IDIAP RESEARCH REPORT

VRBIOM: A NEW PERIOCULAR DATASET FOR BIOMETRIC APPLICATIONS OF HMD

Ketan Kotwal   Gokhan Ozbulak   Sébastien Marcel

Idiap-RR-03-2024

JULY 2024

Centre du Parc, Centre du Parc, Rue Marconi 19, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

VRBIOM: A NEW PERIOCULAR DATASET FOR BIOMETRIC APPLICATIONS OF HMD

Ketan Kotwal[1], **Ibrahim Ulucan**[1], **Gökhan Özbulak**[1,2],
Janani Selliah[1], **and Sébastien Marcel**[1,3]

[1]Idiap Research Institute, Switzerland
[2]École Polytechnique Fédérale de Lausanne, Switzerland
[3]Université de Lausanne, Switzerland
{firstname.lastname}@idiap.ch

ABSTRACT

With advancements in hardware, high-quality head-mounted display (HMD) devices are being developed by numerous companies, driving increased consumer interest in AR, VR, and MR applications. This proliferation of HMD devices opens up possibilities of wide range of applications beyond entertainment. Most commercially available HMD devices are equipped with internal inward-facing cameras to record the periocular areas. Given the nature of these devices and captured data, many applications such as biometric authentication and gaze analysis become feasible. To effectively explore the potential of HMDs for these diverse use-cases and to enhance the corresponding techniques, it is essential to have an HMD dataset that captures realistic scenarios.

In this work, we present a new dataset, called **VRBiom**, of periocular videos acquired using a Virtual Reality headset. The VRBiom, targeted at biometric applications, consists of 900 short videos acquired from 25 individuals recorded in the NIR spectrum. These $10s$ long videos have been captured using the internal tracking cameras of Meta Quest Pro at 72 FPS. To encompass real-world variations, the dataset includes recordings under three gaze conditions: steady, moving, and partially closed eyes. We have also ensured an equal split of recordings without and with glasses to facilitate the analysis of eye-wear. These videos, characterized by non-frontal views of the eye and relatively low spatial resolutions (400×400), can be instrumental in advancing state-of-the-art research across various biometric applications. The VRBiom dataset can be utilized to evaluate, train, or adapt models for biometric use-cases such as iris and/or periocular recognition and associated sub-tasks such as detection and semantic segmentation.

In addition to data from real individuals, we have included around 1100 presentation attacks constructed from 92 PA instruments. These PAIs fall into six categories constructed through combinations of print attacks (real and synthetic identities), fake 3D eyeballs, plastic eyes, and various types of masks and mannequins. These PA videos, combined with genuine (*bona-fide*) data, can be utilized to address concerns related to *spoofing*, which is a significant threat if these devices are to be used for authentication.

The VRBiom dataset is publicly available for research purposes related to biometric applications only.

1 Introduction

The rise of head-mounted displays (HMDs) in recent years has significantly transformed the way we experience digital content. With its Meta Quest series of virtual reality (VR) headsets, the Meta (previously FaceBook) is the largest headset platform as of 2024[1]. Recently, Apple released its Mixed reality (MR) headset named Apple Vision

[1]https://www.meta.com/quest

Pro[2], with specific focus on the spatial computing aspects of the device. The Sony Interactive Entertainment has been actively working on VR for gaming applications with its PlayStation (PS VR) series[3]. Initially, these devices were primarily designed for immersive entertainment, offering users experiences in augmented, virtual, or mixed reality. However, the applications of HMDs extend far beyond gaming and entertainment, encompassing diverse fields such as education [1, 2], professional training [3, 4], healthcare [5, 6, 7], and biometric authentication [8, 9].

Besides its core functionality, an interesting feature of HMDs is the integrated multiple internal cameras. These cameras, typically located in the surroundings of the user's eye region, enhance the immersive experience by tracking user's eye movements and capturing images and videos of the user's eyes and surrounding regions. This source of data opens up a wide range of potential applications to be explored, including biometric authentication. In biometrics, the HMDs can be instrumental in user recognition based on their iris and/or periocular traits [9]. Such applications can be employed for both identification and verification purposes ensuring that the user accessing the device is indeed who they claim to be. In addition to the recognition accuracy, in biometric applications, safeguarding against attacks and ensuring robustness is critical. The presentation attacks (PA), also known as *spoofing*, pose a serious challenge to biometric systems. These attacks can involve the use of masks, synthetic eyes, contact lenses, or printed images to deceive the system into granting unauthorized access [10, 11]. Therefore, developing effective presentation attack detection (PAD) or *anti-spoofing* mechanisms is essential for the practical use of biometric authentication systems in HMDs. Beyond biometrics, HMDs have significant potential in other applications, such as semantic segmentation of the eye region [8, 12]. This is particularly relevant for industrial, entertainment, and biomedical applications where continuous tracking of the gaze and eye movement is necessary. HMDs can facilitate near-continuous gaze tracking, which has applications ranging from interactive gaming to medical diagnostics [13, 14, 15]. Fig. 1 depicts some of the possible biometric use-cases of the HMD data.

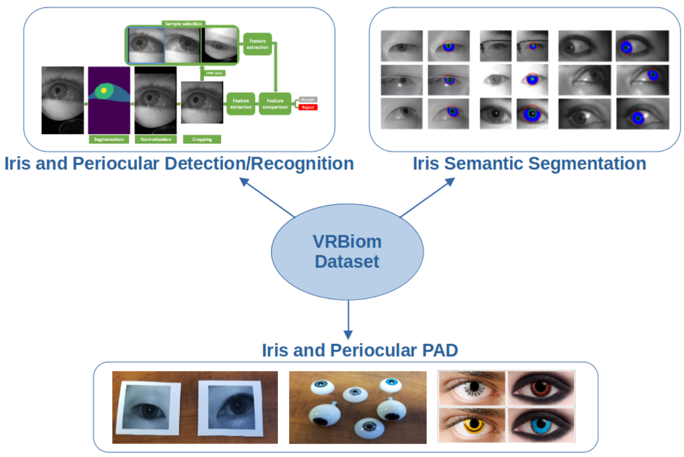

Figure 1: Examples of biometric use-cases offered by the HMD data, in particular by the VRBiom dataset (The example images are obtained from [16, 17, 18]).

