Eye-tracked Virtual Reality: A Comprehensive Survey on Methods and Privacy Challenges

Efe Bozkir, Süleyman Özdel*, Mengdi Wang*, Brendan David-John, Hong Gao, Kevin Butler, Eakta Jain, and Enkelejda Kasneci

***Abstract*—Latest developments in computer hardware, sensor technologies, and artificial intelligence can make virtual reality (VR) and virtual spaces an important part of human everyday life. Eye tracking offers not only a hands-free way of interaction but also the possibility of a deeper understanding of human visual attention and cognitive processes in VR. Despite these possibilities, eye-tracking data also reveal privacy-sensitive attributes of users when it is combined with the information about the presented stimulus. To address these possibilities and potential privacy issues, in this survey, we first cover major works in eye tracking, VR, and privacy areas between the years 2012 and 2022. While eye tracking in the VR part covers the complete pipeline of eye-tracking methodology from pupil detection and gaze estimation to offline use and analyses, as for privacy and security, we focus on eye-based authentication as well as computational methods to preserve the privacy of individuals and their eye-tracking data in VR. Later, taking all into consideration, we draw three main directions for the research community by mainly focusing on privacy challenges. In summary, this survey provides an extensive literature review of the utmost possibilities with eye tracking in VR and the privacy implications of those possibilities.**

***Index Terms*—virtual reality, eye tracking, privacy, security, survey, literature review.**

I. Introduction

Over the last decade, virtual and augmented reality (VR/AR) communities have benefited from developments in computer hardware, graphics, and imaging science. To date, some modern head-mounted displays (HMDs) have already become available for reasonable prices for everyday usage. Furthermore, eye-tracking sensors have become either directly integrated into these HMDs (e.g., HTC Vive Pro Eye[1]) or are available as low-cost add-ons (e.g., Pupil Labs' eye-tracker add-on [1]). Since eye movements are related to human cognition, visual attention, and perception, the information on where users look at a certain point in time can be used to help users in various ways such as providing them with adaptive support during their tasks in VR, increasing their engagement, or improving the usability of the VR applications. At the same time, in the area of VR and HMDs, there is also a lot of room for improvement such as enhancing display resolutions, increasing realism, or preventing cybersickness, and gaze information can potentially be utilized to mitigate some of these technical issues.

It was foreseen by NVIDIA in 2017 that VR was approximately 20 years away from resolutions similar to human eyes [2]. With such technical developments and a wider range of usage, it is likely that future HMDs may become mobile devices like today's mobile phones, tablets, or smartwatches. However, as these devices are situated on the heads of users and have close proximity to the eyes of the users, apart from user-supporting functionalities, they might be perceived as privacy-invasive by users due to the fact that a lot of sensitive attributes about users such as genders, sexual preferences, and personal identities, can be extracted using eye movements [3]. From a privacy perspective, this information should be protected or the user should be provided with control over its release. At the same time, a pleasant virtual experience should be available to all users. In addition to possibilities with eye movements, eye information can also be utilized for authentication purposes with very high accuracy, especially with iris textures as they are treated as visual fingerprints. Therefore, in terms of privacy, there is an essential utility-privacy trade-off that should be taken into account along with user preferences and concerns.

As VR and privacy communities have been working to tackle these issues and with the availability of HMDs to wider communities especially during recent years, the number of works in the intersection of VR, eye tracking, and privacy has been increasing. However, until now, none of the previous works focused on comprehensive coverage of eye tracking in VR and its privacy implications. In this work, **(1)** we first cover the major research between 2012 and 2022 by first discussing how eye movements are extracted and the possibilities of using eye movements in immersive VR. Then, **(2)** we discuss security and privacy implications including eye-based authentication and privacy-preserving eye tracking. Lastly, **(3)** we draw and discuss three different directions for future research especially considering the privacy aspects of immersive eye-tracked VR.

A. Paper Structure and A Guide on How to Read This Paper

As we comprehensively survey research conducted in eye tracking, VR, and privacy domains, and due to the interdisciplinary and disparate nature of each domain, we present the

* indicates equal contribution.

E. Bozkir, S. Özdel, M. Wang, H. Gao, and E. Kasneci are with Human-Centered Technologies for Learning of Technical University of Munich, Germany (E-mails: {efe.bozkir, ozdelsuleyman, mengdi.wang, hong.gao, enkelejda.kasneci}@tum.de).

E. Bozkir is also with Human-Computer Interaction of University of Tübingen, Germany (E-mail: efe.bozkir@uni-tuebingen.de).

B. David-John is with Virginia Tech, USA (E-mail: bmdj@vt.edu).

K. Butler and E. Jain are with the University of Florida, USA (E-mails: butler@ufl.edu, ejain@cise.ufl.edu).

[1]https://vr-compare.com/headset/htcviveproeye, Last access 05/03/2023

structure of this paper and guidelines on how people with different backgrounds can benefit from it. We first discuss existing works in the literature in Section II. We then discuss the most important academic venues where research in these domains is published and how we filtered relevant papers from each venue to include them in our survey in Section III. These sections are appropriate for all readers especially to understand existing surveys and our methodology to form our own work.

Section IV discusses a very wide range of topics from computer vision-based approaches to track the eyes and further process eye region data, to eye-based human-computer interaction and understanding human visual attention, cognition, and perception mostly in an offline fashion. Section IV-A will benefit most readers interested in computer vision and machine learning to conduct research on how to track eyes and make eye-tracking solutions available for others. Section IV-B focuses on human-computer interaction by utilizing eye movements for different purposes such as real-time interaction in VR, foveated rendering, and how to deal with technical issues to improve immersive experiences. This section is more suitable for researchers and practitioners who work on the technical aspects of human-computer interaction and eye-tracked VR. Furthermore, Section IV-C combines eye tracking and VR from human visual attention, cognition, and perception point of view and is more suitable for researchers who have cognitive and experimental psychology backgrounds or are interested in these fields. In Section V, considering all the works and insights provided in Section IV, we provide works that revolve around the privacy and security implications of eye-tracked VR including authentication possibilities using eye data, why privacy-preserving methods are needed for this domain, and how current literature has addressed the privacy issues in a computational way. This section is relevant both for privacy and security researchers and for human-computer interaction researchers who are interested in privacy and security aspects of eye-tracked VR. Lastly, despite the fact that we provide future directions and discussions mainly on privacy implications of eye-tracked VR in Section VI, we note that this section is very relevant for all the researchers and practitioners whose work includes VR as privacy issues should be handled carefully regardless of any research area or application domain. In Section VII, we conclude our paper.

