# Medusa3D: The Watchful Eye Freezing Illegitimate Users in Virtual Reality Interactions

AOCHEN JIAO*, DI DUAN*, and WEITAO XU†, City University of Hong Kong, China

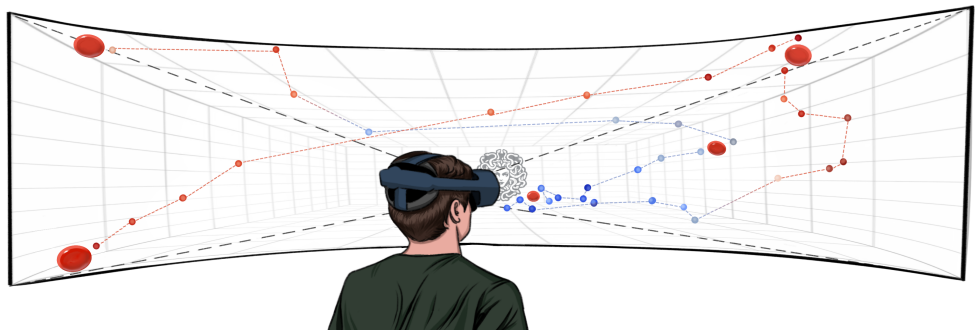


Fig. 1. Medusa3D offers interactive 3D visual stimuli within a VR environment, designed to elicit reflexive eye movements among users. By analyzing these eye responses, Medusa3D authenticates users and freezes unauthorized individuals.

The remarkable growth of Virtual Reality (VR) in recent years has extended its applications beyond entertainment to sectors including education, e-commerce, and remote communication. Since VR devices contain user's private information, user authentication becomes increasingly important. Current authentication systems in VR, such as password-based or static biometric-based methods, are either cumbersome to use or vulnerable to attacks such as shoulder surfing. To address these limitations, we propose Medusa3D, a challenge-response authentication system for VR based on reflexive eye responses. Unlike existing methods, reflexive eye responses are involuntary and effortless, offering a secure and user-friendly credential for authentication. We implement Medusa3D on an off-the-shelf VR and conduct evaluations with 25 participants. The evaluation results show that Medusa3D achieves 0.21% FAR and 0.13% FRR, demonstrating high security under various ocular conditions and resilience against attacks such as zero-effort attack, replay attack, and mimicry attack. A user study indicates that Medusa3D is user-friendly and well-adopted among participants.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Security and privacy** → **Authentication**.

Additional Key Words and Phrases: VR, user authentication, gaze

*Both co-primary authors contributed equally to this paper.
†Weitao Xu is the corresponding author.

Authors' address: Aochen Jiao, aochen.jiao@cityu.edu.hk; Di Duan, dduan5-c@my.cityu.edu.hk; Weitao Xu, weitaoxu@cityu.edu.hk, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong SAR, China, 999077.


ACM 2573-0142/2024/9-ART270
https://doi.org/10.1145/3676515

**ACM Reference Format:**
Aochen Jiao, Di Duan, and Weitao Xu. 2024. Medusa3D: The Watchful Eye Freezing Illegitimate Users in Virtual Reality Interactions. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 270 (September 2024), 21 pages. https://doi.org/10.1145/3676515

## 1 INTRODUCTION

Virtual Reality (VR) is a technology that simulates a computer-generated environment and allows people to interact with it in a very realistic way. It has become increasingly popular in recent years, with applications in fields such as education [62], healthcare [40], e-commerce [15] and remote communication [65]. Similar to mobile devices, VR devices contain crucial personal sensitive information such as bank account and location data. Therefore, user authentication is critical for the security and data protection of VR devices, as it ensures the identity of the users and can prevent unauthorized or malicious access.

However, current user authentication methods in VR devices still adopt the traditional authentication methods that are used in mobile devices such as passwords, digital PINs, and pattern lock. Since VR devices have no keyboard or touchscreen, users have to use hand controllers to input the credentials, which is cumbersome and inefficient [77]. Moreover, these methods are vulnerable to shoulder-surfing attack because attackers can easily observe the user's hand movements and recover the credentials [35, 77]. To address the problems, biometric-based authentication has emerged as a viable method for VR devices. Recent works have studied a variety of biometrics, including muscle [23], iris [84], and head [48, 82]. Despite these efforts, several limitations still exist which are summarized below:

- **Low user experience.** A line of previous methods need to present stimuli to users. The commonly used stimuli include electric [23], vibration [46], and acoustic reverberation [82]. These stimuli are not natural for user interaction with VR devices and will decrease the immersive experience. Furthermore, such stimuli can even be harmful to health in the long term [53, 86].
- **Low security.** A large body of recent systems are vulnerable to various attacks. For example, static physical biometrics such as iris are vulnerable to replay attack [57]. A number of works [54, 61, 83] use body motion behaviors as active features for user authentication. Users need to perform specific tasks to complete authentication, such as grabbing, typing, walking, and nodding. However, these tasks are pre-defined and attackers can easily learn and mimic the legitimate user's behaviors through pre-recorded video.
- **Low practicality.** Iris is a stable and accurate biometric for VR authentication but iris scanners are expensive to deploy. Currently, only high-end products like Apple Vision Pro [1] are equipped with such hardware. Several other works use uncommonly used biometrics, such as brain electrical signal [48] and electrooculography [50], as features for authentication. However, these methods require professional measurement tools that are not readily available in commercial off-the-shelf VR devices.

The above problems motivate us to ask:"*Is it possible to design a user-friendly and secure biometric-based authentication system for VR devices using the hardware available in commercial devices?*" We notice that the eye tracker has become an indispensable component of newly released VR headsets such as Varjo [6], HTC VIVE [2], Pico Neo series [4, 5], and Meta Quest Pro [3]. Compared to other biometrics, using eye movements to authenticate users is more natural and can be seamlessly integrated with the head-mounted display ecosystem. Furthermore, since an increasing number of commercial VR support eye trackers, eye movement-based authentication does not incur high deployment costs.

In terms of eye movement-based authentication, many studies have been conducted but they mainly consider the scenarios in which a user uses a flat screen, such as laptop or desktop computer [13, 34]. It is nontrivial to apply these methods directly to VR scenarios because the eye tracker of VR devices has a significantly lower sampling rate. For example, the sampling rate of EyeLink 1000 Plus [32], which is widely used in monitor-based scenarios, can achieve a sampling rate of 2,000 Hz; however, the sampling rate of eye trackers integrated in popular VR devices are usually in the range of 100 Hz–200 Hz. Recently, several eye response-based authentication systems specially designed for VR devices have been proposed. For example, Zhu *et al.* proposed SoundLock [89] which uses the auditory-pupillary response as biometric. However, the auditory stimuli used to elicit the response are limited, and ambient noise greatly affects its authentication performance.

In this paper, we propose Medusa3D, a practical, secure, and user-friendly user authentication system for VR devices via reflexive eye responses. Specifically, Medusa3D presents 3D visual stimuli to users to elicit reflexive eye responses, including reflexive saccades and pupil diameter changes, which will be recorded by the integrated eye tracker. We design a novel algorithm to extract reflexive parts from the raw data and use a Graph Neural Network (GNN) feature extractor to obtain user-specific features. Finally, a K-Nearest Neighbors (KNN) classifier is utilized to authenticate the current user. Compared with existing methods, Medusa3D offers several advantages. First, it is effortless for VR users because the reflexive eye responses are the user's inherent biometric credentials. Second, the biometric credential is anti-spoofing. Even if the credential is leaked, the user can regenerate a new one easily and the attackers cannot mimic it. Third, Medusa3D can be deployed easily in VR devices equipped with an eye tracker and it does not need additional hardware that is not supported by commercial VR devices.

To evaluate Medusa3D, we implement Medusa3D on an HTC VIVE Pro Eye VR. We carried out an extensive evaluation by recruiting 25 participants with diverse backgrounds. Evaluation results show that Medusa3D achieves 0.21% False Accept Rate (FAR) and 0.13% False Reject Rate (FRR) with authentication time in around 5 s. The evaluation demonstrates high robustness of Medusa3D for different people with diverse ocular conditions and resilience against attacks including zero-effort attack, replay attack, and mimicry attack. Additionally, we conducted a user study to show that Medusa3D is user-friendly and well-adopted by participants.

The contributions of this paper can be summarized as follows:

- We propose Medusa3D, which provides practical, secure, and user-friendly user authentication for VR devices based on reflexive eye responses. In Medusa3D, we design novel 3D visual stimuli in VR that can dynamically interact with the user's gaze. The stimuli effectively elicit reflexive saccades and pupil diameter changes by changing positions unpredictably in both directions and distances.
- To extract the reflexive saccades in VR scenario, we propose a novel algorithm to detect and extract reflexive saccades from the raw gaze data. Our algorithm involves pinpointing the time domain reflexive saccades may occur, adaptively extracting saccades from fixations and verifying the reflexivity.
- To address the information loss due to low sampling rate, we propose a GNN-based feature extractor to extract user-specific features from the reflexive eye responses. Our feature extractor embeds the spatial information of reflexive saccades as a graph, then extracting the deep features through a graph-oriented neural network.
- We conduct a comprehensive evaluation with 25 participants to demonstrate that Medusa3D can achieve high authentication accuracy for individuals with a variety of ocular and other conditions. Moreover, a user study shows that Medusa3D is user-friendly and well-adopted by the participants.