The aforementioned applications have been a part of research and commercial deployments for several years. However, most existing research and user-applications have focused on data captured by sensors positioned to capture a frontal view of the eye. The design of recent popular HMDs typically places displays in front of the user's eyes for an immersive experience, and thereby necessitates the eye tracking cameras to be positioned in the surrounding areas. This placement results in an oblique, non-frontal view of the eye region, posing unique challenges for biometric applications. Additionally, the cameras integrated into these HMDs are often of lower resolution and must operate within constrained resources. Moreover, the fitment of HMDs varies based on the individual's inter-eye distance and nose bridge shape, resulting in variations in the captured regions for different users.

These factors indicate the need for specific studies and datasets acquired directly from HMD devices to address the above challenges. Such datasets can be invaluable for multiple purposes: (*i*) benchmarking the performance of existing

[2] https://www.apple.com/apple-vision-pro

[3] https://www.playstation.com/en-us/ps-vr2

methods/models on HMD data, (*ii*) fine-tuning models to address domain shifts (introduced by different capturing angle, devices, resolutions, etc), and (*iii*) developing new methods tailored to the unique requirements of HMDs.

To address these challenges, in the realm of biometrics, we have created **VRBiom**: a new ***V****irtual* ***R****eality dataset for* ***Biom****etric Applications* acquired from the Meta Quest Pro. One subset of this dataset comprises 900 short videos, each lasting nearly 10 seconds, collected from 25 individuals under various conditions, including steady gaze, moving gaze, and half-closed eye settings. To ensure evaluation and development of various use-cases in realistic scenarios, half of these videos were recorded with subjects wearing glasses, while the remaining half were recorded without glasses. These videos present potential for a wide range of applications, such as biometric recognition and semantic segmentation. It should be noted that conventional biometric recognition pipelines comprise of sequence of tasks, including region of interest (RoI) detection, segmentation, and orientation. The presented dataset may be harnessed to enhance each of these tasks separately. We have also captured recordings using different types of 3D masks, combined with synthetic eyes such as 2D print-outs and 3D fake eyes. These 1104 recordings, also called as presentation attacks (PAs), comprise other subset of the VRBiom dataset meant for addressing concerns related to PAD and anti-spoofing measures. The overall VRBiom dataset consists of 2004 recordings of periocular (left and right) regions captured in near-infrared (NIR) spectrum for $10s$ at 72 FPS.

The present release of the VRBiom consists of videos and subject-level (identity) labels. It also provides details of the scenarios and PA instruments used. We do not provide pixel-level annotations for semantic details. The contributions of our work can be summarized as below:

- We have created and publicly released the VRBiom dataset[4], a collection of more than 2000 periocular videos acquired from a VR device. To the best of our knowledge, this is the first publicly available dataset featuring realistic, non-frontal views of the periocular region for a variety of biometric applications.
- The VRBiom dataset offers a range of realistic and challenging variations for the benchmarking and development of biometric applications. It includes recordings under three gaze conditions: steady gaze, moving gaze, and partially closed eyes. Additionally, it provides an equal split of recordings with and without glasses, facilitating the analysis of the impact of eyewear on iris/ periocular recognition.
- As a part of VRBiom, we have also released more than 1100 PA videos constructed from 92 attack instruments. These attacks, incorporating different combinations of PAIs such as prints, fake 3D eyeballs, and various masks, provide a valuable resource for advancing PAD research in a VR setup.

The organization of this paper is structured as follows: Section 2 provides a brief review of iris/periocular datasets and HMD-based datasets. Section 3 explains details of the VRBiom dataset, including data collection methods and associated challenges. In Section 4, we discuss potential use cases of the dataset. Finally, we summarize the work with Section 5.

2 Related Datasets

In this section, we briefly present some commonly used datasets for iris/periocular biometrics, followed by review of recent datasets collected using HMD devices.

2.1 Datasets for Iris/ Periocular Biometrics

One of the most prominent and commonly used iris datasets is the CASIA-Iris series[5], collected by the Chinese Academy of Sciences, which consists of four main versions denoted with suffixes V1–V4 [19]. The first version, V1, released in 2002, comprised 756 images taken from 108 subjects at a resolution of 320×280. In 2004, CASIA-IrisV2 was released, containing 2,400 images from 120 individuals at VGA resolution (640×480). The third version included three different subsets: Interval, Lamp, and Twins. The Casia-Interval subset was captured from 249 subjects in an indoor environment, while the Lamp subset provided $16k$ images at VGA resolution. The Twins subset contained recordings of 200 twins, primarily children, in an outdoor environment [20]. The latest version, CASIA-IrisV4, is the most comprehensive, including 54,601 images from 2,800 subjects distributed across six subsets namely CASIA-Iris-Interval, Lamp, Twins, Distance, Thousand, and Syn [19, 21].

The Iris Liveness Detection Competition (LivDet-Iris) [22, 23] dataset combines multiple datasets designed for liveness detection in the context of iris PAD. The 2017 competition utilized four different datasets created by Clarkson, Warsaw, Notre Dame (ND), Indraprastha Institute of Information Technology Delhi (IIIT-Delhi), and West Virginia University (WVU) [22]. The Clarkson and Warsaw datasets mainly focus on print attacks, consisting of 8,095 and

[4] The VRBiom dataset can be downloaded from https://www.idiap.ch/dataset/vrbiom.
[5] http://www.cbsr.ia.ac.cn/english/Databases.asp

12,013 images, respectively. The ND CLD 2015 and IIITD-WVU datasets comprise more than $7k$ images each, acquired at VGA resolution. The recent edition of the competition, held in conjunction with IJCB 2023, aimed at creating more challenging attacks [23]. This dataset included attack instruments such as print, contact lenses, electronic displays, fake/prosthetic eyes, and synthetic iris of varying quality.

Researchers at the University of Beira have released a series of UBIRIS datasets of iris images. The UBIRIS v1 dataset provides 1,877 images acquired under less constrained imaging conditions [24], whereas UBIRIS v2 offers more than $11k$ images captured at a distance [25]. Another version, UBIPr, derived from UBIRIS v2, provides wider regions suitable for periocular recognition. They have also released pixel-level masks of semantic elements such as the iris and sclera, making it a useful resource for segmentation applications.