II. Related work

There are several works that systematically analyze relevant research for eye tracking, VR/AR, privacy, and security. Despite this, none of the works review eye tracking in VR setups in a comprehensive and systematic way, the privacy issues that such works could lead to for the end users, and how to mitigate such risks. Previously, Lappi [4] analyzed eye tracking in the wild by discussing the advantages, disadvantages, and technical terms to achieve valid and high-quality results from human eye-related experiments. Duchowski [5] discussed gaze applications and gaze-based interactions in graphical systems by focusing on topics like eye movement analytics, foveated rendering, and visual attention in a SIGGRAPH course. In another work, Plopsky et al. [6] covered gaze-based interaction and eye tracking in head-worn extended reality. Furthermore, Silva et al. [7] discussed the foundations of eye-tracking support for visual analytics systems, relevant applications, and challenges by considering five different themes including privacy protection as an important and challenging issue to be solved as even disorders could be detected with high accuracies, whereas Ens et al. [8] analyzed the challenges in immersive analytics by considering virtual/mixed/augmented realities and drew future directions. Merino et al. [9] presented a systematic literature review on mixed and augmented reality also discussing the possibilities with eye-tracking data. As our focus is rather on VR setups in this work, for mixed and augmented reality solutions, we refer the reader to the previous work in this area [10], [11], [12], [13], [14]. Clay et al.'s work [15] is related to VR-based eye tracking and they examined eye-tracking research in VR along with a pilot user study by focusing on experimental setups, common problems of virtual environments such as vergence-accommodation conflict [16], [17], eye-tracking calibration details, and visual region of interests. However, their study is relatively small scale as they did not study the works in eye tracking and VR in depth. More recently, Adhanom et al. [18] provided a broad overview of eye-tracking applications in VR and their challenges. Similar to Silva et al.'s work [7], they also identified privacy and security as important discussion points. Gressel et al. [19] also provided a brief overview of the privacy aspects of eye tracking and provided recommendations for privacy-aware eye tracking. Furthermore, overviews on considerations for high-quality data collection [20] and measurement of data quality of HMD-based eye tracking based on accuracy and precision [21] were also studied in the literature, which is important when privacy is considered because as when privacy-preserving solutions are introduced, a privacy-utility trade-off is often regarded and assessed.

To improve data analytics, quality, and evaluations, while useful information about users can be extracted from their visual scanning patterns and eye-tracking data, such information also treats users' privacy, which was considered by a few. Liebling and Preibusch [3] discussed that privacy loss may not be very understandable by the end-users and argued that gaze and pupillometry data should be protected carefully. Kröger et al. [22] pointed out the attributes that can be inferred using gaze data similar to Liebling and Preibusch's work [3] and discussed the privacy implications and possible societal impacts of those. Lebeck et al. [23] considered the security and privacy aspects of multi-user AR. The authors found that some of the users are concerned about the powerful abilities of the eye-tracking enabled interfaces and behavioral tracking such as an AR HMD understanding that the user is being attracted to someone due to not being able to stop looking at them, which might also apply for VR. This statement partly overlaps with the findings of more recent work by Steil et al. [24] that users agree to share their eye-tracking data if the co-owner of the data is a governmental health agency or the purpose is research. Roth et al. [25] also considered privacy and security issues as one of the important discussion points for social augmentations in user-embodied VR. Furthermore, Katsini et al. [26] provided an extensive survey on gaze-based authentication and like other previous works [18], [7], further research was encouraged for privacy-preserving eye tracking.

Despite several recent attempts, none of the works in the literature focused on systematically analyzing the research that studies eye-tracking data analytics in immersive VR setups considering the full processing pipeline of eye-tracking methods from pupil detection and gaze estimation to human visual attention and cognition, privacy-preserving manipulations of eye movements in such setups, and further novel research directions that should be pursued to help VR technology to be used everyday life altogether. In our survey, we focus on these aspects and provide a very comprehensive and systematic overview for the research community.

III. Methodology

As a starting point, we analyzed the relevant surveys and the papers that study eye tracking, VR, AR, human-computer interaction, and privacy-preserving eye tracking [4], [7], [8], [6], [9], [3], [22], [26]. Following, we considered the papers between 2012-2022 since the VR devices and eye trackers have been becoming prevalent in the daily life lately. We used Google Scholar to query papers from 33 renowned venues including ACM CHI, IEEE VR, ACM ETRA, IEEE ISMAR, ACM IUI, ACM UIST, ACM ICMI, ACM Mobile-HCI, ACM MM, ACM VRST, IEEE AIVR, AAAI Conference on Artificial Intelligence, IEEE CVPR, IEEE ICCV, ECCV, NeurIPS/NIPS, ACM Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), Journal of Eye Movement Research (JEMR), IEEE Transactions on Visualization and Computer Graphics (TVCG), ACM Transactions on Graphics (ToG), PoPETS/PETS, USENIX Security, USENIX SOUPS, NDSS, IEEE Symposium on S&P, IEEE S&P Magazine, ACM SIGSAC Conference on Computer and Communications Security (CCS), Journal of Privacy and Confidentiality, IEEE Transactions on Information Forensics and Security (TIFS), IEEE Transactions on Dependable and Secure Computing, Computers & Security, ACM Transactions on Privacy and Security, and ACM Transactions on Cyber-Physical Systems. For the conference proceedings, we only considered the main conference proceedings, leaving companion and workshop proceedings aside, apart from the works that present concrete results or that are highly related to eye tracking and VR. We cross-checked the query results from the DBLP.

In order to filter the relevant papers, we used ("eye tracking" AND "virtual reality") as the main query where each phrase could appear anywhere in the article. In addition, to further track privacy- and biometrics-related papers, we used the following sub-queries: ("eye tracking" AND "virtual reality" AND "privacy") and ("eye tracking" AND "virtual reality" AND "biometrics"). We also applied forward and backward tracking by checking the citations and references of the queried papers, respectively. The main query yielded 1327 query results, whereas two sub-queries led to 212 and 95 results, respectively. Figure 1 shows the resulting distributions of the main query according to venues. If the aforementioned venues do not appear in Figure 1, it means that no paper was found at those venues with our main query.

IV. Eye-tracked Virtual Reality

Eye tracking can be considered one of the key technologies, especially for VR setups, as it is possible to track eyes more accurately within HMDs compared to remote setups due to the reduced distance between eye trackers and eyes in HMDs. In addition, as eyes do not move completely voluntarily, it is likely that one may get more objective measurements for user behaviors compared to self-reported data such as questionnaires, which are often used in psychology-related studies in VR. While this rich source of information could be utilized in various manners, there are different technical aspects. Firstly, in order to use this data, eye regions and gaze directions are extracted. These tasks are mostly carried out by using different computer vision and machine learning techniques. Later, especially estimated gaze directions are utilized for interaction purposes, for instance, to support users in a variety of ways, such as by predicting user intents and providing them with context-sensitive aid or by purely technical ways such as foveated rendering. While such cases should work in real-time in practice, collected data from different experimental setups can be used ad-hoc to understand humans and how they behave in various application domains such as education or medicine. These application domains indeed show that VR can also be considered as a research tool with its provided immersion to create scenarios where it is almost impossible to do in the real world due to privacy and safety issues or simply because of practical infeasibility.