## 2 RELATED WORK

### 2.1 VR Authentication

There are many research works looking for safer mechanisms in VR authentication. The existing methods can be categorized into five classes [77]: (1) knowledge-based methods [7, 33, 52, 59]; (2) physical biometrics [11, 23, 48, 68]; (3) behavioural biometrics [50, 56, 61, 72, 87]; (4) token-based methods [20] and (5) multi-factor methods [51, 88]. Based on the discussion in [77], current methods have some limitations to being widely deployed in reality: they either require additional cognitive loads to remember the passwords, or require extra hardware like EEG and EoG instruments that are not part of the commercial VR devices. Additionally, some methods like iris scanning [84] may not be used in general VR devices but only on high-end products due to the expensive instrument needed [89]. With the development of VR, it is a trend to integrate an eye tracker within VR to improve the user experience and provide better interaction [25]. The eye tracker has become increasingly common in newly released VR headsets [2–6]. Therefore, Medusa3D utilizes the user's eye responses within VR to provide a novel authentication method, positioning it as a promising method with wide applicability.

### 2.2 Gaze-based Authentication

Since the non-intrusive nature of gaze, many studies [31, 37] have explored the feasibility of serving gaze as a biometric for authentication. Previous studies can be categorized as two classes: explicit and implicit gaze-based authentication [41].

Explicit gaze-based authentication is a category of methods that require users to register a system by performing specific eye movements consciously and remembering them. During the inference phase, only legitimate users can get access to the system by recalling their predefined passwords (*i.e.*, eye movements). These methods may use fixations [70], gaze gestures [42], and smooth pursuit eye movement [9, 27, 66]. These methods are easy to be combined with existed hardware for widespread deployment, but their authentication process is slower than the traditional PIN method [41].

Implicit gaze-based authentication, which does not require users to remember any secrets, leverages inherent gaze behaviors and can occur throughout an activity. Research in the field mainly focused on assessing unique eye movements when performing activities with varying visual stimuli [41]. Previous works have used text-based tasks [13, 34], image-based tasks [17, 24, 55, 75], dynamic stimuli like moving targets [8, 12, 87] or video [19, 71]. These methods are effortless compared with the explicit one. However, they rely on desktop eye trackers with high sampling rates or active eye movements for task completion, which attackers may mimic. Different from these systems, Medusa3D leverages reflexive eye movements for authentication in VR.

### 2.3 Challenge-Response Authentication

According to the typology of authentication methods stated in previous wrok [60], current authentication methods can be categorized as users possession, users knowledge and users characteristic. Challenge-response authentication is a method that belongs to users characteristic. However, different from static biometrics like fingerprints [38] and iris [84], challenge-response authentication relies on the active biometrics of a user's physiological response to a given stimulus (referred as a "challenge"). The underlying assumption is that the response of each user to a specific challenge is distinctive, making each challenge-response pair akin to a biometric password. Previous research on challenge-response authentication has utilized various types of active biometrics, such as finger responses to vibrations [47, 49], brain wave signals generated by brain responses to visual stimuli [11, 48], and acoustic features from ultrasound in response to facial articulatory gestures [28]. In

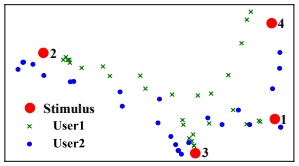



Fig. 2. Saccade patterns.

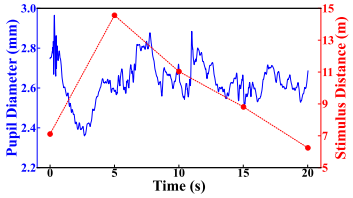



Fig. 3. Impact of changes in depth.

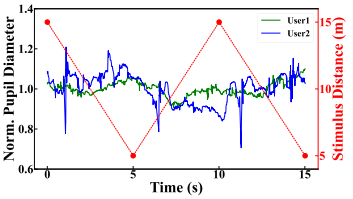



Fig. 4. Pupil change patterns.

VR scenarios, Chen *et al.* utilized muscle response to electrical stimulation [23]; Luo *et al.* proposed OcuLock [50] using electrooculography signals response to visual stimuli; Wang *et al.* and Li *et al.* used head vibrations or reverberations in VR headset [46, 82].

Compared with the methods using static biometrics [8, 36, 38], challenge-response authentication systems are more robust against data theft and replay attack, which are common problems for passive biometrics [30, 43, 85]. This is because the credentials created under challenge-response protocols are revocable. However, the challenge-response methods listed above either need additional hardware or are not suitable for usage in VR devices. Recently, SoundLock [89] utilized the auditory-pupillary response as an authentication method in VR. They used specific types of audio to elicit the change of pupil size. However, this method can be easily disturbed by the ambient noise. Instead, Medusa3D uses 3D visual stimuli in VR to trigger reflexive saccades and pupil diameter changes for robust authentication.

## 3 PRELIMINARY STUDY

### 3.1 Primer on Reflexive Eye Responses

Reflexive eye movement is an activity that is driven by visual stimulation but does not require volitional control [45, 69]. Previous research demonstrates that when a noticeable change occurs within human view, the user's eyes naturally initiate reflexive saccades to adapt to the change [44, 74], which serves as the cornerstone of the extraction of high-level cognitive information.

Previous studies [74] have investigated using reflexive saccades as biometric to authenticate users on flat panel display (*i.e.*, screen), since these saccades are produced reflexively and are effortless neuronal responses. Recent studies have proposed that pupil diameter also plays an important role in the unique gaze pattern [31], however, the aforementioned methods are screen-based, and thus cannot simulate changes in depth. Furthermore, the user's open-field view is not conducive to the elicitation of stimulus-induced reflexive saccades. Given these shortages, using VR as the implementation platform is superior because it covers the whole eyes and avoids the effect of external light on pupil diameter.

To verify the feasibility of using reflexive eye movement for user authentication, we conduct a pilot study. We implement an interactive ball as a stimulus in VR and record the user's gaze trajectory to analyze the feasibility of using reflexive saccades to authenticate users. The ball will flash to arbitrary positions once it is gazed at by the user. As Fig. 2 shows, although the same series of visual stimuli is given (fixed random seed), there are different patterns among the saccades of different users. To investigate the influence of changes in depth, we fix the direction in which the stimulus appears and only change the depth of the stimulus in the virtual space. As Fig. 3 shows, with the mutations of stimulus distances, the user's pupil diameter has corresponding dilation or constriction to adjust the changes. Furthermore, given the same series of distance mutations, changes in pupil diameter also exhibit significant differences among different individuals but remain similar for the same individual on different trials (Fig. 4). Therefore, we can conclude that the unique intrinsic patterns of reflexive eye movement can be used as biometric for user authentication.

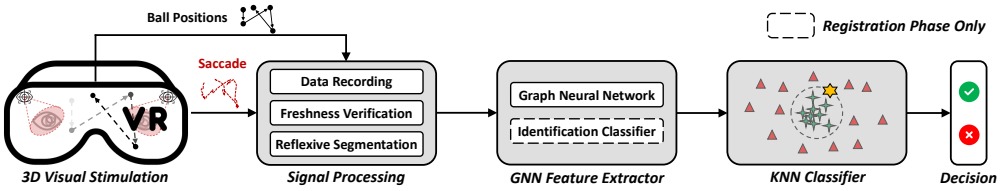



Fig. 5. System overview.

### 3.2 Design Rationale

To utilize the reflexive responses, we adopt the commonly used challenge-response mechanism. We need to carefully design the challenge to elicit the responses we need. The pilot study indicates that unexpected salient changes of the stimulus position can elicit reflexive saccades of people, while distance mutations of the stimulus within a distance of 5 m–15 m will trigger changes in pupil diameter to focus on it. Therefore, based on the intrinsic patterns of different people discovered in pilot study, we exploit the salient position changes of a ball, varying in both directions and distances within VR environment as visual stimuli. Additionally, it is crucial to design the visual stimuli to be unpredictable to prevent users from anticipating the next position, which would inhibit their reflexive responses. The detailed design of the visual stimulation will be discussed in Section 4.2.

After presenting the visual stimulation, we need to extract the reflexive responses, *i.e.*, reflexive saccades, from the raw eye movement data. Previous research works [74, 81] have shown that after an unexpected challenge, the corresponding reflexive saccade is usually initiated within a latency of 250 ms and is physically impossible to initiate less than 80 ms. The saccade usually lasts for 20 ms–100 ms [74]. Based on these facts, the time domain that a reflexive saccade may occur is the time interval of 80 ms–350 ms after the stimulus presents. During visual tasks like search or scene perception, our eyes switch between fixations and saccades [67]. Thus, within the time domain, we can extract the saccadic part from fixations based on their differences in angular velocities of eyes. Finally, we need to verify the reflexivity of the extracted saccades based on the initiation time to make sure it is the corresponding responses we need. The detailed design for reflexive saccades extraction will be elaborated in Section 4.3.

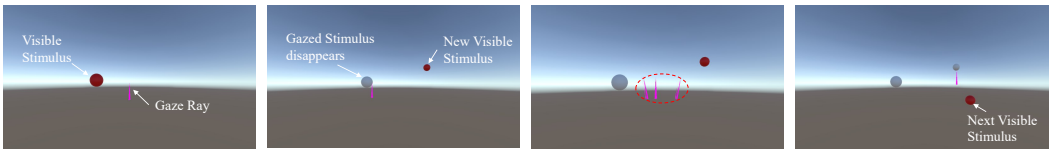
## 4 SYSTEM DESIGN

### 4.1 Overview

Fig. 5 shows the overview of Medusa3D, which consists of two phases: registration and authentication.