Additionally, the Multimedia University (MMU) released two iris image datasets in 2010 that are publicly available. The first version, MMU-v1, consists of 450 images obtained from 45 subjects, while the subsequently released MMU-v2 provides 995 images from 100 subjects [26, 27].

Table 1 summarizes the iris and periocular biometric datasets, providing details related to samples and attacks where applicable.

Datasets and Subsets		#Subjects	#Images	Resolution	Types of Attacks	Collection Year	Institution(s)
CASIA-Iris series [19]	CASIA-IrisV1	108*	756	320×280	No	2002	Chinese Academy of Sciences
	CASIA-IrisV2	120**	2,400	640×480	No	2004	
	CASIA-IrisV3	700	22,034	Various	No	2005	
	CASIA-IrisV4	2,800	54,601	Various	No	2010	
LivDetIris 2017 [22]	ND CLD 2015	-	7,300	640×480	contact lens (live)	2015	University of Notre Dame
	IIITD-WVU	-	7,459	640×480	Print	2017	IIIT Delhi
	Clarkson	50	8,095	640×480	Print and contact lens (live)	2017	Clarkson University
	Warsaw	457	12,013	640×480	Print and contact lens (print, live)	2017	Warsaw University
UBIRIS [24, 25]	UBIRIS-V1	241	1,877	400×300	No	2004	University of Beira
	UBIRIS-V2	261	11,102	800×600	No	2009	
MMU [26, 27]	MMU-V1	45	450	-	No	2010	Multimedia University
	MMU-V2	100	995	-	No	2010	
LivDet-Iris 2023 [23]	-	-	13,332***	Various	Print, contact lens, fake/ prosthetic eyes, and synthetic iris	2023	Multiple Institutions

Table 1: Overview of iris and periocular biometric datasets. (*number of classes, **number of eyes, ***number of test data).

2.2 HMD Datasets for Iris/ Periocular Biometrics

Due to the inherent challenges in collecting data using HMD devices and the limitations of hardware quality, there have been fewer attempts to gather such datasets. However, with recent advancements in hardware technology and the increasing number of companies releasing HMD headsets, these types of datasets are gaining traction in both research and end-user applications.

Released in 2012, the Point of Gaze (PoG) dataset is one of the HMD datasets created for gaze detection and head pose studies [28]. It comprises data from 20 subjects recorded at a resolution of 768×480. The Labelled Pupils in the Wild (LPW) dataset contains 66 eye-region videos at VGA resolution captured at 95 FPS [29]. As its name suggests, the LPW dataset features several variations in illumination, eyewear (glasses and contact lenses), and makeup, along with gaze directions. Kim *et al.* introduced the NVGaze dataset, which was collected from 35 subjects [30]. This dataset focuses on gaze estimation and includes approximately 2.5 million *bona-fide* images at VGA resolution.

The Open Eye Dataset (OpenEDS) has two editions: OpenEDS 2019 and 2020, both released by Facebook Reality Labs, which have significantly advanced the field of HMD-based eye-tracking research. The OpenEDS 2019 dataset [31] comprises videos and images from 152 subjects and is partially annotated for semantic segmentation. The 2020 version [32] includes two sub-datasets (the Gaze Prediction and Eye Segmentation Datasets) from 80 subjects, comprising approximately $580k$ images with a resolution of 640×400 pixels, captured at 100 FPS.

The creators of OpenEDS also organized a competition (challenge) to improve performance in tracking and segmentation tasks. While OpenEDS is similar to our presented work, it differs in two main aspects: first, OpenEDS data is acquired from eye-facing cameras, yielding frontal eye views, whereas our dataset uses a consumer-level HMD capturing eyes and periocular regions from non-frontal angles. Second, OpenEDS includes recordings only from real (*bona-fide*) subjects, while our dataset incorporates a variety of presentation attacks (PAs), making it valuable for PAD research as well [31, 32].

Our dataset, VRBiom, consists of 900 short videos from 25 participants and more than 1100 videos across six categories of attack instruments. Each video, with a spatial resolution of 400×400, is approximately 10 seconds long and captured at 72 FPS. Table 2 provides a brief comparison of the publicly available HMD datasets for research.

Datasets	**Collection Year**	**#Subjects**	**#Images**	**Resolution**	**FPS** (Hz)	**Presentation Attacks**	**Profile**	**Synthetic Eye Data**
Point of Gaze (PoG) [28]	2012	20	–	768×480	30	✗	Frontal	✗
LPW [29]	2016	22	130,856	640×480	95	✗	Frontal	✗
NVGaze [30]	2019	35	2,500,000	640×480	30	✗	Frontal	✓
OpenEDS2019 [31]	2019	152	356,649	640×400	200	✗	Frontal	✗
OpenEDS2020 [32]	2020	80	579,900	640×400	100	✗	Frontal	✗
VRBiom (this work)	2024	25	1,262,520	400×400	72	✓	Non-Frontal	✓

Table 2: Summary of the iris/ periocular datasets acquired by the HMDs.

3 The VRBiom Dataset

In this section, we describe the process of acquiring both *bona-fide* and PA data, along with relevant statistics. We also discuss some of the challenges encountered in the creation of the VRBiom dataset.

3.1 Collection Setup

A total of 25 subjects, aged between 18 and 50 and representing a diverse range of skin tones and eye colors, participated in the data collection process. Each participant was briefed on the project objectives and provided their consent through a signed consent form. The data acquisition for each participant was conducted in a single session lasting approximately 20 minutes. We divided the recordings into two sub-sessions: the first without the subject wearing glasses and the second with glasses. For each sub-session, we captured three recordings involving different gaze variations: steady gaze, moving gaze, and partially closed eyes. Each video had a duration of approximately 10 seconds, recorded at a frame rate of 72 frames per second (FPS). The videos have a spatial resolution of 400×400 pixels and were captured in a single-channel in the NIR spectrum. It was often observed that for nearly first second of the recording, the recordings were over-exposed to the NIR illumination. Therefore, we recommend discarding the first 70–80 frames from processing.

The iris/periocular videos in the VRBiom dataset were captured by the inward facing cameras located in the Meta Quest Pro headset. For *bona-fide* recordings, the subjects wore the HMD device, whereas for recording of PAs, we systematically positioned the Meta Quest Pro on the attack instrument (as described later in this section). The Meta Quest Pro was connected to a workstation via USB-C cable which facilitated both charging and the transfer of recording commands and data. Figs. 2a and 2b depict the recording setups for *bona-fide* and PAs, respectively.

