 2005, pp. 780–789.

[4] M. Brown, M. Marmor, E. Zrenner, M. Brigell, M. Bach *et al.*, “Iscev standard for clinical electro-oculography (eog) 2006,” *Documenta ophthalmologica*, vol. 113, no. 3, pp. 205–212, 2006.

[5] A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, “Eye movement analysis for activity recognition using electrooculography,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 4, pp. 741–753, 2010.

[6] J. Chauhan, H. J. Asghar, A. Mahanti, and M. A. Kaafar, “Gesture-based continuous authentication for wearable devices: The smart glasses use case,” in *International Conference on Applied Cryptography and Network Security*. Springer, 2016, pp. 648–665.

[7] Y. Chen and W. S. Newman, “A human-robot interface based on electrooculography,” in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA’04. 2004*, vol. 1. IEEE, 2004, pp. 243–248.

[8] V. Di Lollo, J.-i. Kawahara, S. S. Ghorashi, and J. T. Enns, “The attentional blink: Resource depletion or temporary loss of control?” *Psychological research*, vol. 69, no. 3, pp. 191–200, 2005.

[9] C. Ding and H. Peng, “Minimum redundancy feature selection from microarray gene expression data,” *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

[10] Q. Ding, K. Tong, and G. Li, “Development of an eog (electro-oculography) based human-computer interface,” in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006, pp. 6829–6831.

[11] A. T. Duchowski, “Eye tracking methodology,” *Theory and practice*, vol. 328, no. 614, pp. 2–3, 2007.

[12] S. Eberz, K. Rasmussen, V. Lenders, and I. Martinovic, “Preventing lunchtime attacks: Fighting insider threats with eye movement biometrics,” in *Proceedings of Network and Distributed System Security Symposium*. NDSS, 2015.

[13] M. Funk, K. Marky, I. Mizutani, M. Kritzler, S. Mayer, and F. Michahelles, “Lookunlock: Using spatial-targets for user-authentication on hmds,” in *CHI Conference on Human Factors in Computing Systems Late Breaking Work*, 2019.

[14] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof attacks on gait authentication system," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 491–502, 2007.

[15] C. Galdi, M. Nappi, D. Riccio, and H. Wechsler, "Eye movement analysis for human authentication: a critical survey," *Pattern Recognition Letters*, vol. 84, pp. 272–283, 2016.

[16] N. Garun, "Amazon will soon refund up to $70 million of in-app purchases made by children," 2019, http://www.theverge.com/2017/4/4/15183254/amazon-ends-appeal-refund-70-million-in-app-purchases.

[17] C. George, M. Khamis, E. von Zezschwitz, M. Burger, H. Schmidt, F. Alt, and H. Hussmann, "Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality," in *Proceedings of Network and Distributed System Security Symposium*. NDSS, 2017.

[18] Google, "Introducing daydream standalone vr headsets," https://vr.google.com/daydream/standalonevr/.

[19] C. Holland and O. V. Komogortsev, "Biometric identification via eye movement scanpaths in reading," in *2011 International joint conference on biometrics (IJCB)*. IEEE, 2011, pp. 1–8.

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

[21] M. Joukal, *Anatomy of the Human Visual Pathway*. Springer, 04 2017, pp. 1–16.

[22] M. Juhola, Y. Zhang, and J. Rasku, "Biometric verification of a subject through eye movements," *Computers in biology and medicine*, vol. 43, no. 1, pp. 42–50, 2013.

[23] T. Kinnunen, F. Sedlak, and R. Bednarik, "Towards task-independent person authentication using eye movement signals," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 2010, pp. 187–190.

[24] C. Krapichler, M. Haubner, R. Engelbrecht, and K.-H. Englmeier, "Vr interaction techniques for medical imaging applications," *Computer methods and programs in biomedicine*, vol. 56, no. 1, pp. 65–74, 1998.

[25] D. Kumar and E. Poole, "Classification of eog for human computer interface," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*, vol. 1. IEEE, 2002, pp. 64–67.

[26] T. B. Kuo and C. C. Yang, "Frequency domain analysis of electrooculogram and its correlation with cardiac sympathetic function," *Experimental neurology*, vol. 217, no. 1, pp. 38–45, 2009.

[27] Lenovo, "Mirage solo with daydream," https://www.lenovo.com/us/en/daydreamvr/.