**Registration phase.** In this phase, the legitimate user first needs to complete a five-point calibration after putting on the VR headset to ensure that the eye tracker can precisely capture the user's eye movements. Then, a random stimulation that contains 900 stimuli will be presented in the user's view. During this process, the eye tracker continuously records the user's reflexive eye responses elicited by the stimuli. After obtaining the raw data, Medusa3D segments reflexive saccades (elaborated in Section 4.3) and trains a universal GNN feature extractor and a user-specific KNN classifier (Section 4.4 and 4.5) for each legitimate user.

**Authentication phase.** When a user tries to unlock the VR device secured by Medusa3D, a short stimulation (consists of 10 random stimuli, takes 5 s on average) and the corresponding eye movements will be coupled for signal processing. Medusa3D first verifies the freshness of the collected data to ensure its validity. If valid, the system will segment the user's reflexive saccades and further feed them into the well-trained GNN feature extractor to extract user-specific




(a) Stim. appears in FOV. (b) Gaze intersects stim. (c) Saccades catch stim. (d) New stim. is gazed.

Fig. 6. Design of visual stimulation.

features. By feeding the features into the pre-trained KNN classifier, Medusa3D obtains the final decision—whether it is a legitimate user or not.

### 4.2 3D Visual Stimulation Design

Recall the discussion in Section 3.2, to elicit reflexive eye responses, the visual stimuli used in Medusa3D should satisfy the following requirements: **(1) salient change:** the stimuli should cause salient change in the field of view to elicit reflexive saccades to search for the target; **(2) variable depth:** the visual stimuli should change in depth to elicit the variations of the user's pupil diameters as 3D features; **(3) unpredictability:** the visual stimuli should always be unseen and unpredictable for users such can exclude the interference from memory and subjective forecasting.

Based on these requirements and analysis in Section 3.2, we design the "stimuli" used in Medusa3D as an interactive red ball, which unpredictably shifts in random directions within a range of 5 m to 15 m from the user. The diameter of the ball is designed as 0.7 m in virtual 3D space, as to make the gaze interaction effect similar to that in previous work [74]. As Fig. 6b shows, once the user's gaze ray intersects with the red ball, it will flash to a new random position. The process will continue until $N$ stimuli have been presented. Here, $N$ means the number of stimuli used during a single authentication. Furthermore, considering that the user might not have focused on the red ball, the ball will change its position after 1,000 ms. Note that if the proportion of valid points (*i.e.*, gazed by user) does not reach a certain threshold ($\tau$), it will be eliminated in the freshness verification (elaborated in Section 4.3). In practice, for each 10-stimulus challenge, the whole process takes about 5 s on average.

### 4.3 Signal Processing

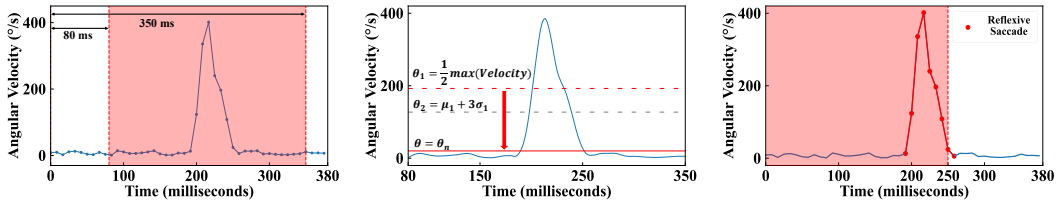
**Freshness verification.** Upon completion of the interactive data recording, we combine the stimuli positions with the user's saccades for further analysis. Given that each series of visual stimuli is fresh for the user, it is necessary to verify the freshness of the recorded eye movements. It means that Medusa3D only segments reflexive saccades from valid samples where the user's gaze intersects with a sufficient proportion ($> \tau$) of the visual stimuli, discarding any invalid samples. The threshold $\tau$ represents the ratio of successfully gazed stimuli to the total number of stimuli presented. In the authentication phase, invalid samples will result in an invalid attempt, necessitating a re-authentication.

Specifically, when a visual stimulus appears within the field of view, it is successfully gazed at only if the user's gaze ray intersects with the ball within a time not exceeding $t_{\max}$. Otherwise, the system will automatically transit the stimulus to a new position after $t_{\max}$, and the point of stimulus will be marked as an anomaly (*i.e.*, bad data point). Through practice, we set the values of $t_{\max}$ and $\tau$ to 1,000 ms and 0.5, respectively.

**Reflexive saccade segmentation.** We will now proceed to the segmentation of reflexive saccades from the raw gaze data. Previous methods [10, 76] cannot be applied to Medusa3D because they are not suitable for the extremely low sampling rate and reflexivity attribution we need. Therefore,

**Algorithm 1** Reflexive Saccade Extraction

**Require:** $\boldsymbol{\omega}$: angular velocity time series, $s$: the occurrence time of each stimulus
**Ensure:** Extracted reflexive saccades
  **for** each stimulus $s_i$ **do**
    $P \leftarrow \textsc{Slice}(\boldsymbol{\omega}, s_i + 80\text{ ms}, s_i + 350\text{ ms})$ ▷ Extract data between 80-350 ms post-stimulus
    $P_{\text{interp}} \leftarrow \textsc{Upsample}(P, 1200\text{ Hz})$ ▷ Interpolate data for high-frequency simulation
    $P_{\text{smooth}} \leftarrow \textsc{SavitzkyGolay}(P_{\text{interp}})$ ▷ Process interpolated data with Savitzky-Golay filter
    $\theta \leftarrow \frac{1}{2}\max(P_{\text{smooth}})$ ▷ Initiate the threshold value
    $\mu, \sigma \leftarrow \textsc{Statistics}(P_{\text{smooth}} < \theta)$ ▷ Compute AVG, SD of the data lower than threshold
    **while** $|(\mu + 3\sigma) - \theta| > 1$ **do**
      $\theta \leftarrow \min(\mu + 3\sigma, \theta - 0.1)$ ▷ Iterate threshold value until convergence
      $\mu, \sigma \leftarrow \textsc{Statistics}(P_{\text{smooth}} < \theta)$ ▷ Update $\mu, \sigma$ for convergence check and next loop
    **end while**
    $I \leftarrow \textsc{FindIndices}(P_{\text{smooth}} \geq \theta)$ ▷ Obtain the indices of saccade part
    $I_{\text{orig}} \leftarrow \textsc{MapToOriginal}(I, \text{step of } P_{\text{interp}})$ ▷ Map the indices to original data before upsampling
    **if** $\textsc{Latency}(I_{\text{orig}}, s_i) < 250$ ms AND $\textsc{IsPeak}(I_{\text{orig}})$ **then**
      **return** $I_{\text{orig}}$ ▷ Check if saccade is reflexive or not
    **end if**
  **end for**




(a) Reflexive saccade timing. (b) Adaptive saccade segmentation. (c) Reflexivity verification.

Fig. 7. Process to extract reflexive saccade.

we propose a novel algorithm to extract reflexive saccades in Medusa3D. A pseudo code of the algorithm is presented in Algorithm 1.

We first interpolate the data gaps caused by the eye closure, then differentiate the time series of gaze angle to obtain the angular velocity time series. Then, for each visual stimulus, we focus only on the data from the period beginning 80 ms after the stimulus appears to 350 ms when it ends (Fig. 7a), as this is the duration of time in which reflexive saccades are likely to occur as discussed in Section 3.2. To address the low sampling rate issue, we upsample the extracted data to a frequency of 1,200 Hz. Then, we employ a Savitzky-Golay filter [64] to smooth the signal.

Subsequently, the proposed algorithm can adaptively search the threshold $\theta$ to distinguish saccade and fixation. As Fig. 7b shows, $\theta$ is initialized as half of the highest peak value. During each iteration, the value of $\theta$ will be updated to $\mu + 3\sigma$, where $\mu$ represents the mean and $\sigma$ represents the standard deviation of all data points below the current value of $\theta$. This iteration process continues until the value difference of two adjacent $\theta$ is less than 1 °/s, at which the process is considered to have converged. However, for some special situations, the angular velocity data of the user's fixation fluctuates drastically. The updated $\mu + 3\sigma$ may exceed the existing $\theta$, thus the threshold will move towards an undesirable direction. To this end, we optimize the segmentation algorithm [58]

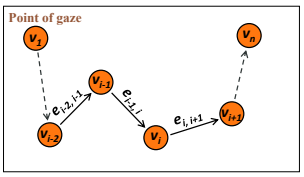



Fig. 8. The structure of gaze graph.

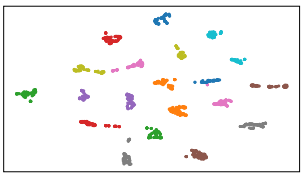


Fig. 9. t-SNE of 20 participants.

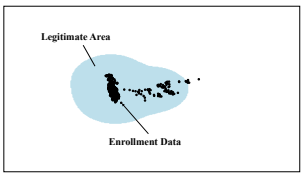



Fig. 10. KNN decision boundary.

proposed by Nyström *et al.* by adding a "one-way valve" mechanism. By selecting the minimal of $\theta - 0.1$ and $\mu + 3\sigma$, we can locate the optimization direction of $\theta$ and hence address this issue.