To streamline the data collection, we developed the Meta Quest Pro Data Capture App (app) (Refer to Fig. 2c for GUI). This app enabled us to set file names, manage recordings, and transfer files from the Meta Quest Pro to our workstation efficiently. Additionally, the app provided tools to convert the video recordings from the proprietary *vrs* format to the conventional *avi* format, which is compatible with most common video players and softwares. Through the capture app, we also ensured a specific naming convention for each recoded video. The naming convention consisted of the following fields: Type (*bona-fide* or PA), subject ID (or PA ID), Eye (left or right), glasses (without or with), type of gaze (Steady, Moving, or Partially closed), claimed ID (for PAs), details of PAI (for PAs), recording ID, and a random suffix. These fields, with their mapped values, were used to create the filename for each video. More details on this can be found in the actual dataset.

(a) (b) (c)

Figure 2: Setup for the dataset collections: (a) for *bona-fide* recordings, the subjects wore the HMD devices, (b) for PA recordings, the HMD device was carefully placed on the temple region of the attack instrument (mannequin, in this example), and (c) the Meta Quest Pro Data Capture App used for data collection.

3.2 *bona-fide* Recordings

Before starting the recording, we ensured that the HMD (Meta Quest Pro) is securely and comfortably fitted around the subject's head, providing them with an immersive virtual experience. Considering the diverse array of use-cases, we captured videos under three different gaze variations:

- **Steady Gaze:** The subject maintains a nearly fixed gaze position by fixating their eyes on a specific (virtual) object.
- **Moving Gaze:** The subject's gaze moves freely across the scene.
- **Partially Closed Eyes:** The subject keeps their eyes partially closed without focusing on any particular gaze.

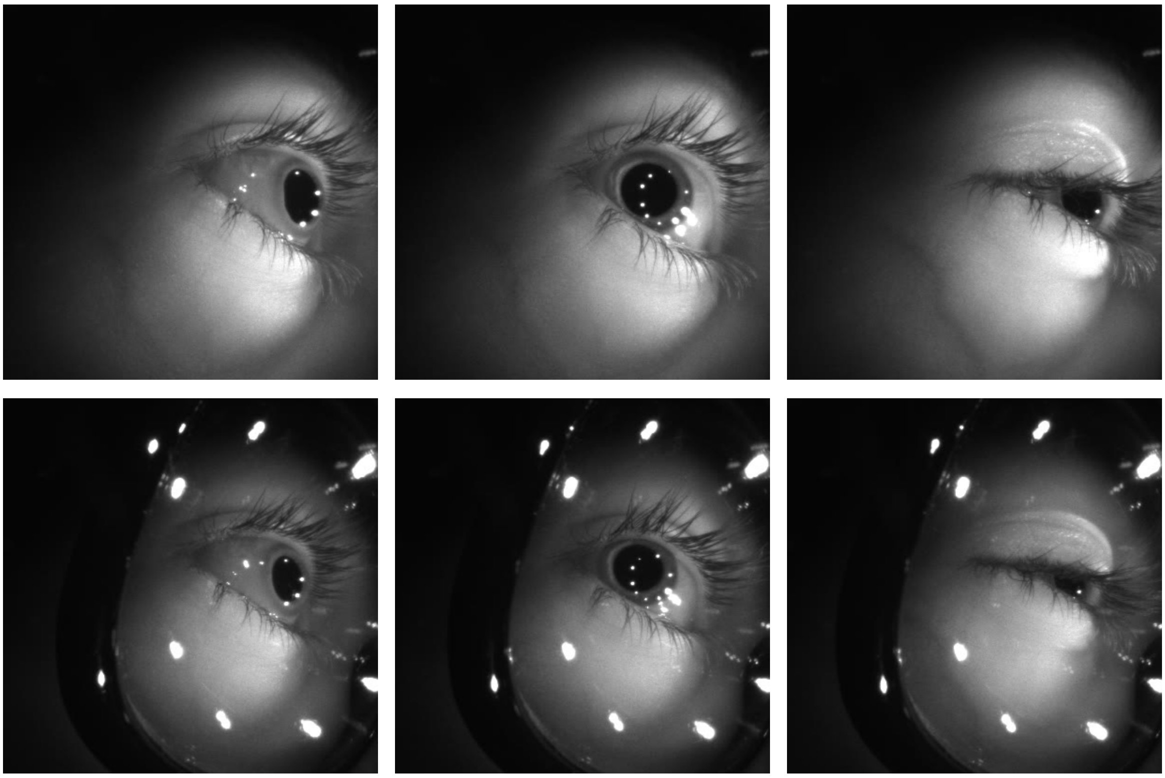

Figure 3: Samples of *bona-fide* recordings from VRBiom dataset. Each row presents a sample of steady gaze, moving gaze, and partially closed eyes (from left to right). Top and bottom rows refer to recordings without and with glasses, respectively.

#Subjects	#gaze	#glass	#repetitions	#eyes	#total
25	3	2	3	2	900

Table 3: Summary of *bona-fide* recordings in the VRBiom dataset.

These variations were recorded under two eye-wear conditions: with glasses and without glasses. If the subjects did not have their own medical glasses, they were provided with a pair of fake glasses. Each gaze variation and glasses condition was repeated three times to ensure a robust dataset.

Fig. 3 provides samples of each of the aforementioned variations. The HMD is equipped with cameras that simultaneously record videos of both eyes. Therefore, for each recording, two videos are captured concurrently- one for the left eye and one for the right eye. In total, we collected 900 *bona-fide* videos from 25 subjects, which can be summarized as follows:

Thus, the structure and coverage of the VRBiom ensure that each subject's iris/periocular region is recorded across various realistic scenarios, providing a valuable resource for researchers to analyse and develop biometric applications.

3.3 PA Recordings

Most of the PAs were constructed by combining two categories of attack instruments: those targeting the eyes and those targeting the periocular region. For the eyes, we used a variety of instruments including fake 3D eyes (eyeballs), printouts from synthetic and real identities, and plastic-made synthetic eyes. For the periocular region, we employed mannequins and 3D masks made of different materials. These masks also served as a fake head where the HMD could be securely placed for recording.