Considering these, we organize this section as follows. We first discuss major works that utilize computer vision and machine learning to estimate gaze and to detect eye regions as these are the initial steps of all, also including the methods for eye movement event detection in VR, which differs from the conventional settings. Then, we provide a detailed overview of eye-based interaction and its different applications. As the last step, we cluster the works that conduct offline data analyses particularly to understand human visual attention, cognition, and perception and provide an overarching view of these works. The structure of this section is given in Figure 2.

A. Hardware, Datasets, and Algorithms

In this section, we first lead insights into the works about hardware and datasets that are relevant to eye-tracked VR in Sections IV-A1 and IV-A2, respectively. Then in Sections IV-A3, IV-A4, and IV-A5, we proceed to discuss eye-tracking-related algorithms for eye region segmentation, gaze estimation, and eye event classification, respectively. A brief summary of the papers discussed in this section is given in Table I. It should be noticed that eye tracking in immersive VR using HMDs is essentially different from other real-world scenarios as users mostly wear HMDs on their heads, and sensors only see the eye regions rather than the whole face. Therefore, if not explicitly mentioned, we reduce the focus to such setups in this work. For an extended evaluation of eye tracking, such as gaze estimation models regardless of immersive VR or real-world scenarios, we refer the reader to the survey papers by Cheng et al. [27] for appearance-based gaze estimation with deep learning, by Akinyelu and

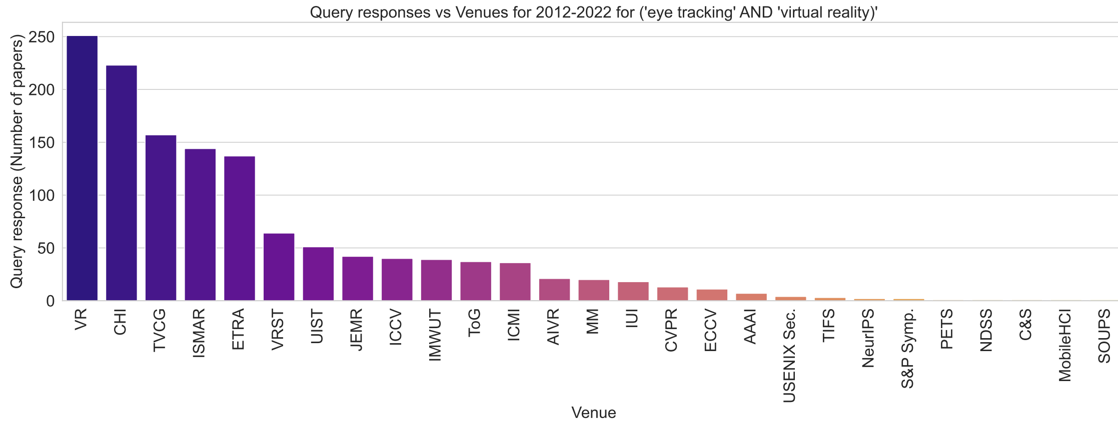


Fig. 1. Query responses according to venues for the query ('eye tracking' AND 'virtual reality') between 2012-2022.

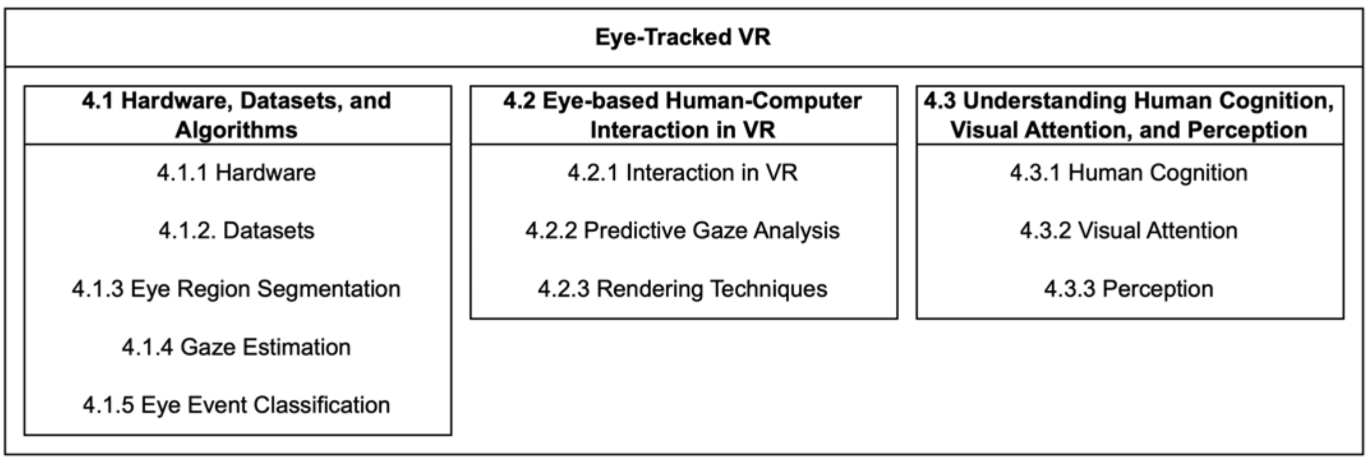


Fig. 2. Organization of Section IV.

Blignaut [28] for convolutional neural network (CNN)-based gaze estimation, and by Kar and Corcoran [29] for gaze estimation in consumer platforms.

1) Hardware: We begin by examining the hardware necessary for implementing eye-tracked VR. Formerly, the majority of VR headsets like Oculus Rift[2] and HTC Vive[3] did not deliver built-in eye tracking functionality by themselves. To approach eye tracking on these devices, add-on eye-tracking extensions like the ones provided by SMI[4], Pupil Labs[5], and Tobii[6] were needed. As eye tracking and VR both grow at an astounding rate, an increasing amount of HMDs like HTC Vive Pro Eye[7], Fove-0[8] and Varjo XR-3 support eye tracking on their own. An example of an HMD device, the Varjo XR-3, which comes with an integrated eye tracker, is shown in Figure 3, along with eye images and a part of a sample raw data. In the meanwhile, a growing body of research highlights the importance of evaluating the performance of built-in and add-on eye trackers for HMD devices. For instance, Lohr et al. [32] measured the quality of eye data acquired with the SMI add-on eye tracker in the HTC Vive from multiple facets, including spatial accuracy, spatial precision, temporal precision, linearity, and crosstalk. Using the same metrics, Aziz and Komogortsev [36] assessed the eye-tracking performance of the Microsoft HoloLens 2[9]. The quality of eye-tracking data recorded by the SMI extension has also been analyzed by Roth et al. [33] in combination with Oculus Rift DK2[10] regarding tracking precision and fixation accuracy in the context of foveated rendering. Adhanom et al. [30] reported an average eye-tracking accuracy of 1.23° and a root mean square (RMS) precision of 0.62° on the HTC Vive Pro Eye with their newly developed Unity package called GazeMetrics. In a recent work

[2]https://vr-compare.com/headset/oculusrift, discontinued, Last access 01/19/2023.