Finally, once $\theta$ is determined, we extract the component beyond $\theta$ as a saccade and judge whether the segmented saccade is reflexive. As Fig. 7c shows, only saccades whose starting points fall within the red area are defined as reflexive saccades. If the starting point exceeds the delay upper limit (250 ms), it will be considered as volitional [81] and will be discarded in the subsequent pipeline.

### 4.4 GNN Feature Extractor

Since GNNs excel in capturing complex relationships in non-Euclidean data structures, they are particularly adept at modeling the intricate spatial relationships inherent in gaze analysis [16]. Furthermore, their ability to handle dynamic data makes them suitable for real-time or sequential saccade tracking, a critical aspect of gaze-related tasks. Moreover, GNNs offer enhanced context understanding by considering the nodes and edges representing the surrounding environment, providing richer insights into the user's focus and interactions. Therefore, we design a GNN model to extract discriminative features from the saccade trajectory for different individuals.

Fig. 8 shows the basic idea of how to model the user's gaze as a graph, which reflects the geometric and pairwise relations among nodes (*i.e.*, gaze points). We define the segmented reflexive gaze points as a set of nodes $V = \{v_1, v_2, \ldots, v_k\}$. Then, to embed spatial information into the graph, we build an edge $E = \{e_{i,i+1}\}_{i=1}^{k-1}$ between each pair of adjacent gaze points. Considering the temporal information, we set the weight scalar $s_{i,j} > 0$ only if $j > i$ while we do not link the nodes (*i.e.*, $s_{i,j} = 0$) if $j < i$. Therefore, the graph used in Medusa3D is a directed graph $\mathcal{G} = (V, E)$. Finally, we formulate the development of an effective GNN feature extractor as a user identification problem (*i.e.*, $M$-way classification), where $M$ is the number of users involved in model development. We design a graph-oriented network that consists of a feature extractor and a classifier to achieve this goal. Three graph convolutional layers (GCNConv) are included in the feature extractor with a ReLU activation function and dropout layer ($p = 0.5$) followed by each GCNConv. Here, the hidden channel number is 64. The features extracted after the third GCNConv are utilized for intermediate representation and classification, with the final output derived by global mean pooling followed by a softmax function. This structure efficiently captures and utilizes graph-based spatial and temporal relationships for distinguishable feature learning.

Once the network is trained, the classifier will be discarded. Then, we can get a GNN feature extractor, which can obtain user-specific features from a directed graph. Fig. 9 shows the t-distributed Stochastic Neighbor Embedding (t-SNE) result of deep features in latent space. It is easy to find that each individual's cluster is separated from others, which provides the feasibility to delimit the boundary of the legitimate user using a KNN model.

### 4.5 KNN Classifier

To develop an authentication system, it is crucial to find a classifier capable of distinguishing the legitimate user from others. To this end, we implement a KNN model to delimit the boundary of the legitimate user in latent space. In practice, the efficacy of the boundary delimited by the KNN

Table 1. Participants demography.

| Gender | # | % | Age | # | % | Myopia diopters | # | % | Eyewear type | # | % | VR Ex. | # | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Female | 14 | 56 | 18–21 | 2 | 8 | [0.00 D, -2.00 D) | 5 | 20 | None | 4 | 16 | Yes | 13 | 52 |
| Male | 11 | 44 | 22–24 | 9 | 36 | [-2.00 D, -4.00 D) | 7 | 28 | Contact lenses | 3 | 12 | No | 12 | 48 |
| | | | 25–27 | 10 | 40 | [-4.00 D, -6.00 D) | 10 | 40 | Eyeglasses | 18 | 72 | | | |
| | | | 28–30 | 4 | 16 | [-6.00 D, -8.00 D] | 3 | 12 | | | | | | |

Fig. 11. A participant during the study.

model depends on two critical hyperparameters: $k$ and $\alpha$. Generally, the value of $k$ determines the number of nearest neighbors to consider, affecting the model's sensitivity to noise and its ability to generalize, while the value of $\alpha$ can influence the weighting of these neighbors, impacting the balance between bias and variance in the model's predictions. The essence of developing a KNN model is to find an effective combination of $(k, \alpha)$ that can package legitimate user samples in the latent space as much as possible while excluding samples from others.

We extract the deep features of the legitimate user using the GNN feature extractor, then project them into a 2D latent space so that we can visualize the boundary as shown in Fig. 10. The boundary defined by the KNN model tightly wraps around the cluster of legitimate user samples, with only a very few samples falling outside the boundary. In such situations, when a new attempt is made to unlock the device, it will be granted if the projected position of the new sample falls in the boundary. Otherwise, Medusa3D will regard the user as a spoofer and deny the access request.

## 5 EVALUATION

### 5.1 Experimental Setup

*5.1.1 Implementation.* We implement Medusa3D on HTC VIVE Pro Eye, as this model with its integrated eye tracker has been proved can be used as an assessment tool for saccadic eye movement [39]. The model has dual OLED 3.5-inch diagonal screen with 1440 × 1600 pixels per eye resolution. The eye tracker integrated within the VR has a sampling frequency of 120 Hz with a trackable field of view of 110 degrees. The VR is powered by a desktop PC with an Intel Core i7-8700K CPU and an Nvidia Geforce RTX 2080Ti GPU. The prototype is implemented through the Unity platform with version 2021.3.30f1. The eye movement data is output through the SRanipal API [80]. The GNN model and user-specific KNN classifier are trained on the same PC using the RTX 2080Ti GPU and the PyTorch framework with version 1.13. All experiments adopt a 7:2:1 ratio to split training, validation, and testing data in the time order.

*5.1.2 Data Collection.* To evaluate the performance of Medusa3D, we invited 25 participants (14 females, 11 males), with ages ranging from 18 to 30 years old (AVG=24.42, SD=2.87)[1]. We considered

[1]Ethical approval has been obtained to conduct this study (No. H002554)

the diversity of the users from different perspectives, such as the eyewear type, the severity of myopia, and the VR experience. The details of participants' demography are shown in Tab. 1.

The experiment setup for participants is shown in Fig. 11. Each participant was informed of the experiment's purpose and the data that would be collected. An experimenter assisted the participants in wearing and adjusting the VR devices. Then, the participants calibrate wearing the VR with a 5-point calibration [79] provided by HTC. It is worth noting that calibration was necessary after each re-wearing to ensure that the hardware can track the user's gaze precisely. During data collection, the participants were challenged by 1,000 visual stimuli (*e.g.*, the number of balls presented to the participant). The total duration of data collection for each participant was about 10 minutes. Out of all the 25 participants, we randomly chose five participants as unseen attackers, *i.e.*, the five users' data will not be used in any experiments except security analysis against unseen attackers.

*5.1.3 Threat Model.* We assume that the communication between the eye tracker and the CPU cannot be surveilled. Similar to previous studies [22, 89], we consider the following attacks:

- **Zero-effort attack:** an attacker attempts to unlock the VR headset using their biometric responses to visual stimuli, impersonating the legitimate user.
- **Replay attack:** an attacker tries to replay or inject the previously recorded eye movement samples to Medusa3D.
- **Mimicry attack:** an attacker observes and mimics the legitimate user's eye movement from previously recorded samples and try to spoof Medusa3D by imitating the legitimate user.

*5.1.4 Metrics.* We use the following metrics, which are widely used in previous works [22, 23]:

- **Identification Error Rate:** the probability that the identity of a user is incorrectly classified.
- **False Accept Rate (FAR):** the probability that an attacker is authenticated as legitimate.
- **False Reject Rate (FRR):** the probability that a legitimate user is authenticated as illegitimate.
- **Equal Error Rate (EER):** the point at which the FAR and FRR are equal.

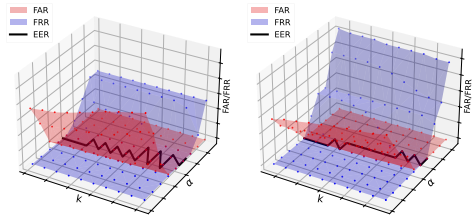
### 5.2 Overall Performance

In this experiment, we aim to evaluate the overall authentication performance of Medusa3D. We first use all participants' training data and validation data to train the GNN feature extractor. The trained model has achieved an overall identification accuracy rate of 99.7% on test data. This result indicates that the trained model can effectively extract the features of different people. We then use the feature values extracted by the feature extractor to develop the user-specific KNN classifier. For each participant, we train the classifier using the training dataset's feature values and determine the optimal hyperparameters $k$ and $\alpha$ through the performance in validation data. As shown in Fig. 12, the error rate on the validation set of different participants may show different patterns for $k$ and $\alpha$. There are many EER points as illustrated. For each participant, we find the lowest EER point and use its corresponding $k$, $\alpha$ as the optimal hyperparameters to train a user-specific classifier for this participant. Finally, for every participant, we use the test data as legitimate attempts and all other participants' test data as illegal attempts to calculate the FAR and FRR, respectively, to evaluate the user-specific classifier. We then calculate the FAR and FRR for all the participants.