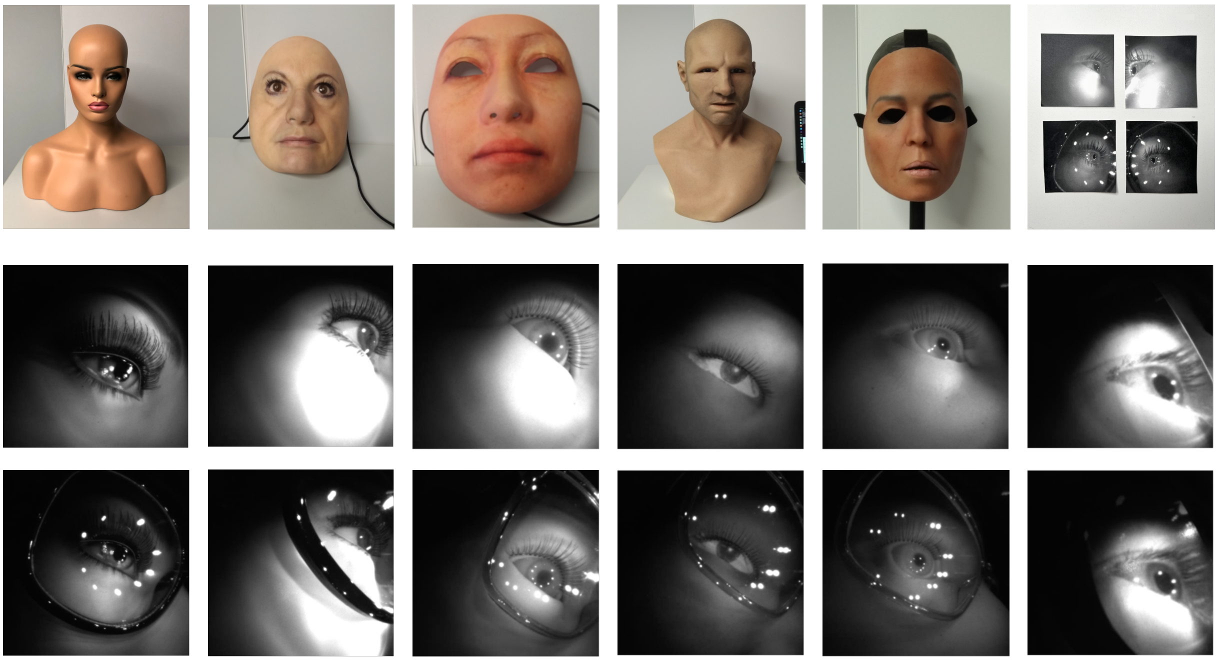

Figure 4: Samples of PA recordings from VRBiom. The top row represents the PAI captured in RGB (visible) spectrum, while middle and bottom rows depict the NIR recordings without and with glasses, respectively, as acquired by the internal (right) camera of the Meta Quest Pro. From left to right, each column presents a sample of the type of PAIs belonging to the attack series from 2–7.

Using these combinations, we created six categories of PAIs, each with a variable number of attack instruments. Samples of these categories are provided in Fig. 4, with a brief description as follows:

- **Mannequins:** We used a collection of seven mannequins made of plastic, each featuring its own eyes. These mannequins are generic (fake heads) as they do not represent any specific *bona-fide* identity. Unlike the requirements of additional components for other PAIs to provide a stable platform or extra equipment to stabilize HMDs, these mannequins are particularly advantageous due to their integrated and stable platform for placing HMDs.

- **Custom rigid masks (type-I):** This category includes ten custom rigid masks, comprising five men and five women. The term custom implies that the masks were modeled after real individuals. Manufactured by REAL-f Co. Ltd. (Japan), these masks are made from a mixture of sandstone powder and resin. They represent real persons (though not part of the *bona-fide* subjects in this work) and include eyes made of similar material with synthetic eyelashes attached.
- **Custom rigid masks (type-II):** We used another collection of 14 custom rigid masks to construct the PAs for the VRBiom. These masks differ from the previous category in two ways: first, they are made of amorphous powder compacted with resin material by Dig:ED (Germany); second, they have empty spaces at the eye locations, where we inserted fake 3D glass eyeballs to construct an attack.
- **Generic flexible masks:** This category includes twenty flexible masks made of silicone. These full-head masks do not represent any specific identity, hence termed generic. The masks have empty holes for eyes, where we inserted printouts of periocular regions from synthetic identities. Using synthetic identities alleviates privacy concerns typically associated with creating biometric PAs.
- **Custom flexible masks:** These silicone masks represent real individuals. Similar to the previous attack categories, these masks have holes at the eye locations, where we inserted fake 3D eyes to create the attacks. We have 16 PAIs in this category.
- **Vulnerability Attacks:** For each subject in the *bona-fide* collection, we created a print attack. For both eye-wear conditions (without and with glasses), we manually selected an appropriate frame from each subject's *bona-fide* videos and printed it at true-scale on a laser printer. The printouts were cut into periocular crops and placed on a mannequin to resemble a realistic appearance when viewed by the tracking cameras of the HMD. The so-obtained 25 print PAs can be used to assess the vulnerability of the corresponding biometric recognition system. These attacks simulate scenarios where an attacker gains access to (unencrypted) data of authorized individuals and constructs simple attacks using printouts.

Table 4 summarizes the details of the VRBiom dataset comprising 900 *bona-fide* and 1104 PA videos. It also provides the naming conventions used to indicate the type of PAI.

Type	**PA Series**	**Subtype**	**# Identities**	**# Videos**	**Attack Types**
bona-fide	–	[steady gaze, moving gaze, partially closed] × [glass, no glass]	25	900	–
Presentation Attacks	2	Mannequins	7	84	Own eyes (same material)
	3	Custom rigid mask (type I)	10	120	Own eyes (same material)
	4	Custom rigid mask (type II)	14	168	Fake 3D eyeballs
	5	Generic flexible masks	20	240	Print attacks (synthetic data)
	6	Custom silicone masks	16	192	Fake 3D eyeballs
	7	Print attacks	25	300	Print attacks (real data)

Table 4: Details of *bona-fide* and different types of PAs from the VRBiom dataset. Each video was recorded at 72 FPS for approximately $10s$.