[3]https://vr-compare.com/headset/htcvive, discontinued, Last access 01/19/2023.

[4]SensoMotoric Instruments, acquired by Apple.

[5]https://pupil-labs.com/, Last access 01/19/2023.

[6]https://www.tobii.com/, Last access 01/19/2023.

[7]https://www.vive.com/us/product/vive-pro-eye/overview/, Last access 01/19/2023.

[8]https://fove-inc.com/product/fove0/, Last access 01/19/2023.

[8]https://varjo.com/products/xr-3/, Last access 01/31/2023.

[9]https://www.microsoft.com/en-us/hololens, AR HMD, Last access 01/19/2023.

[10]https://vr-compare.com/headset/oculusriftdk2, discontinued, Last access 02/05/2023.

TABLE I

A NON-EXHAUSTIVE OVERVIEW OF PAPERS DISCUSSED IN SECTION IV-A.

Contribution	Characteristics
Eye-tracker performance assessment	[30], [31] HTC Vive Pro Eye [32], [33] SMI add-on eye tracker + HTC Vive / Oculus Rift DK2 [34], [35] HTC Vive Pro Eye, Varjo VR-1, Fove-0 [36] Microsoft HoloLens 2 (AR)
Low-cost eye trackers	[37] Using dichroic mirrors and personalizable lens [38] Using phone selfie camera + Google Cardboard [39] EyeSpyVR: Using phone selfie camera + VR Box headset [40] EyeMR: Using USB camera + IR-LED + Cardboard
Non-VOG eye trackers	[41], [42] Photosensor-oculography (PSOG) [43], [44] Electro-oculography (EOG) [45] Scleral search coil (SSC)
Datasets	[46], [47] OpenEDS: 152 users, for eye region segmentation [48] NVGaze: 35 users, synthetic + real data [49] IQVA: 14 users, question-driven visual attention [50] 13 users, automatic eye event classification [51] TEyeD: 54 users, largest unified eye dataset [52] SynchronEyes: 15 users, stationary + HMD eye tracking [53] 100 users, quantitative taxonomy for videos, multi-modal [54] OpenNEEDS: 44 users, multi-modal [55] VREED: 34 users, for emotion recognition, multi-modal [56] EgoBody: 2 users, for human body reconstruction [57] HE-Gaze: 15 users, AR dataset, multi-modal [58] 32 users, real-time emotion annotation
Eye region segmentation	[59] EyeNet: Using residual blocks and convolutional attention [60] Using CycleGANs for segmentation, refinement, and generation [61] EllSeg: Robust against occlusions [62] Semi and unsupervised domain adaptation [63] Eye-MMS: Using multi-scale inter-connected CNN, lightweight [64] EyeSeg: Using generalized dice loss function, lightweight [65] RITnet: Using U-Net + DenseNet, lightweight
Generating eye image from segmentation	[66] D-ID-Net: Two-phase image generation [67] Seg2Eye: content from segmentation + style from person [68] GeoMaskGAN: Maintaining geometric consistency
Gaze estimation	[69] Using changes in pixel brightness, lightweight [70] Using eye motion event, extremely high frame rate [71] Neural3DGaze: 3D pupil localization [72] Robust against head pose changes [73] Using unsupervised representation + gaze redirection [74] Using CNN-recurrent model for temporal gaze trace [75] ARE-Net: Asymmetric regression of eyes [76] Using depth information
Calibration	[77] Using correlation between fixation and hand interaction
Eye event classification	[78] Data/algorithm translation between 2D-monitor and HMD [50] Rule-based classifier [79] Estimation of remaining time until next saccade

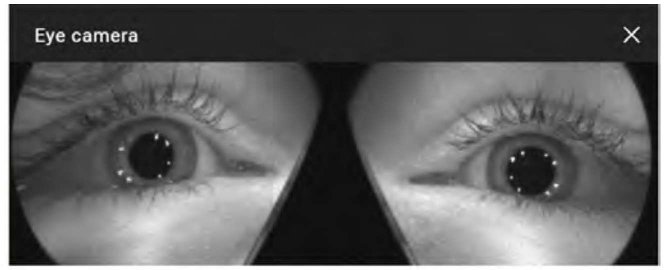
by Schuetz and Fiehler [31], an average accuracy of 1.08° and mean standard deviation (SD)/RMS precisions of 0.36°/0.2° were reported on the same device after outlier correction. The authors also found a significant decrease in accuracy and precision when participants wore vision correction glasses, while the effect of contact lenses was more elusive. Stein et al. [35] conducted a comparison between Fove-0, Varjo VR-1[11], and HTC Vive Pro Eye using another two metrics: eye-tracking delay and latency. With a delay of 15-52 ms and a latency of 45-81 ms, the Fove-0 outperformed the other two HMDs to a large extent. The assessment of eye-tracked HMD can also go beyond statistical performance. For instance, Maraj et al. [34] compared the HTC Vive Pro Eye with the Varjo VR-1 in three eye-tracking scenarios. The comparison results were reported in non-parametric forms that covered immersion, simulation sickness, and visual discomfort.

a) Low-cost HMD Eye Trackers: While add-on eye-tracking extensions and built-in tracking modules have become more commercially affordable, they can still be expensive due to additional hardware and software requirements. To address this issue, Stengel et al. [37] proposed an implementation utilizing dichroic mirrors and a personalizable lens positioning system at an estimated cost of merely 450$. The authors deployed dichroic mirrors to support eye-tracking cameras outside of the field of view (FOV), whereas the lens locating module was applied to account for different interocular distances. With a model-based gaze estimation algorithm, the system achieved a gaze angle error of 0.5°-3.5°. An alternative, more affordable system was suggested by Greenwald et al. [38], which uses a smartphone and an inexpensive Google Cardboard[14] that costs approximately 15$. In this system, gaze estimation is approached through the capture of Purkinje

[11] https://vr-compare.com/headset/varjovr-1, discontinued, Last access 01/19/2023.

[13] https://varjo.com/products/xr-3/, Last access 01/31/2023.