Fig. 13 plots the FAR and FRR of all the 20 participants. Most participants can achieve 0% FAR and FRR, while others have low FAR or FRR, indicating effective authentication with user-specific classifiers. The highest values of FAR and FRR are 3.54% (*i.e.*, User 7) and 1.52% (*i.e.*, User 20). Overall, Medusa3D achieves 0.21% FAR and 0.13% FRR, demonstrating its high reliability and effectiveness across a diverse demographic. Compared with the state-of-the-art authentication methods for VR

Table 2. Comparison with state-of-the-arts.

| Methods | FAR (%) | FRR (%) | Authentication time (s) |
|---|---|---|---|
| OcuLock [50] | 3.55 | 3.55 | ≤10 |
| SkullConduct [68] | 6.90 | 6.90 | ≤23 |
| Brain Password [48] | 2.50 | 2.50 | ≈4.80 |
| ElectricAuth [23] | 0.83 | 2.00 | ≈1.30 |
| SoundLock [89] | 0.76 | 0.91 | ≤7 |
| VibHead [46] | ≈5 | ≈5 | ≤1 |
| **Medusa3D** | **0.21** | **0.13** | **≈5** |




(a) (b)

Fig. 12. FAR, FRR and EER for specific user with respect to $k$ and $\alpha$ under KNN classifier.

devices, Medusa3D almost achieves the best performance among all biometric-based methods in both FAR and FRR while maintaining a reasonable authentication time as shown in Tab. 2.

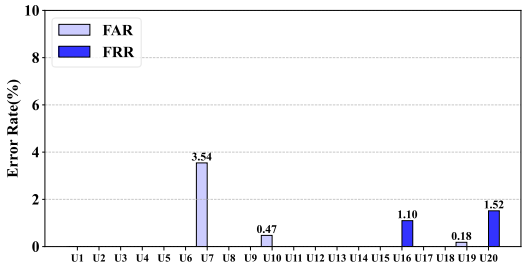



Fig. 13. Authentication error rates for 20 participants.

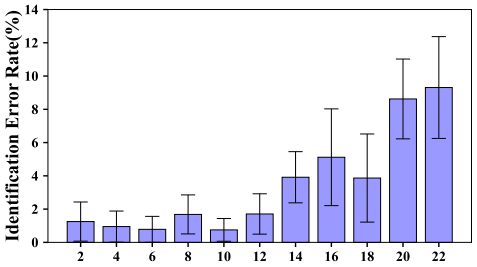



Fig. 14. The impact of visual stimuli amount ($N$).

## 5.3 Micro-benchmarks

In this subsection, we will evaluate the impact of various factors that may affect the performance of Medusa3D.

*5.3.1 Impact of Visual Stimuli Amount.* A crucial factor affecting the performance of Medusa3D is the amount of visual stimuli presented to users because it balances the security and user experience. Intuitively, the more visual stimuli we present to the user, the more information we can obtain and the lower error rate Medusa3D will achieve. Therefore, this experiment aims to evaluate the impact of visual stimuli amount on authentication performance. To this end, we adjust the number of visual stimuli and calculate the identification error rate. Fig. 14 shows the results of using different numbers of stimuli ($N$). To our surprise, the identification error rate does not decrease gradually

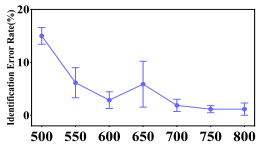



Fig. 15. Impact of train.

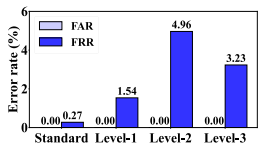



Fig. 16. Impact of fatigue.

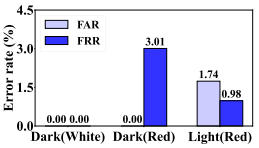



Fig. 17. Impact of colors.

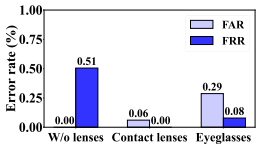



Fig. 18. Impact of lenses.

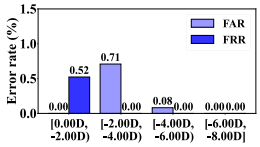



Fig. 19. Impact of myopia.

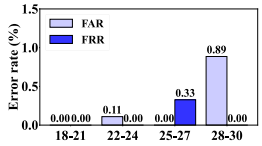



Fig. 20. Age diversity.

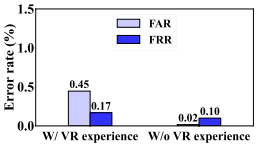



Fig. 21. VR Ex. diversity.

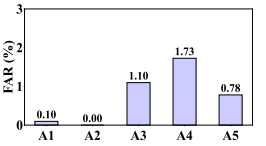



Fig. 22. Zero-effort attack.

when more visual stimuli are used but instead increases. After analysis, we find that larger $N$ leads to a more complicated graph which integrates more information but fewer training samples since the total amount of training data is fixed. Meanwhile, we also need to note that we should not directly choose the lowest $N$ since lower $N$ results in less information (*e.g.*, simple gaze graph). For example, the error rates of $N = 4$ and $N = 6$ are lower than that of $N = 2$. Thus we need to find the balance between these factors to find the best value of $N$ for Medusa3D. Finally, we choose $N = 10$ as the best visual stimuli amount setting for Medusa3D because it has shown the best mean error rate as well as the lowest standard deviation, which means a stable performance.

*5.3.2 Impact of Training Dataset Size.* Training dataset size is an important factor when building up Medusa3D. It cannot be as large as possible as it is correlated with the registration time when the user first registers. To evaluate the impact of training dataset size on Medusa3D, we adjust the training size from 500 to 800 visual stimuli and calculate the identification error rate. As shown in Fig. 15, with the training data increasing, the error rate falls to around 2% and comes to a plateau phase. Although increasing the dataset size can achieve better authentication performance, it also leads to longer registration time. As the identification error rate comes to a plateau phase, finally we choose 700 visual stimuli as a trade-off. This setting means that users need to be challenged by 700 training stimuli and additional validation part for registration. Thus the total registration time of Medusa3D will need around 7 minutes.

*5.3.3 Impact of Fatigue.* Prolonged and intense eye movements may induce fatigue in users, potentially affecting the performance of Medusa3D. To evaluate the influence of fatigue on authentication accuracy, we invited five participants to participate in an extended half-hour data collection following the standard data collection. We equally divide the data into three parts based on their timestamps so that different parts simulate different levels of fatigue after using a VR device for a certain period. We use these data as test data to calculate the authentication performance of the model trained on the standard dataset. The error rates corresponding to varying levels of fatigue are depicted in Fig. 16. The data shown in "Standard" is calculated with the test data in standard dataset. We can observe that as fatigue level increases, FAR remains consistent with that of standard level, while FRR increases to approximately 5% at Level-2 and falls back to about 3% at Level-3. These results indicate that fatigue has an impact on user-specific feature and affects the authentication performance of Medusa3D.

*5.3.4 Impact of Stimuli Colors.* To evaluate the impact of colors on the visual stimuli, we invite five participants to conduct an additional experiment. We set three different color settings for visual stimulation: (1) light background with a red ball; (2) dark background with a red ball and (3) dark background with a white ball. For each setting, we present 500 visual stimuli for each participant, *i.e.*, around 5 minutes of data per setting per participant. The result is depicted in Fig. 17. We observe that the setting of a dark background with a white ball has shown the best performance while the performance on a light background with a red ball is acceptable. Considering the interaction need in the VR scenario, the light background will provide better potential to integrate Medusa3D as a pervasive authentication tool. Thus we finally choose the light background with a red ball as our visual stimuli setting for Medusa3D.

*5.3.5 Impact of Contact Lenses and Glasses.* We now investigate whether the user wears glasses or not has an impact on the authentication performance. We divide the participants into three groups: (1) those wearing conventional eyeglasses (14 people); (2) those wearing contact lenses (3 people); (3) those without any corrective lenses (3 people). Fig. 18 plots the FAR and FRR of different groups. The highest error rate for all three groups is around 0.5%. We find that the participants with contact lenses and glasses obtain similar performance while they achieve better performance compared

with the people without any lenses. The reason is that the features of different eyewears may be included during the eye tracker calibration procedure and the unique eyewear features are helpful to achieve better authentication performance. These findings suggest that Medusa3D achieves high authentication rates regardless of whether the user is wearing glasses or not, demonstrating the wide applicability of Medusa3D.

*5.3.6 Impact of Myopia Diopters.* High myopia may affect eye tracking performance [79] and further affect the authentication performance. To evaluate the impact of myopia diopters on Medusa3D, we examine the influence of varying diopters of myopia on the system's error rate. Participants are divided into four groups based on their myopia severity: [0.00 D, -2.00 D) with four people; [-2.00 D, -4.00 D) with five people; [-4.00 D, -6.00 D) with nine people and [-6.00 D, -8.00 D] with two people, where D is diopter. We calculate the FAR and FRR of all the groups and plot the results in Fig. 19. First, we notice that the error rate decreases with the increment of myopia diopters. This is because people with high myopia always wear glasses or contact lenses, and the data captured by the eye tracker may include information of the unique lenses. Additionally, the error rates of all groups are below 1.0% and the highest error rate is around 0.7%, indicating the high robustness of Medusa3D for different myopia diopters.

*5.3.7 Impact of Other Factors.* In this experiment, we evaluate the influence of other factors, including age and prior experience with VR. The results of these factors are depicted in Fig. 20 and Fig. 21, respectively. In terms of age, the results in Fig. 20 indicate that the younger participants seem to achieve better performance than the older participants. The reason may be the responsiveness difference as the age increases. In terms of prior experience with VR, the results in Fig. 21 show that the people with VR experience achieve slightly higher FAR than those without experience. This may be because users without VR experience exhibit better reflexive responses compared with those with VR experience. The results also demonstrate that Medusa3D can achieve high authentication accuracy for a wide range of users.