3.4 Challenges

Here we briefly discuss some of the challenges encountered during data collection, and the corresponding solutions employed.

- Eyelashes: During the initial data capture experiments, we observed that the subjects' eyelashes appeared as a prominent feature, especially when capturing the periocular region. The absence of eyelashes in most PAIs, where prints or 3D eyeballs were used to construct the attack, can be considered as an easy to distinguish yet unrealistic feature. To create a more realistic scenario, we decided to use false eyelashes from standard makeup kits.
- PA recordings: Recording the PAs was challenging due to the lack of real-time feedback from the device or capture app. To achieve the correct positioning and angle for various PAIs, it often required multiple attempts and recording trials for data collectors to ensure the desired quality of the captured data.

- NIR Camera's Over-exposure: The proximity between the attack instrument and the NIR illuminator/receiver often resulted in over-exposed recordings. Given that VR headsets are designed to fit closely on the forehead, we had a limited room for adjustment. However, we attempted minor adjustments to the HMD's orientation and the placement of the PAI to mitigate exposure issues to a considerable extent.
- Synthetic Eyes: One challenge we faced was the variability in the identity of synthetic eyes. To ensure accurate biometric measurements, we needed to carefully select synthetic eyes that closely resemble human eyes.
- Print Attack: The print-out attacks designed for vulnerability assessment are inherently 2-dimensional, whereas the periocular region's appearance is 3D. To simulate realistic appearance, we employed small paper balls to roll the printed attacks, thereby providing a structural view akin to 3D representation.

4 Potential Use-Cases of the VRBiom Dataset

This section outlines the potential use-cases of the newly created VRBiom dataset, emphasizing its applicability in biometric applications such as iris and periocular detection/recognition, presentation attack detection (PAD), as well as intriguing tasks (or sub-tasks) associated with biometrics such as semantic segmentation focusing on the eye, iris, and sclera regions. The overview of these use-cases is illustrated in Fig. 1

The use cases discussed here, although well-established in research and development, are highlighted more as possibilities than substantiated claims of direct applicability. The VRBiom dataset can serve as a benchmark to evaluate the effectiveness of current HMD data for these applications. Alternatively, it can be instrumental in developing (or fine-tuning) applications specifically for data obtained from HMD devices. The present data can be characterised by non-frontal views of periocular regions, relatively low spatial resolutions, and limited resources. Thus, some applications discussed in this section may prove to be quite challenging. This, however, presents interesting research problems to transform this data into better quality as well as to develop robust applications.

Needless to mention, the advances in HMD technology will not only enhance the quality of the data but also expand the scope of applications.

4.1 Iris and Periocular Recognition

The VRBiom dataset comprises 900 video recordings from 25 *bona-fide* subjects. Each subject participated in sessions combining three gaze and two eye-wear variations, thus providing a valuable resource for sampling and analyzing iris and periocular regions for detection and recognition tasks. When sampled at the frame level, videos with moving gaze can be treated as individual frames with varied gazes. These individual frames facilitate the localization and enrollment of the eye region, making them suitable for both detection and recognition purposes. Detection involves localizing specific regions of interest (RoI), such as the iris, eye, or periocular regions, while recognition can be performed in both *1:1* verification and *1:N* identification scenarios.

Based on the design of the pipeline, recent iris/periocular recognition methods can be categorized into feature-extraction based or classification based. Feature-extraction based methods employ models, often deep convolutional neural network (CNN) architectures, as feature extractors to obtain a compact representation (*embedding*) of a pre-processed input, (*i.e.*, iris/periocular image). In these methods, preprocessing and matching/scoring are typically separated from feature extraction. Classification based methods, on the other hand, treat the recognition problem in an end-to-end manner, training the iris dataset in a supervised learning setting.

An overview of iris recognition methods using handcrafted features and classical machine learning techniques can be found in [33]. A recent work by Nguyen *et al.* provides a systematic survey of deep learning (DL)-based methods for iris recognition [27]. Another comprehensive review of iris recognition methods is presented in [34], which details methods from both categories and various preprocessing techniques for the first category (non end-to-end methods). Alonso-Fernandez and Bigun discuss several handcrafted feature-based methods for periocular recognition, categorizing them into texture, shape, and color-based methods [35].

Iris recognition remains a topic of significant interest within the biometric community. Over the past decade, several challenges have been organized to advance iris recognition performance [36, 37]. The challenge described in [37] was specifically dedicated to mobile iris recognition. Boutros *et al.* benchmarked the OpenEDS [31]- an HMD dataset– for iris recognition [16, 9]. Although this dataset has been acquired from an HMD device, it captures nearly frontal view of the eye regions making it easy to detect, localize, and recognize an individual.

The use of HMD data for biometric recognition is still in its early stages and is yet to gain mainstream attention. A major challenge posed by HMD-based iris detection is the oblique view of the RoI, which often results in localization failures and leads to failure to acquire (FtA). The non-frontal view, even after normalization, may cause feature/data distortion. This, combined with the relatively small size of the RoI, results in the loss of fine, subtle features that may present discriminatory information. Several quality metrics for iris and periocular images, including low-quality images, have been presented in [33]. A study of the quality assessment of HMD-based data for recognition purposes may provide useful insights towards designing new methods.

4.2 Iris and Periocular PAD

For a biometric authentication system to be practically deployable, its resistance to presentation attacks (PAs) is crucial. The importance of PAD for iris and periocular traits has been well recognized by the biometric community. Several works by A. Czajka, K. Bowyer, and colleagues present comprehensive reviews of iris PAD methods and datasets [11, 38]. However, these works do not discuss datasets or PAD methods for HMD-based datasets, likely due to a lack of relevant data. With VRBiom, we provide the first PAD dataset acquired using HMD devices, containing approximately 1100 short videos of PAs. Through a combination of attack instruments for the eyes and surrounding regions (refer to Section 3), we have created a variety of attack scenarios for PAD tasks. The inclusion of print attacks of the *bona-fide* samples allows for the vulnerability analysis of different PAD methods.

Several studies have also proposed approaches that combine both iris and periocular regions for PA detection [39, 18]. Such approaches can be explored for the VRBiom dataset, if the detection and localization of iris does not yield satisfactory results.