[14] https://arvr.google.com/cardboard/, Last access 01/26/2023.



gaze_forward_x	gaze_forward_y	gaze_forward_z	gaze_origin_x	...	right_projected_y
0.020967	-0.249046	0.968265	0.0	...	-0.127911
0.023332	-0.256099	0.966369	0.0	...	-0.130731
0.023825	-0.267165	0.963356	0.0	...	-0.134594
0.025230	-0.269331	0.962717	0.0	...	-0.136075
0.022590	-0.269124	0.962841	0.0	...	-0.130524
...	...	...	...	...	...
0.109950	-0.248506	0.962370	0.0	...	-0.101884
0.109999	-0.246964	0.962761	0.0	...	-0.099792
0.106294	-0.247206	0.963115	0.0	...	-0.100414
0.107338	-0.245482	0.963440	0.0	...	-0.100361
0.108383	-0.245633	0.963285	0.0	...	-0.100110

Fig. 3. Left: Varjo XR-3[13], which is a high-end HMD device for both VR and AR. Upper right: eye images taken by cameras inside Varjo XR-3. Lower right: some eye features recorded by Varjo XR-3.

images (i.e., reflections of on-screen images on the eyes) using the smartphone's front-facing camera, and an average error of 5.0° can be achieved. In parallel to them, Ahuja et al. [39] presented a similar framework named EyeSpyVR, which also uses the smartphone selfie camera and a 10$ VR Box headset to achieve coarse gaze tracking. Having tested the framework with 70 subjects, the researchers reported an average gaze estimation error of 10.8° and 12.9°, respectively for the use case with and without calibration. Another system deployed on Google Cardboard called EyeMR was developed by Stratmann et al [40]. In contrast to the aforementioned systems that are realized using the smartphone selfie camera, EyeMR uses USB cameras and extended IR-LED circuit boards to function gaze estimation and supports both mono- and binocular eye tracking.

b) Non-Video-oculography (VOG) HMD Eye Trackers: Common camera-based eye-tracking systems fall into the category of video-oculography (VOG). VOG-based methods often struggle to find an optimal balance among factors like power consumption, computational cost, latency, and accuracy. There also exist other types of eye trackers, which can be categorized into photosensor-oculography (PSOG), electro-oculography (EOG), and scleral search coil (SSC) according to Duchowski et al. [80] and Rigas et al. [81]. PSOG is akin to VOG in the sense that they both often require light sources and optical sensors to capture reflections, whereas PSOG usually uses a sparse grid of photodiodes/photosensors and records significantly fewer pixel values compared to VOG. As a result, PSOG can outperform VOG on the sampling rate, computational cost, and power consumption. The works by Li et al. [41] and by Katrychuk et al. [42] fall into this category. In comparison to VOG and PSOG, EOG-based methods such as by Shimizu and Chernyshov [43] and by Bernal et al. [44], as well as SSC-based methods like by Whitmire et al. [45], commonly put no requirement on light emitters or receivers. The former approach eye tracking by placing electrodes around the eyes and then measuring voltage changes during eye movement, whereas the latter realize gaze estimation by requiring users to wear wired contact lenses and then tracking the lenses in a magnetic field. As stated by Adhanom et al. [18], EOG-based methods can be easily implemented in HMDs and are capable of functioning eye tracking even when eyes are closed though being less accurate, while SSC-based methods provide excellent accuracy but are often hard to deploy. Other than the aforementioned methods, eye tracking may also be approached with quite special devices and algorithms. One example would be the work by Shenoy et al. [82] where the authors achieved gaze estimation using adaptive optics scanning laser ophthalmoscope (AOSLO), which is a device that can image retina at high resolution and high frame rate. The authors modeled eye tracking as a joint estimation of retina motion and appearance similar to simultaneous localization and mapping (SLAM [83]) and achieved an accuracy of below $\frac{1}{60^\circ}$ at a sampling rate of 1 kHz.

2) Datasets: As data quality, sample size, and context-dependent information are very important for the performance and robustness of the gaze estimators, Kim et al. [48] proposed two datasets satisfying such criteria, including a synthetic one using anatomically informed eye and face models and a real-world dataset from 35 participants for near-eye gaze estimation. The authors showed that their trained CNNs perform approximately with accuracy losses of 2° and 0.5° for person-independent and personalized setups, respectively. Garbin et al. [46], [47] introduced a well-known dataset called OpenEDS that can be used for not only gaze estimation but also eye region segmentation. The dataset was collected from 152 participants in an immersive virtual environment and covers diverse data types, including pixel-level annotations of eye regions, unlabeled eye images, video sequences, and point clouds. In most cases, visual attention is raised in a bottom-up pattern driven by visual stimuli during data collection. Jiang et al. [49] presented a novel dataset called IQVA (i.e., immersive question-directed visual attention) in which visual attention was driven by tasks in a top-down style. The dataset consists of eye-tracking data from 975 HMD video clips with questions, with each being annotated by 14 participants. The authors also incorporated a neural network to predict the correctness of attention maps. Agtzidis et al. [50] published an eye-tracking dataset for 360° videos collected from 13 observers together with a novel two-stage eye event annotation pipeline and a rule-based eye event classifier for automatic event labeling. Fuhl et al. [51] introduced the largest ever unified public eye dataset, which was gathered using seven different head-mounted devices in the real world, VR (54 participants), and AR. The dataset covers various types of eye data, such as 2D and 3D landmarks, eye segmentation information, and 3D eyeball annotation.

More recently, Aziz et al. [52] published a new dataset called SynchronEyes, in which eye movements of 15 participants were recorded simultaneously with both a stationary eye tracker EyeLink 1000[15] and a wearable glass eye tracker AdHawk MindLink[16]. Jin et al. [53] presented a new quantitative taxonomy for VR videos that rely on three metrics: camera motion, video quality, and dispersion of region of interest (ROI). Based on this taxonomy, the researchers created a dataset of both head and gaze behavior from 100 participants and 27 videos. There also exist other multi-modal datasets available. An example would be the OpenNEEDS dataset by Emery et al. [54], which was collected from 44 participants exploring two virtual spaces and it includes data of not only eyes but also head, hands, and scenes. Tabbaa et al. [55] introduced a dataset called VREED for emotion recognition where data from different modalities, including eye movement, electrocardiogram (ECG), and galvanic skin response (GSR) were recorded from 34 subjects while experiencing 360° videos. Recently, Zhang et al. [56] published a new multi-modal dataset called EgoBody, where data including eye gaze, head motion, and gestures were collected while an emphasis on the work is human body reconstruction from an egocentric view in VR. Chen et al. [57] also presented a multi-modal dataset called HE-Gaze recently, which was captured from 15 participants using an AR HMD device Nreal Light[17]. In addition to eye movement data, the dataset also contains head movement data and near-eye images for both eyes. Along with the dataset, the authors also introduced an appearance-based gaze estimation algorithm called HE-Tracker that utilizes head movement and is capable of approaching eye tracking in real-time at 48 Hz with an accuracy of 3.655°.

a) Synthetic Data: Collecting high-quality eye-gaze data can be expensive in terms of both time and cost. This has led to the development of synthetic 3D models such as UT Multi-View [84], SynthesEyes [85], and UnityEyes [86], as well as generative algorithms like [87] and data augmentation approach like gaze redirection [88]. While few of these works are specifically tailored for the use of HMDs, some concepts can be adapted for eye data generation in immersive virtual environments.

b) Data Annotation: An essential step during and after a data collection process is data annotation, which is often costly. Xue et al. [58] addressed the issue of collecting emotion annotations along with eye-tracking data in VR environments. Typically, emotional feedback during the VR experience is annotated retrospectively. As a consequence, emotion labeling can be discrete and time-consuming. By leveraging their previously introduced peripheral visualization frameworks HaloLight and DotSize [89], the authors were able to achieve real-time continuous emotion annotation during data collection from 32 participants and provided temporally precise labels for downstream tasks.