[28] S. Li, A. Ashok, Y. Zhang, C. Xu, J. Lindqvist, and M. Gruteser, "Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns," in *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2016, pp. 1–9.

[29] F. Lin, K. W. Cho, C. Song, W. Xu, and Z. Jin, "Brain password: A secure and truly cancelable brain biometrics for smart headwear," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2018, pp. 296–309.

[30] Z. Ling, Z. Li, C. Chen, J. Luo, W. Yu, and X. Fu, "I know what you enter on gear vr," in *2019 IEEE Conference on Communications and Network Security (CNS)*, June 2019, pp. 241–249.

[31] B. R. Manor and E. Gordon, "Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks," *Journal of neuroscience methods*, vol. 128, no. 1-2, pp. 85–93, 2003.

[32] J. MEME, "Jine meme eye sensing," 2019, https://jins-meme.com/en/products/es/.

[33] A. Nguyen, Z. Yan, and K. Nahrstedt, "Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction," in *2018 ACM Multimedia Conference on Multimedia Conference*. ACM, 2018, pp. 1190–1198.

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

[35] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado, "Fingerprint liveness detection using convolutional neural networks," *IEEE transactions on information forensics and security*, vol. 11, no. 6, pp. 1206–1213, 2016.

[36] Oculus, "Getting started with your oculus quest," https://support.oculus.com/855551644803876/.

[37] ——, "Our first all-in-one gaming headset," https://www.oculus.com/quest/.

[38] ——, "Rift store: VR games, apps & more," https://www.oculus.com/experiences/rift/.

[39] K. Pfeil, E. M. Taranta II, A. Kulshreshth, P. Wisniewski, and J. J. LaViola Jr, "A comparison of eye-head coordination between virtual and physical realities," in *Proceedings of the 15th ACM Symposium on Applied Perception*, 2018, p. 18.

[40] K. Pfeuffer, M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt, "Behavioural biometrics in vr: Identifying people from body motion and relations in virtual reality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 110.

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

[42] P. Qvarfordt and S. Zhai, "Conversing with the user based on eye-gaze patterns," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2005, pp. 221–230.

[43] H. L. Ramkumar, "Electrooculogram," 2019, https://eyewiki.aao.org/Electrooculogram#Testing_process.

[44] C. E. Rogers, A. W. Witt, A. D. Solomon, and K. K. Venkatasubramanian, "An approach for user identification for head-mounted displays," in *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 2015, pp. 143–146.

[45] I. Sluganovic, M. Roeschlin, K. B. Rasmussen, and I. Martinovic, "Using reflexive eye movements for fast challenge-response authentication," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 1056–1067.

[46] V. R. Society, "Virtual reality air force training," https://www.vrs.org.uk/virtual-reality-military/air-force-training.html.

[47] C. Song, A. Wang, K. Ren, and W. Xu, "Eyeveri: A secure and usable approach for smartphone user authentication," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[48] Y. Song, Z. Cai, and Z.-L. Zhang, "Multi-touch authentication using hand geometry and behavioral information," in *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2017, pp. 357–372.

[49] R. Steinberg, R. Linsenmeier, and E. Griff, "Retinal pigment epithelial cell contributions to the electroretinogram and electrooculogram," *Progress in retinal research*, vol. 4, pp. 33–66, 1985.

[50] H. Steiner, S. Sporrer, A. Kolb, and N. Jung, "Design of an active multispectral swir camera system for skin detection and face verification," *Journal of Sensors*, 2016.

[51] Viar360, "Virtual reality market size in 2018 with forecast for 2019," 2019, https://www.viar360.com/virtual-reality-market-size-2018/.

[52] A. Walker, "Potential security threats with virtual reality technology," 2017, https://learn.g2.com/security-threats-virtual-reality-technology.

[53] J.-G. Wang and E. Sung, "Study on eye gaze estimation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 32, no. 3, pp. 332–350, 2002.

[54] Wikipedia, "Virtual reality therapy," https://en.wikipedia.org/wiki/Virtual_reality_therapy.

[55] J. Yi, S. Luo, and Z. Yan, "A measurement study of youtube 360 live video streaming," in *Proceedings of the 29th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*. ACM, 2019, pp. 49–54.

[56] Y. Zhang, W. Hu, W. Xu, C. T. Chou, and J. Hu, "Continuous authentication using eye movement response of implicit visual stimuli," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. 177, 2018.