LivDet-Iris[6] is a well-known competition regularly organized to compare liveness detection methods (*i.e.*, to differentiate between real human and fake samples). The fifth installment of this competition was held in conjunction with IJCB in 2023 [23]. Despite the participation of reputable academic institutions, the average classification error rates were as high as 22–37% (the variation refers to the weighing mechanism of different PAIs). Although the quality of data and types of attacks in the competition datasets and VRBiom dataset differ, this highlights the challenging and unresolved nature of the iris and/or periocular PAD problem.

4.3 Semantic Segmentation

Semantic segmentation of the eye to localize components such as the iris and sclera has demonstrated its usefulness in both biometric and non-biometric tasks. This segmentation process is often an integral crucial stage in recognition and PAD-related tasks discussed in the previous section. The creators of the OpenEDS dataset have also highlighted segmentation as a key application of interest [31]. Segmentation of individual traits, such as the iris [40] and sclera [41], has received considerable attention in the literature. Recently, multi-class semantic segmentation–which involves the simultaneous detection and localization of various parts of the eye region– has gained interest as well [42, 43, 44].

Similar to the previous use-cases, the topic of eye segmentation has been further advanced by numerous competitions. For instance, challenges such as NIR-ISL [17] and SSRBC [45] focus specifically on iris and sclera segmentation, respectively. The VRBiom dataset (after annotations) can also be used to benchmark the performance of existing segmentation methods, with reference to biometric applications, on realistic HMD data captured in the NIR spectrum. Given the variation in the exact angle of capture due to differences in subjects' face shapes and inter-eye distances, the frames in the VRBiom dataset present a challenging dataset. As typical segmentation methods are usually trained on frontal views of the eye, one may be required to address this domain gap by designing implicit or explicit affine transformations.

5 Summary

In this work, we introduced the VRBiom dataset, a novel dataset of periocular videos captured using the VR device, Meta Quest Pro. This dataset is the first publicly available resource offering realistic, non-frontal views of the periocular region, comprising 900 short videos from 25 subjects and 1104 presentation attack (PA) videos using 92 different attack instruments. The dataset includes diverse and challenging scenarios such as steady gaze, moving gaze, and partially closed eyes; with an equal split of recordings with and without glasses.

The HMD-captured data are relatively novel and present significant potential for various applications that are yet to be fully explored. However, these also pose unique challenges such as non-frontal views, user-specific fitment issues, and low resolutions. With the VRBiom dataset, we hope to provide a valuable resource for understanding and advancing biometric applications associated with AR and VR devices. The inclusion of various attack recordings also provides an opportunity to enhance *anti-spoofing* measures for VR-based authentication systems.

[6] https://livdet.org

We have outlined the data collection process and associated challenges, providing useful insights for researchers aiming to acquire similar data from VR devices. The VRBiom dataset is publicly available to support advancements in biometric research, including authentication, semantic segmentation, and presentation attack detection (PAD), along with related sub-tasks.

Acknowledgements

Authors would like to thank Yannick Dayer (Idiap Research Institute) for creating a data capture application. We would like to gratefully acknowledge funding support from the Meta program "Towards Trustworthy Products in AR, VR, and Smart Devices" (2022)[7] and the Swiss Center for Biometrics Research and Testing.

References

[1] Jaziar Radianti, Tim A Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Computers & education*, vol. 147, pp. 103778, 2020. 2

[2] Dorota Kamińska et al., "Virtual reality and its applications in education: Survey," *Information*, vol. 10, no. 10, pp. 318, 2019. 2

[3] Biao Xie et al., "A review on virtual reality skill training applications," *Frontiers in Virtual Reality*, vol. 2, pp. 645153, 2021. 2

[4] Feng-Kuang Chiang, Xiaojing Shang, and Lu Qiao, "Augmented reality in vocational training: A systematic review of research and applications," *Computers in Human Behavior*, vol. 129, pp. 107125, 2022. 2

[5] Andy Wai Kan Yeung et al., "Virtual and augmented reality applications in medicine: analysis of the scientific literature," *Journal of medical internet research*, vol. 23, no. 2, pp. e25499, 2021. 2

[6] Mohd Javaid and Abid Haleem, "Virtual reality applications toward medical field," *Clinical Epidemiology and Global Health*, vol. 8, no. 2, pp. 600–605, 2020. 2

[7] Irene Suh, Tess McKinney, and Ka-Chun Siu, "Current perspective of metaverse application in medical education, research and patient care," in *Virtual Worlds*. MDPI, 2023, vol. 2, pp. 115–128. 2

[8] Fadi Boutros, Naser Damer, Kiran Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper, "Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation," *Image and Vision Computing*, vol. 104, pp. 104007, 2020. 2

[9] Fadi Boutros, Naser Damer, Kiran Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper, "On benchmarking iris recognition within a head-mounted display for ar/vr applications," in *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2020, pp. 1–10. 2, 9

[10] Aythami Morales, Julian Fierrez, Javier Galbally, and Marta Gomez-Barrero, "Introduction to iris presentation attack detection," *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*, pp. 135–150, 2019. 2

[11] Adam Czajka and Kevin W Bowyer, "Presentation attack detection for iris recognition: An assessment of the state-of-the-art," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–35, 2018. 2, 10

[12] Zhimin Wang, Yuxin Zhao, Yunfei Liu, and Feng Lu, "Edge-guided near-eye image analysis for head mounted displays," in *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 11–20. 2

[13] Viviane Clay, Peter König, and Sabine Koenig, "Eye tracking in virtual reality," *Journal of eye movement research*, vol. 12, no. 1, 2019. 2

[14] Sebastian Kapp, Michael Barz, Sergey Mukhametov, Daniel Sonntag, and Jochen Kuhn, "ARETT: augmented reality eye tracking toolkit for head mounted displays," *Sensors*, vol. 21, no. 6, pp. 2234, 2021. 2

[15] Efe Bozkir et al., "Eye-tracked virtual reality: A comprehensive survey on methods and privacy challenges," *arXiv preprint arXiv:2305.14080*, 2023. 2

[16] Fadi Boutros, Naser Damer, Kiran Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper, "Periocular biometrics in head-mounted displays: A sample selection approach for better recognition," in *International Workshop on Biometrics and Forensics (IWBF)*, 2020, pp. 1–6. 2, 9