3) Eye Region Segmentation: Having broadly viewed hardware and datasets for eye-tracked VR, we now proceed to provide an overview of algorithms related to gaze estimation on HMDs, beginning with the algorithms for eye region segmentation. Semantic segmentation of the eye region is crucial for gaze estimation and is often the starting point for many gaze estimation algorithms. According to Shen et al. [62], early eye region segmentation algorithms like [90], [91] were mainly driven by iris texture and sclera extraction. Nowadays, an increasing shift towards multi-class (multi-region) eye segmentation that is typically realized through deep learning has been experienced. Unlike other semantic segmentation tasks, eye region segmentation is often challenged by images being low resolution, blurring, off-axis, etc. Kansal and Nathan [59] presented an efficient encoder-decoder network called EyeNet that addresses these issues by using residual blocks [92] in both encoder and decoder to improve gradient flow and using convolutional block attention modules (CBAM [93]) to enhance boundary sharpness and accuracy. The network accomplishes a total score of 0.974 on the EDS evaluation metric. Fuhl et al. [60] explored the applicability of CycleGANs [94] for eye segmentation and suggested three different generative adversarial networks (GANs) for segmentation, image refinement, and image generation, respectively. In particular, these models were trained with cyclic loss to prevent discriminator overfitting. Conventionally, eye segmentation is vulnerable to occlusion caused by eyelids and eyelashes. Kothari et al. [61] addressed this challenge with a segmentation framework called EllSeg that is robust against occlusions and can be implemented jointly with other pupil and iris ellipse segmentation methods in an encoder-decoder pattern.

[15] https://www.sr-research.com/eyelink-1000-plus/, discontinued, Last access 01/25/2023.

[16] https://www.adhawkmicrosystems.com/adhawk-mindlink, Last access 01/25/2023.

[17] https://www.nreal.ai/light/, AR HMD, Last access 01/23/2023.

a) Lightweight Eye Segmentation: Computational resources, such as computational power, memory, and storage, are often bottlenecks of HMD devices. Boutros et al. [63] alleviated the computation burden on embedded systems with their miniature multi-scaled segmentation network (Eye-MMS), which is based on multi-scale interconnected CNN. The authors achieved a 3% loss in accuracy compared to the original model while successfully reducing the number of model parameters from 6574k to 80k. Perry and Fernandez [64] proposed a lightweight encoder-decoder segmentation network named EyeSeg, which adopts a customized loss function combining categorical cross entropy and generalized dice loss function (GDL [95]) to tackle the shortage of labeled data. EyeSeg contains merely 190k parameters while yielding a 94.5% mean intersection over union (mIOU) on the 2020 OpenEDS dataset. Another lightweight segmentation network called RITnet that combines U-Net [96] and DenseNet [97] was developed by Chaudhary et al. [65]. The network has a size of only 0.98 MB and allows eye segmentation at 300 Hz in real-time with an mIOU of 95.3% on the OpenEDS dataset.

b) Dealing with Data Shortage: Another major challenge in eye region segmentation is the lack of high-quality eye images. Damer et al. [66] overcame this challenge by generating eye images from semantic segmentation with their novel model called D-ID-NET. The pipeline of this model consists of two phases: in the first phase, a domain network (D-Net) synthesizes identity-irrelevant images from semantic labels, and in the second phase, identity-specific information is introduced into the images by an identity-specific network (ID-Net). Both networks are CNNs with the same structure but different training strategies. A similar concept was implemented by Buehler et al. [67] with their model Seg2Eye, which generates content-preserving eye images from semantic segmentation. The work resembles style transfer [98] in the sense that semantic segmentation defines the content of generated images while their styles are controlled by style features extracted from images of the target person. Shen et al. [62] extensively studied domain adaption for eye segmentation in the case when only a few source images are annotated while most data are not labeled. They systematically investigated the impact of annotated data by training the model in supervised, unsupervised, and semi-supervised manners respectively, and varying the amount of labeled eye images in the target domain during training. More recently, Lu et al. [68] proposed an image-to-image translation network called GeoMaskGAN that accounts for geometric consistency. The network consumes as input a pair of the eye image and eye segmentation mask and outputs a new pair while reducing the geometric gap between translated images and original ones.

4) Gaze Estimation: Gaze estimation is the core focus of most eye-tracking works. In this subsection, we do not necessarily differentiate between stationary eye tracking and HMD eye tracking, as most gaze estimation algorithms are not customized for the use of HMDs. A broad categorization of gaze estimation methods is provided in Figure 4.

a) Model-based Gaze Estimation: According to Hansen and Ji [101], gaze estimation algorithms can be broadly classified into two categories: model-based and appearance-based. Model-based methods are sometimes also referred to as feature-based. As their name implies, model-based techniques rely on local geometric eye features such as contours and eyeball models. Model-based methods often provide high accuracy, while a common disadvantage of them is that they often have a limited range of operation in accordance with Zhang et al. [102], which is presumably not a central concern for HMD devices since sensors are close enough to eyes. As surveyed by Zhang et al. [103], model-based algorithms can be further categorized into corneal reflection-based (glint-based) and shape-based (glint-free), depending on whether additional light sources are needed or not. Glint-based methods like [104], [105] often compute the location of the cornea center using the Purkinje reflection of infrared light, while a typical workflow of glint-free methods [106], [107] often begins with eye landmark detection and then geometric features are fitted to service gaze estimation in downstream.

Formerly, gaze estimation methods were often heavyweight and hence could not function with eye tracking at a high rate. In recent work, Feng et al. [69] addressed this issue with their novel model-based algorithm that is based on event-driven eye segmentation. The model tracks events (changes in pixel brightness level) to predict ROIs of near-eye images, whose resolutions reduce to 18-32% compared to original images. Then, eye region segmentation is carried out on the reduced eye images and gaze estimation is realized upon segmentation. The whole model is capable to operate at 30 Hz on a mobile device at an accuracy of 0.1°-0.5°. Angelopoulos et al. [70] also made use of motion events to promote the frame rate of gaze estimation by placing event cameras close to the eyes. While conventional model-based pupil detection algorithms are used to approach basic pupil tracking, the motion events are utilized to update pupil location at high frequency. On an event-based eye dataset, the whole system reaches an accuracy of 0.45°-1.75° at a frame rate of over 10 kHz. Another deficiency of most model-based algorithms is that they do not take pupil location in 3D space into consideration, which can in turn negatively influence eye-tracking accuracy. Lu et al. [71] proposed a solution to this with their 3D pupil localization model that utilizes an advanced anatomical eyeball model and accounts for the error caused by corneal reflection. With the proposed model, the authors achieved error reductions of 47.5% and 18.7%, respectively for 3D pupil localization and for gaze estimation in comparison to prior works on their newly collected dataset, which contains ground truth of relative 3D pupil locations.