### 5.4 Security Analysis

*5.4.1 Zero-effort Attack.* In a zero-effort attack, attackers attempt to unlock the device with Medusa3D by using their own biometrics as credentials. The assumption here is that the attacker has obtained physical access to the victim's VR headset. This could occur, for instance, if the device is misplaced or stolen, or in scenarios where it is inadvertently left in a public space, such as a library, a classroom, or a cafe. We randomly choose 5 out of 25 participants as attackers and treat the remaining participants as legitimate users. The data of the legitimate users are used to train their own user-specific classifiers after which the data of the attackers are used to attack the system. The aggregate FAR of five unseen attackers is recorded at 0.74%. A detailed breakdown of the FAR for each attacker is illustrated in Fig. 22. Notably, the highest attacker's FAR is below 2%. Such findings provide a robust evaluation of the system's defensive capabilities against zero-effort attack.

*5.4.2 Replay Attack.* In a replay attack, an adversary could replay a pre-recorded eye movement response to bypass the authentication system. However, spoofing Medusa3D via this approach is non-trivial, if not impossible, due to several factors. First, the VR headset occludes the eyes part, precluding external recording of eye movements. Second, even if an attacker successfully captures the eye movements of a legitimate user through malware, the feasibility of replaying these movements on a standard display to Medusa3D is highly improbable. This is because the eye-tracking technology relies on the reflection of near-infrared light from the eyes to build its model and calculate the associated gaze directions for interactive stimulation [18]. Such reflections cannot be emulated by a screen or other materials. Last, when an attacker tries to inject the pre-recorded

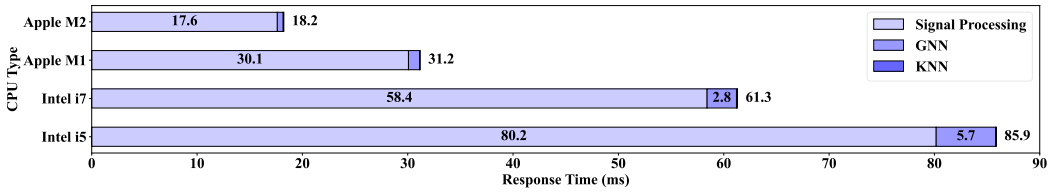



Fig. 23. Computational delay on different devices.

eye movement data via a software virus, our challenge-response mechanism will always present a fresh series of visual stimuli that is different from before. The previous recorded eye movement cannot be coupled with the new visual stimuli. Thus Medusa3D is resistant to the replay attack.

*5.4.3 Mimicry Attack.* Attackers may also attempt a mimicry attack, wherein an attacker acquires and imitates the eye movement patterns of the legitimate user's behavior to bypass Medusa3D. Nevertheless, Medusa3D is designed to thwart such attacks. First, the visual stimuli used in Medusa3D are dynamic and different in each attempt, which leads to different eye responses in different attempts. It is difficult for attackers to make a similar response when facing a new unseen challenge based on the old eye movement response that they have learned and imitated. Second, Medusa3D exclusively recognizes reflexive responses as biometrics for authentication, while disregarding voluntary eye movements, *i.e.*, the imitation eye movement is voluntary and will be excluded from the reflexive part for authentication. Therefore, the design of Medusa3D fundamentally prevent mimicry attack.

## 5.5 System Overhead

The authentication process should be efficient to ensure an optimal user experience within VR environments. To assess the time efficiency of Medusa3D in authenticating users under various settings, we conduct evaluations using four distinct types of CPUs: Intel i5-8265U, i7-8700K, Apple M1 and M2. It is important to mention that our experimental procedures are conducted entirely on the CPU, without leveraging GPU resources. We measure the response times for 100 authentication attempts and report the average time.

The whole pipeline of Medusa3D includes three main components: signal processing, GNN feature extractor, and KNN classifier. The running time of each component is depicted in Fig. 23. The overall running time varies greatly on different hardware devices, ranging from 18.2 ms–85.9 ms. On the i5 CPU, which has the lowest computing capability, the running time of Medusa3D to complete one authentication attempt is 85.9 ms, demonstrating the high efficiency of our design. We also find that the most time-consuming component is signal processing, which takes about 93%–97% of the whole processing time. This is because signal processing part includes the reflexive part extraction task which utilizes the iteration method that is time-consuming. Recall the stimulus amount determined in Section 5.3.1, *i.e.*, 10 stimuli each time. In practice, for each authentication attempt, the stimulation process takes around 5 s and our evaluation indicates the computational time does not exceed 100 ms on prevalent hardware devices. In summary, combining the stimulation process and computational time together, the required authentication time for Medusa3D is approximately 5 s.

## 5.6 User Study

To evaluate the usability of Medusa3D, we invite all of the participants to conduct a user study. These participants experience the usage of Medusa3D and they are told to evaluate the usability of Medusa3D by answering multiple questions based on their own experience. We design seven specific questions to understand the users' feedback. The questionnaire is shown in Fig. 24. The

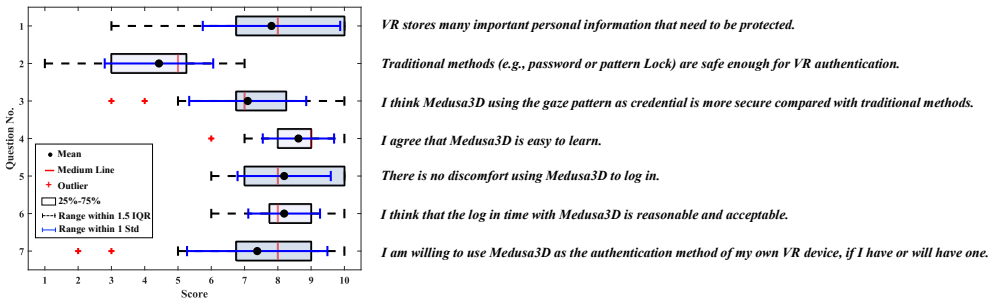



Fig. 24. Results of user study.

participants answer these questions with scores ranging from 1 to 10 which represent from strongly disagree to strongly agree.

Fig. 24 shows the results of user study. For the opinions on current VR privacy and security situations (Q1 & Q2), most participants agree that there are private information stored in VR and current authentication methods are not enough to protect it. For the comparison of new method with traditional methods (Q3), most participants agree that using gaze patterns as credential is more secure than password or pattern lock. While two participants express skepticism about the security enhancement provided by using gaze patterns as credentials. They argue that the choice of authentication method is less significant compared to the broader issue of network security, which they believe plays a more crucial role in overall protection. When coming to the user-friendliness of Medusa3D (Q4–6), all of the participants agree with these questions and most of them strongly agree (the averages and medians of all three questions are higher than 8). The results show that most of participants agree Medusa3D is easy to learn and use without discomfort while the authentication time is acceptable for them. The responses from participants indicate that Medusa3D is a user-friendly system. Finally, we inquire the participants if they are willing to use Medusa3D for their own VR devices (Q7). Participants give a wide range of responses. Overall, the majority of participants are willing to use Medusa3D. While there are still two participants who are not willing or even strongly unwilling to use Medusa3D, attributing their reluctance to habituation with passwords or skepticism towards the maturity of this novel technology.

Overall, the majority of participants recognize the potential security issue in VR and are not satisfied with traditional authentication methods. Medusa3D provides a user-friendly and secure method for VR authentication and most participants display openness and acceptance towards the new method.

## 6 LIMITATION AND FUTURE WORK

Although the previous evaluation demonstrates the superior performance of Medusa3D under different conditions, there still exist several limitations that need to be tackled.

**Performance degradation over time.** The principle of Medusa3D is that the response to the visual stimuli of every user is unique. However, previous research [63] has shown that the template acquired during the registration phase poorly represents the newly collected data samples, leading to degraded performance. This will pose a significant challenge to our system. In spite of the limitation, there are some promising solutions to address it. For example, previous work [29] achieves stable sensing performance during a long time period by combining unsupervised domain adaptation and self-supervised learning. Furthermore, the application of incremental learning on biometrics-based authentication [21, 73] indicates that it is a promising solution to solve the

long-period performance degradation problem. The investigation of integrating these techniques into our system will be included in our future work.

**Different nervous system status.** Medusa3D relies on the reflexive response to visual stimuli and the reflexive responses are determined by the nervous system. However, people's nervous system status is affected by many factors such as tiredness and drunk [14, 26]. In this case, the response time and response pattern of a user may be different from that in normal status. To address the limitation, we will integrate user state detection algorithms to detect the user's nervous status and accordingly combine adaptive algorithms to adapt to different nervous conditions.

**Different eye conditions.** Previous evaluation has shown that Medusa3D can perform well on participants with different eye conditions. However, if the user has gone through an eye surgery where scars were left on the cornea, the eye tracker may not track the eye movement correctly [78]. As a result, the authentication performance will drop significantly. Additionally, a user may wear different glasses during training data collection and authentication. To address the problem, we may apply domain adaptation techniques to make sure we can learn features that are independent of eye conditions.

## 7 CONCLUSION

In this paper, we propose Medusa3D, a challenge-response authentication system for VR based on reflexive eye responses. Medusa3D employs a GNN feature extractor to extract user-specific features from the reflexive saccades and pupil diameter changes, elicited by meticulously designed visual stimuli. Then, we use KNN to build a user-specific authentication classifier. We implemented a prototype of Medusa3D on an HTC VIVE Pro Eye VR and conducted extensive evaluations to evaluate the performance of Medusa3D under different conditions. Evaluation results show that Medusa3D can provide reliable authentication on users with diverse conditions while incurring low computational costs. Our user study also demonstrates that Medusa3D is user-friendly and well-adopted by the participants for use as an authentication method on their own VR devices.