[7] Meta Research Program

[17] Caiyong Wang et al., "Nir iris challenge evaluation in non-cooperative environments: Segmentation and localization," in *2021 IEEE International joint conference on biometrics (IJCB)*. IEEE, 2021, pp. 1–10. 2, 10

[18] Steven Hoffman, Renu Sharma, and Arun Ross, "Iris + Ocular: Generalized iris presentation attack detection using multiple convolutional neural networks," in *International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–8. 2, 10

[19] Chinese Academy of Sciences, "Casia iris image database," 2002, http://biometrics.idealtest.org. 3, 4

[20] Luiz A Zanlorensi, Rayson Laroca, Eduardo Luz, Alceu S Britto Jr, Luiz S Oliveira, and David Menotti, "Ocular recognition databases and competitions: A survey," *Artificial Intelligence Review*, vol. 55, no. 1, pp. 129–180, 2022. 3

[21] Lubos Omelina, Jozef Goga, Jarmila Pavlovicova, Milos Oravec, and Bart Jansen, "A survey of iris datasets," *Image and Vision Computing*, vol. 108, pp. 104109, 2021. 3

[22] David Yambay et al., "Livdet iris 2017–iris liveness detection competition 2017," in *IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 733–741. 3, 4

[23] Patrick Tinsley et al., "Iris liveness detection competition (livdet-iris)–the 2023 edition," in *IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2023, pp. 1–10. 3, 4, 10

[24] Hugo Proença and Luís A Alexandre, "Ubiris: A noisy iris image database," in *International Conference on Image Analysis and Processing (ICIAP)*. Springer, 2005, pp. 970–977. 4

[25] Hugo Proença, Silvio Filipe, Ricardo Santos, Joao Oliveira, and Luis A Alexandre, "The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2009. 4

[26] Chuan Chin Teo, Han Foon Neo, GKO Michael, Connie Tee, and KS Sim, "A robust iris segmentation with fuzzy supports," in *International Conference on Neural Information Processing (ICONIP)*. Springer, 2010, pp. 532–539. 4

[27] Kien Nguyen, Hugo Proença, and Fernando Alonso-Fernandez, "Deep learning for iris recognition: A survey," *ACM Computing Surveys*, vol. 56, no. 9, pp. 1–35, 2024. 4, 9

[28] Christopher D McMurrough, Vangelis Metsis, Dimitrios Kosmopoulos, Ilias Maglogiannis, and Fillia Makedon, "A dataset for point of gaze detection using head poses and eye images," *Journal on Multimodal User Interfaces*, vol. 7, pp. 207–215, 2013. 4, 5

[29] Marc Tonsen, Xucong Zhang, Yusuke Sugano, and Andreas Bulling, "Labelled pupils in the wild: a dataset for studying pupil detection in unconstrained environments," in *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications*, 2016, pp. 139–142. 4, 5

[30] Joohwan Kim et al., "Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation," in *Proceedings of the CHI conference on human factors in computing systems*, 2019, pp. 1–12. 4, 5

[31] Stephan J Garbin, Yiru Shen, Immo Schuetz, Robert Cavin, Gregory Hughes, and Sachin S Talathi, "Openeds: Open eye dataset," *arXiv preprint arXiv:1905.03702*, 2019. 4, 5, 9, 10

[32] Cristina Palmero, Abhishek Sharma, Karsten Behrendt, Kapil Krishnakumar, Oleg V Komogortsev, and Sachin S Talathi, "Openeds2020: Open eyes dataset," *arXiv preprint arXiv:2005.03876*, 2020. 4, 5

[33] Kevin W Bowyer and Mark J Burge, *Handbook of iris recognition*, Springer, 2016. 9, 10

[34] Yimin Yin, Siliang He, Renye Zhang, Hongli Chang, Xu Han, and Jinghua Zhang, "Deep learning for iris recognition: a review," *arXiv preprint arXiv:2303.08514*, 2023. 9

[35] Fernando Alonso-Fernandez and Josef Bigun, "A survey on periocular biometrics research," *Pattern Recognition Letters*, vol. 82, pp. 92–105, 2016. 9

[36] Man Zhang et al., "The first ICB* competition on iris recognition," in *IEEE International Joint Conference on Biometrics*. IEEE, 2014, pp. 1–6. 9

[37] Man Zhang, Qi Zhang, Zhenan Sun, Shujuan Zhou, and Nasir Uddin Ahmed, "The btas competition on mobile iris recognition," in *8th international conference on biometrics theory, applications and systems (BTAS)*. IEEE, 2016, pp. 1–7. 9

[38] Aidan Boyd, Zhaoyuan Fang, Adam Czajka, and Kevin W Bowyer, "Iris presentation attack detection: Where are we now?," *Pattern Recognition Letters*, vol. 138, pp. 483–489, 2020. 10

[39] Kiran B Raja, R Raghavendra, Jean-Noel Braun, and Christoph Busch, "Presentation attack detection using a generalizable statistical approach for periocular and iris systems," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2016, pp. 1–6. 10

[40] BO Adegoke, EO Omidiora, SA Falohun, and JA Ojo, "Iris segmentation: a survey," *International Journal of Modern Engineering Research (IJMER)*, vol. 3, no. 4, pp. 1885–1889, 2013. 10

[41] Peter Rot, Matej Vitek, Klemen Grm, Žiga Emeršič, Peter Peer, and Vitomir Štruc, "Deep sclera segmentation and recognition," *Handbook of vascular biometrics*, pp. 395–432, 2020. 10

[42] Aayush K Chaudhary, Rakshit Kothari, Manoj Acharya, Shusil Dangi, Nitinraj Nair, Reynold Bailey, Christopher Kanan, Gabriel Diaz, and Jeff B Pelz, "Ritnet: Real-time semantic segmentation of the eye for gaze tracking," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3698–3702. 10

[43] Jonathan Perry and Amanda Fernandez, "Minenet: A dilated cnn for semantic segmentation of eye features," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 0–0. 10

[44] Andres Valenzuela, Claudia Arellano, and Juan E Tapia, "Towards an efficient segmentation algorithm for near-infrared eyes images," *IEEE Access*, vol. 8, pp. 171598–171607, 2020. 10

[45] Abhijit Das et al., "Sclera segmentation and joint recognition benchmarking competition: Ssrbc 2023," in *IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2023, pp. 1–10. 10