b) Appearance-based Gaze Estimation: An increasing shift towards appearance-based methods has been experienced in recent years. In contrast to model-based techniques, appearance-based methods are conceptually and structurally simpler: they typically use machine learning techniques to directly learn gaze direction from photographic eye appearance (images). As surveyed by Akinyelu and Blignaut [108], prior appearance-based methods like [109], [110], [111] generally relied on basic machine learning models such as linear regression, support vector machine, and random forest. Nowadays, appearance-based methods are often approached through deep learning and CNNs. One example of advanced

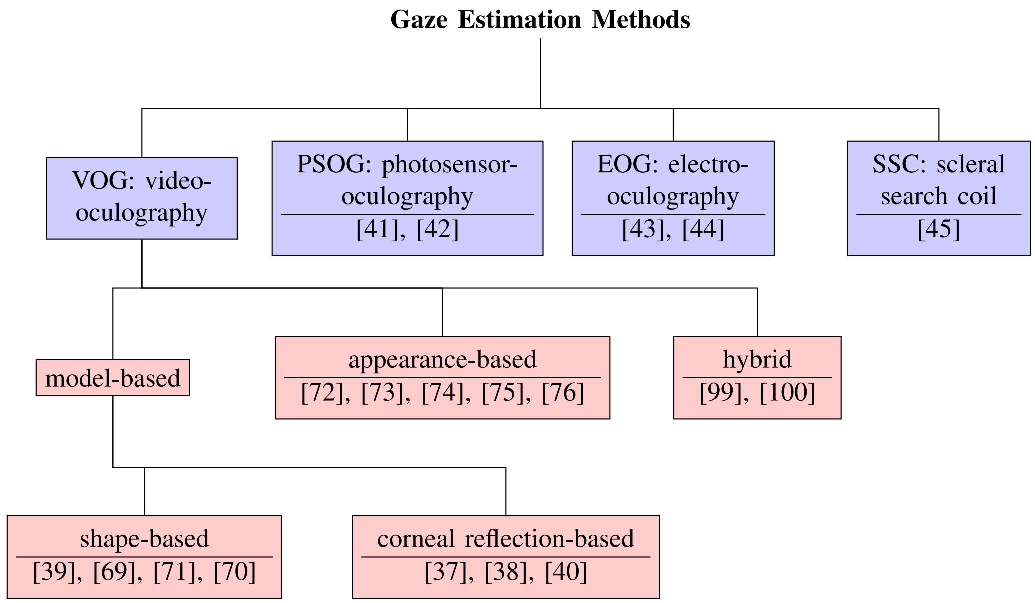


Fig. 4. A general taxonomy of gaze estimation methods that are discussed in Section IV-A.

appearance-based models is the work by Ranjan et al. [72], which improves the robustness against head pose by head post clustering. In contrast to former works in which one dedicated network is needed for each head pose, the authors designed a branched structure where the majority of layers are shared while only a few final dense layers are specified for each head pose. Moreover, they trained the model by transfer learning from the object viewpoint estimation CNN [112] that is intrinsically more related to gaze estimation than object detection. To overcome the problem of data shortage, Yu and Odobez [73] suggested an unsupervised representation for gaze estimation for which unlabelled eye images are used, and calibration works in a few-shot manner. The authors jointly trained a network for learning gaze representation and a network for gaze redirection while using warping field regularization to prevent overfitting and distortion. Different from prior works which learn embeddings in high dimensional space, this novel model captures 2D embeddings which are linked to clear physical meaning, namely eyeball yaw and pitch. With solely 100 calibration samples, the model was capable to reach an accuracy of 7°-8°.

Most appearance-based gaze estimation methods only utilize static eye images and often ignore temporal trace of gaze, albeit it contains important information. Palmero et al. [74] demonstrated the benefit of temporal gaze sequence by implementing a many-to-one CNN-recurrent model for gaze estimation. By bringing into temporal information, the researchers realized an error reduction of up to 19.78% on average, respectively 16.91% for the horizontal axis and 23% for the vertical axis. In another promising appearance-based model, Cheng et al. [75] introduced a new asymmetric regression-evaluation network (ARE-Net) that exploits the asymmetry of human eyes. The model consists of an asymmetric regression network (AR-Net) that estimates 3D gaze direction and evaluation networks (E-Net) which assess the performance of each eye and adjust the regression strategy accordingly. While a model considering a single eye can achieve an accuracy of 6.3° on the MPIIGaze [113] dataset, the proposed method decreased the error to 5.0°. Returning to the context of eye tracking in immersive virtual environments, a common problem is the HMD slippage, which can dramatically downgrade the gaze estimation accuracy of appearance-based models due to their sensitivity to camera placement. To mitigate this issue, Stojanov et al. [76] recently incorporated depth information in parallel to eye appearance while using two approaches to improve the robustness against fitment and slippage. While the first approach simply combines features from the eye image and depth map channel-wise, the second approach exploits the so-called transformer-based cross-modal attention (CMA) block [114] and increases the generalizability of the model to a large extent.

c) Hybrid Gaze Estimation: Besides appearance-based and model-based techniques, in the last years, gaze estimation techniques that combine both have also drawn attention. Such models are often called hybrid models. As the name implies, these methods like [99], [100] commonly use deep learning models to extract eye geometric features and then map the features to gaze position.

d) Calibration: A procedure that is closely related to gaze estimation is calibration, which serves as a prerequisite for an eye-tracking device to function properly. However, when deployed in HMDs, calibration can be disrupted by movements of the head and body. Therefore, it is often necessary to re-calibrate during usage. There has been some work dedicated to addressing this problem. For instance, to provide a more immersive and consistent user experience, Sidenmark and Lundstroem [77] analyzed the timing and probability of fixations by exploiting the correlation between gaze location, virtual objects, and hand interaction. Based on the analyzed gaze patterns, the authors suggested an implicit and consecutive re-calibration during the use of HMDs.

5) Eye Event Classification: For the sake of data visualization and further analysis (e.g., offline analyses in psychology experiments that utilize eye tracking), eye movement is often divided into a sequence of consecutive events of different types like fixations, saccades, and smooth pursuits. This procedure is often referred to as eye movement event classification, eye movement segmentation, or eye movement event detection. Broadly speaking, fixations are defined as stable eye movements when eyes are focused on a specific area or volume, whereas saccades are fast and ballistic switches between consecutive fixations. These two eye movements are often combined in the temporal dimension to create visual scanpaths. In contrast to fixations and saccades, there are some other more fine-grained eye movements such as smooth pursuits, which are defined as eye movements that are fixated on a moving target.