## ACKNOWLEDGMENTS

The work described in this paper was substantially sponsored by the project 62101471 supported by NSFC and was partially supported by the Shenzhen Research Institute, City University of Hong Kong. The work described in this paper was partially supported by the Natural Science Foundation of Guangdong Province (Project No. 2024A1515010192), the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 21201420 and CityU 11201422), the Innovation and Technology Commission of Hong Kong (Project No. PRP/037/23FX and MHP/072/23), and NSF of Shandong Province (Project No. ZR2021LZH010). We extend our gratitude to Dr. Yongliang Chen for assistance with the user study and to Dr. Jingbo Hu for valuable advice on optometry. We also thank all the participants for their involvement in the study, as well as the reviewers for their constructive comments.

## REFERENCES

[1] 2023. Apple Vision Pro. https://www.apple.com/apple-vision-pro/.
[2] 2023. HTC VIVE Pro Eye. https://www.vive.com/hk/product/vive-pro-eye/overview/.
[3] 2023. Meta Quest Pro. https://www.meta.com/quest/quest-pro/.
[4] 2023. Pico Neo2 Eye. https://www.tobii.com/products/integration/xr-headsets/device-integrations/pico-neo-2-eye.
[5] 2023. Pico Neo3 Pro Eye. https://www.tobii.com/products/integration/xr-headsets/device-integrations/pico-neo-3-pro-eye.
[6] 2023. Varjo. https://varjo.com/xr-headsets/.

[7] Yomna Abdelrahman, Florian Mathis, Pascal Knierim, Axel Kettler, Florian Alt, and Mohamed Khamis. 2022. Cuevr: Studying the usability of cue-based authentication for virtual reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. 1–9.

[8] Karan Ahuja, Rahul Islam, Varun Parashar, Kuntal Dey, Chris Harrison, and Mayank Goel. 2018. Eyespyvr: Interactive eye sensing using off-the-shelf, smartphone-based vr headsets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–10.

[9] Hassoumi Almoctar, Pourang Irani, Vsevolod Peysakhovich, and Christophe Hurter. 2018. Path Word: A multimodal password entry method for ad-hoc authentication based on digits' shape and smooth pursuit eye movements. In *Proceedings of the 20th ACM international conference on multimodal interaction*. 268–277.

[10] Richard Andersson, Linnea Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. 2017. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior research methods* 49 (2017), 616–637.

[11] Patricia Arias-Cabarcos, Thilo Habrich, Karen Becker, Christian Becker, and Thorsten Strufe. 2021. Inexpensive brainwave authentication: new techniques and insights on user acceptance. In *30th USENIX Security Symposium (USENIX Security 21)*. 55–72.

[12] Gary Bargary, Jenny M Bosten, Patrick T Goodbourn, Adam J Lawrance-Owen, Ruth E Hogg, and John D Mollon. 2017. Individual differences in human eye movements: An oculomotor signature? *Vision research* 141 (2017), 157–169.

[13] Akram Bayat and Marc Pomplun. 2018. Biometric identification through eye-movement patterns. In *Advances in Human Factors in Simulation and Modeling: Proceedings of the AHFE 2017 International Conference on Human Factors in Simulation and Modeling*. Springer, 583–594.

[14] Henri Begleiter and Arthur Platz. 1972. The effects of alcohol on the central nervous system in humans. In *The Biology of Alcoholism: Volume 2: Physiology and Behavior*. Springer, 293–343.

[15] Satish Rupraoji Billewar, Karuna Jadhav, VP Sriram, Dr A Arun, Sikandar Mohd Abdul, Kamal Gulati, and Dr Narinder Kumar Kumar Bhasin. 2022. The rise of 3D E-Commerce: the online shopping gets real with virtual reality and augmented reality during COVID-19. *World Journal of Engineering* 19, 2 (2022), 244–253.

[16] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. 2018. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE transactions on knowledge and data engineering* 30, 9 (2018), 1616–1637.

[17] Virginio Cantoni, Chiara Galdi, Michele Nappi, Marco Porta, and Daniel Riccio. 2015. GANT: Gaze analysis technique for human identification. *Pattern Recognition* 48, 4 (2015), 1027–1038.

[18] Jiani Cao, Chengdong Lin, Yang Liu, and Zhenjiang Li. 2022. Gaze Tracking on Any Surface with Your Phone. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 320–333.

[19] Dario Cazzato, Marco Leo, Andrea Evangelista, and Cosimo Distante. 2015. Soft biometrics by modeling temporal series of gaze cues extracted in the wild. In *Advanced Concepts for Intelligent Vision Systems: 16th International Conference, ACIVS 2015. Proceedings 16*. Springer, 391–402.

[20] Pan Chan, Tzipora Halevi, and Nasir Memon. 2015. Glass otp: Secure and convenient user authentication on google glass. In *Financial Cryptography and Data Security: FC 2015 International Workshops, BITCOIN, WAHC, and Wearable*. Springer, 298–308.

[21] Jagmohan Chauhan, Young D Kwon, Pan Hui, and Cecilia Mascolo. 2020. Contauth: Continual learning framework for behavioral-based user authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 4 (2020), 1–23.

[22] Yongliang Chen, Tao Ni, Weitao Xu, and Tao Gu. 2022. SwipePass: Acoustic-based second-factor user authentication for smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 6, 3 (2022), 1–25.

[23] Yuxin Chen, Zhuolin Yang, Ruben Abbou, Pedro Lopes, Ben Y Zhao, and Haitao Zheng. 2021. User authentication via electrical muscle stimulation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[24] Elena N Cherepovskaya and Andrey V Lyamin. 2017. An evaluation of biometrie identification approach on low-frequency eye tracking data. In *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE, 000123–000128.

[25] Viviane Clay, Peter König, and Sabine Koenig. 2019. Eye tracking in virtual reality. *Journal of eye movement research* 12, 1 (2019).

[26] M Thyge Corfitsen. 1994. Tiredness and visual reaction time among young male nighttime drivers: a roadside survey. *Accident Analysis & Prevention* 26, 5 (1994), 617–624.

[27] Heiko Drewes, Mohamed Khamis, and Florian Alt. 2019. Dialplates: Enabling pursuits-based user interfaces with large target numbers. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. 1–10.

[28] Di Duan, Zehua Sun, Tao Ni, Shuaicheng Li, Xiaohua Jia, Weitao Xu, and Tianxing Li. 2024. F2Key: Dynamically Converting Your Face into a Private Key Based on COTS Headphones for Reliable Voice Interaction. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services (MobiSys)*. 127–140.

[29] Di Duan, Huanqi Yang, Guohao Lan, Tianxing Li, Xiaohua Jia, and Weitao Xu. 2023. EMGSense: A Low-Effort Self-Supervised Domain Adaptation Framework for EMG Sensing. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 160–170.
[30] Simon Eberz, Nicola Paoletti, Marc Roeschlin, Marta Kwiatkowska, I Martinovic, and A Patané. 2017. Broken hearted: How to attack ECG biometrics. In *Network and Distributed System Security (NDSS) Symposium 2017*. Internet Society.
[31] Simon Eberz, Kasper Rasmussen, Vincent Lenders, and Ivan Martinovic. 2015. Preventing lunchtime attacks: Fighting insider threats with eye movement biometrics. In *Network and Distributed System Security (NDSS) Symposium 2015*. Internet Society.
[32] EyeLink. 2023. EyeLink 1000 Plus. https://www.sr-research.com/eyelink-1000-plus/.
[33] Markus Funk, Karola Marky, Iori Mizutani, Mareike Kritzler, Simon Mayer, and Florian Michahelles. 2019. Lookunlock: Using spatial-targets for user-authentication on hmds. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.
[34] Anjith George and Aurobinda Routray. 2016. A score level fusion method for eye movement biometrics. *Pattern Recognition Letters* 82 (2016), 207–215.
[35] Ceenu George, Mohamed Khamis, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and secure vr: Adapting and evaluating established authentication systems for virtual reality. NDSS.
[36] Anhong Guo, Robert Xiao, and Chris Harrison. 2015. Capauth: Identifying and differentiating user handprints on commodity capacitive touchscreens. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. 59–62.
[37] Kenneth Holmqvist and R Andersson. 2017. Eye tracking: A comprehensive guide to methods. *Paradigms and measures* (2017).
[38] Christian Holz and Patrick Baudisch. 2013. Fiberio: a touchscreen that senses fingerprints. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 41–50.
[39] Yu Imaoka, Andri Flury, and Eling D De Bruin. 2020. Assessing saccadic eye movements with head-mounted display virtual reality technology. *Frontiers in Psychiatry* 11 (2020), 572938.
[40] Mohd Javaid and Abid Haleem. 2020. Virtual reality applications toward medical field. *Clinical Epidemiology and Global Health* 8, 2 (2020), 600–605.
[41] Christina Katsini, Yasmeen Abdrabou, George E Raptis, Mohamed Khamis, and Florian Alt. 2020. The role of eye gaze in security and privacy applications: Survey and future HCI research directions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–21.
[42] Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016. Gazetouchpass: Multimodal authentication using gaze and touch on mobile devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2156–2164.
[43] Tomi Kinnunen, Md Sahidullah, Héctor Delgado, Massimiliano Todisco, Nicholas Evans, Junichi Yamagishi, and Kong Aik Lee. 2017. The ASVspoof 2017 challenge: Assessing the limits of replay spoofing attack detection. (2017).
[44] Michael F. Land. 2011. Oculomotor behaviour in vertebrates and invertebrates. In *The Oxford Handbook of Eye Movements*. Oxford University Press.
[45] R John Leigh and David S Zee. 2015. *The neurology of eye movements*. Contemporary Neurology.
[46] Feng Li, Jiayi Zhao, Huan Yang, Dongxiao Yu, Yuanfeng Zhou, and Yiran Shen. 2023. Vibhead: An authentication scheme for smart headsets through vibration. *ACM Transactions on Sensor Networks* (2023).
[47] Jingjie Li, Kassem Fawaz, and Younghyun Kim. 2019. Velody: Nonlinear vibration challenge-response for resilient user authentication. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 1201–1213.
[48] Feng Lin, Kun Woo Cho, Chen Song, Wenyao Xu, and Zhanpeng Jin. 2018. 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*. 296–309.
[49] Jian Liu, Chen Wang, Yingying Chen, and Nitesh Saxena. 2017. VibWrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 73–87.
[50] Shiqing Luo, Anh Nguyen, Chen Song, Feng Lin, Wenyao Xu, and Zhisheng Yan. 2020. OcuLock: Exploring human visual system for authentication in virtual reality head-mounted display. In *2020 Network and Distributed System Security Symposium (NDSS)*.
[51] Florian Mathis, Hassan Ismail Fawaz, and Mohamed Khamis. 2020. Knowledge-driven biometric authentication in virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–10.
[52] Florian Mathis, John H Williamson, Kami Vaniea, and Mohamed Khamis. 2021. Fast and secure authentication in virtual reality using coordinated 3d manipulation and pointing. *ACM Transactions on Computer-Human Interaction*

*(ToCHI)* 28, 1 (2021), 1–44.

[53] Volker Mellert, Ingo Baumann, Nils Freese, and Reinhard Weber. 2008. Impact of sound and vibration on health, travel comfort and performance of flight attendants and pilots. *Aerospace Science and Technology* 12, 1 (2008), 18–25.

[54] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. 2020. Within-system and cross-system behavior-based biometric authentication in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 311–316.

[55] Subhadeep Mukhopadhyay and Shinjini Nandi. 2018. LPiTrack: Eye movement pattern recognition algorithm and application to biometric identification. *Machine Learning* 107 (2018), 313–331.

[56] Tahrima Mustafa, Richard Matovu, Abdul Serwadda, and Nicholas Muirhead. 2018. Unsure how to authenticate on your vr headset? come on, use your head!. In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*. 23–30.

[57] Kien Nguyen, Hugo Proença, and Fernando Alonso-Fernandez. 2022. Deep Learning for Iris Recognition: A Survey. *arXiv preprint arXiv:2210.05866* (2022).

[58] Marcus Nyström and Kenneth Holmqvist. 2010. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods* 42, 1 (2010), 188–204.

[59] Ilesanmi Olade, Hai-Ning Liang, Charles Fleming, and Christopher Champion. 2020. Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (vr). In *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations*. 45–52.

[60] Maria Papathanasaki, Leandros Maglaras, and Nick Ayres. 2022. Modern authentication methods: A comprehensive survey. *AI, Computer Science and Robotics Technology* (2022).

[61] Ken Pfeuffer, Matthias J Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. 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*. 1–12.

[62] Logan Pinter, Marcos Izquierdo, and Mohammad Faridul Haque Siddiqui. 2023. Revolutionizing Learning: An Interactive VR Application for Solids of Revolution. In *Proceedings of the 7th International Conference on Education and Multimedia Technology*. 35–40.

[63] Paulo Henrique Pisani, Abir Mhenni, Romain Giot, Estelle Cherrier, Norman Poh, André Carlos Ponce de Leon Ferreira de Carvalho, Christophe Rosenberger, and Najoua Essoukri Ben Amara. 2019. Adaptive biometric systems: Review and perspectives. *ACM Computing Surveys (CSUR)* 52, 5 (2019), 1–38.

[64] William H Press and Saul A Teukolsky. 1990. Savitzky-Golay smoothing filters. *Computers in Physics* 4, 6 (1990), 669–672.

[65] Huajian Qiu, Paul Streli, Tiffany Luong, Christoph Gebhardt, and Christian Holz. 2023. ViGather: Inclusive Virtual Conferencing with a Joint Experience Across Traditional Screen Devices and Mixed Reality Headsets. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (2023), 1–27.

[66] Vijay Rajanna, Seth Polsley, Paul Taele, and Tracy Hammond. 2017. A gaze gesture-based user authentication system to counter shoulder-surfing attacks. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 1978–1986.

[67] Keith Rayner. 2009. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology* 62, 8 (2009), 1457–1506.

[68] Stefan Schneegass, Youssef Oualil, and Andreas Bulling. 2016. SkullConduct: Biometric user identification on eyewear computers using bone conduction through the skull. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1379–1384.

[69] Caroline KL Schraa-Tam, Phillippus van Broekhoven, Josef N van der Geest, Maarten A Frens, Marion Smits, and Aad van der Lugt. 2009. Cortical and cerebellar activation induced by reflexive and voluntary saccades. *Experimental brain research* 192 (2009), 175–187.

[70] Mythreya Seetharama, Volker Paelke, and Carsten Röcker. 2015. Safetypin: Secure pin entry through eye tracking. In *Human Aspects of Information Security, Privacy, and Trust: Third International Conference, HAS 2015. Proceedings 3*. Springer, 426–435.

[71] Sherif Seha, Georgios Papangelakis, Dimitrios Hatzinakos, Ali Shahidi Zandi, and Felix JE Comeau. 2019. Improving eye movement biometrics using remote registration of eye blinking patterns. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2562–2566.

[72] Yiran Shen, Hongkai Wen, Chengwen Luo, Weitao Xu, Tao Zhang, Wen Hu, and Daniela Rus. 2018. GaitLock: Protect virtual and augmented reality headsets using gait. *IEEE Transactions on Dependable and Secure Computing* 16, 3 (2018), 484–497.

[73] Zhihao Shen, Shun Li, Xi Zhao, and Jianhua Zou. 2023. IncreAuth: Incremental learning based behavioral biometric authentication on smartphones. *IEEE Internet of Things Journal (IOTJ)* (2023).

[74] Ivo Sluganovic, Marc Roeschlin, Kasper B Rasmussen, and Ivan Martinovic. 2016. Using reflexive eye movements for fast challenge-response authentication. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 1056–1067.
[75] Chen Song, Aosen Wang, Kui Ren, and Wenyao Xu. 2016. Eyeveri: A secure and usable approach for smartphone user authentication. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.
[76] Mikhail Startsev and Raimondas Zemblys. 2023. Evaluating eye movement event detection: A review of the state of the art. *Behavior Research Methods* 55, 4 (2023), 1653–1714.
[77] Sophie Stephenson, Bijeeta Pal, Stephen Fan, Earlence Fernandes, Yuhang Zhao, and Rahul Chatterjee. 2022. Sok: Authentication in augmented and virtual reality. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 267–284.
[78] Tobii. 2023. Glasses, lenses and eye surgery. https://help.tobii.com/hc/en-us/articles/210249865-Glasses-lenses-and-eye-surgery.
[79] HTC VIVE. 2023. VIVE Pro Eye Specs. https://www.vive.com/hk/product/vive-pro-eye/specs/.
[80] HTC VIVE. 2023. VIVE Sense: Eye and Facial Tracking SDK. https://developer.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/.
[81] Robin Walker, David G Walker, Masud Husain, and Christopher Kennard. 2000. Control of voluntary and reflexive saccades. *Experimental Brain Research* 130 (2000), 540–544.
[82] Ruxin Wang, Long Huang, and Chen Wang. 2023. Low-effort VR Headset User Authentication Using Head-reverberated Sounds with Replay Resistance. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, 3450–3465.
[83] Xue Wang and Yang Zhang. 2021. Nod to auth: Fluent ar/vr authentication with user head-neck modeling. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.
[84] Richard P Wildes. 1997. Iris recognition: an emerging biometric technology. *Proc. IEEE* 85, 9 (1997), 1348–1363.
[85] Yi Xu, True Price, Jan-Michael Frahm, and Fabian Monrose. 2016. Virtual u: Defeating face liveness detection by building virtual models from your public photos. In *25th USENIX Security Symposium (USENIX Security 16)*. 497–512.
[86] Konstantina G Yiannopoulou, Georgios I Papagiannis, Athanasios I Triantafyllou, Panayiotis Koulouvaris, Aikaterini I Anastasiou, Konstantinos Kontoangelos, and Ioannis P Anastasiou. 2021. Neurological and neurourological complications of electrical injuries. *Neurologia i Neurochirurgia Polska* 55, 1 (2021), 12–23.
[87] Yongtuo Zhang, Wen Hu, Weitao Xu, Chun Tung Chou, and Jiankun Hu. 2018. Continuous authentication using eye movement response of implicit visual stimuli. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–22.
[88] Huadi Zhu, Wenqiang Jin, Mingyan Xiao, Srinivasan Murali, and Ming Li. 2020. Blinkey: A two-factor user authentication method for virtual reality devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–29.
[89] Huadi Zhu, Mingyan Xiao, Demoria Sherman, and Ming Li. 2023. SoundLock: A Novel User Authentication Scheme for VR Devices Using Auditory-Pupillary Response.. In *Network and Distributed System Security Symposium (NDSS)*.

Received February 2024; revised May 2024; accepted June 2024