In accordance with Agtzidis and Dorr [78], traditional classification methods such as [115], [116], [117] which were designed for 2D monitor-based scenarios cannot be directly applied to eye movement event classification in VR HMDs due to the change in frame of reference. The authors suggested two approaches to alleviate such a problem: the first approach aims at the transplantation of 2D monitor-based classification algorithms to 3D Cartesian space, while the second approach projects HMD data onto 2D space such that traditional methods can be used without modification. Similar to other gaze-related tasks, eye event classification in VR suffers from data vacancy. In a subsequent study, Agtzidis et al. [50] tried to address this issue. The authors gave explicit definitions to primary and secondary eye events and then proposed a two-stage annotation pipeline. Furthermore, the authors developed a rule-based eye event classifier for automatic event labeling. Based on these, the authors published an annotated eye-tracking dataset for 360° videos collected from 13 observers. Recently, Rolff et al. [79] introduced a classification strategy that redefines eye event detection as a time-to-event problem. In comparison to traditional algorithms that divide eye movement into discrete events, the newly proposed model is capable of estimating the remaining time until the next saccade and predicting eye behavior in the near future.

B. Eye-based Human-Computer Interaction in VR

In the prior section, we provide a review of the latest advancements in eye-tracking hardware, datasets, and algorithms that form the algorithmic fundamentals of eye-tracked VR. These advancements have opened up numerous opportunities to enhance human-computer interaction in immersive virtual environments. In this section, we provide a comprehensive examination of research in the field of eye-tracking-based human-computer interaction within the context of VR. We begin with exploring recent works related to eye-based interaction, explicitly addressing the challenges and potential solutions in Section IV-B1. Subsequently, we delve into predictive gaze analyses in Section IV-B2 by encompassing prediction and visualization techniques. Lastly, in Section IV-B3, we give an overview of eye-based rendering techniques that improve the real-time performance of VR devices and enhance the user experience by mainly creating more realistic virtual environments.

1) Interaction in VR: Inherent characteristics of virtual environments provide users with novel opportunities for interacting with these environments and settings. Many interaction methods require additional controllers, such as handheld ones, while gaze-based interaction can solely be executed by using built-in or integrated eye trackers in a hands-free way. However, there are numerous challenges with regard to precision, time efficiency, and simplicity. In this section, we focus on works that revolve around eye-based interaction techniques in VR that address existing challenges in this field. An overview of all papers in this section is depicted in Table II.

a) User Interface Design: We first present a review focusing on the design of user interfaces based on eye tracking in VR. The main goal of these interfaces is to allow for robust interaction while minimizing the time and effort required by users. In the field of VR, several works have addressed this issue. For instance, Reiter et al. [118] introduced a hand-attached menu design that can be navigated with wrist rotation and interacted with the eyes, freeing up one hand for other tasks. In another work, Choi et al. [119] proposed a solution to the "Midas Touch" problem [158], a phenomenon that occurs when a user's gaze falsely interacts with an object or menu item during a search task. Choi et al. defined the gazing region between 25° and 45° as the "Kuiper Belt" which is not frequently targeted by the users during interaction due to being outside of the comfortable eye movement region [159], [160], but still within the physically reachable area [161], [162]. By placing menu items within the Kuiper Belt region, Choi et al. [119] showed the number of false inputs during the search task significantly reduced, and the Midas Touch problem was avoided.

Kim et al. [120] suggested another menu design that aims to improve the precision of gaze-based selection and reduce the time required for selection. Incorporating lattices as a guiding mechanism for gaze gestures in the menu structure yielded a lower number of selection errors and shorter selection times as compared to conventional gaze interactive menu designs [163]. Ahn et al. [121] introduced a novel gaze-based VR/AR marking menu named "StickyPie" that addressed the limitations present in existing eye-tracking interaction techniques, including overshooting and false activation. The authors reported that StickyPie, which provides a scale-invariant marking menu, achieved over 10% improvement in efficiency compared to RegularPie, a more conventional scale-variant menu design. An alternative interface design was presented by Wei et al. [137] for the object locating process with gaze in several VR applications with a significant number of objects, resulting in a reduced task load and time compared to traditional methods. Recently, GazeDock was presented by Yi et al. [123] as a view-fixed peripheral menu that is automatically displayed when the user's gaze moves to the menu region. GazeDock, with a personalization and optimized selection algorithm, achieved an average selection time of 471 ms and a false trigger rate of 3.6% while being

TABLE II
A NON-EXHAUSTIVE OVERVIEW OF PAPERS DISCUSSED IN SECTION IV-B1.

Contribution	Characteristics
User interface design	[118] Hand-attached menu interacted with wrist rotation and eye gaze [119] Menu placement on Kuiper Belt region [120] Menu design incorporating lattices as a guiding mechanism [121] StickyPie: Solution for overshooting and false activation [122] Wheelchair control with waypoint navigation [123] GazeDock: View-fixed menu with personalized selection [124] VRDoc: Reading interface design [125] SwiVR-Car-Seat: Car interface design [126] Bionic Tracking : Cell tracking in 3D volumetric data
Target selection	[127] Outline Pursuits: Gaze and object movements based selection [128] Radi-Eye: Discrete and continuous selection [129] EyeSeeThrough: Merging confirmation and selection processes [130] Eye&Head: Head-supported gaze actions [131] Comparison of gaze interaction in selection and drawing tasks [132] Comparison of gaze, foot, head, and mouse interaction [133] Interaction with dwell time and pinch movement [134] Saccade-based confirmation method [135] Weighted Pointer: An error-aware gaze-based interaction [136] Fatigue analysis for eye-gaze and controller-based selection [137] Gaze-based label guidance for object locating task [138] Comparison of hand-free text selection task
Text entry	[139] TapGazer: Finger tapping and eye gaze [140] Interaction methods using blink or neck motions [141] SSVEP: Combines brain-computer interfaces and eye-tracking [142] Keyboard visualization and dwell&click-based interaction
Disambiguation & Depth	[143] VOR-gain-based depth estimation [144] Regression model with vergence-based features [145] Depth estimation with VOR gain [146] Comparison of disambiguation techniques
Intent	[147] Predict tasks based on eye movements [148] Logistic regression with focal attention and eye gaze [149] LSTM model based on sequence of eye gaze
Hand redirection	[150] Sparse Haptic Proxy: Mimic haptic feedback [151] REACH+: Physical interaction in virtual spaces to enhance realism [152] Hand redirection exploiting blink motion of the eye [153] Object manipulation tasks beyond arm reach [154] Error detection in gesture input with the help of gaze
Walking redirection	[155] Discrete rotation technique for both blind and open eyes [156] Prediction for the future positions of user with LSTMs [157] Prediction for short- and long-term future positions of user with LSTMs

preferred over dwell- and pursuit-based approaches. A virtual interface, VRDoc, was proposed by Lee et al. [124] for a reading task that might be useful for office workers. VRDoc incorporates eye-tracking-based interaction methods to reduce document selection and navigation time as well as required effort compared to existing VR reading interfaces. Günther et al. [126] designed an eye-tracking-based user interface called Bionic Tracking to facilitate object tracking by exploiting smooth pursuits of eyes in a virtual environment. Based on experiments conducted with 7 cell tracking experts, the interface, designed for tracking cells in 3D volumetric data using eye tracking achieved a speed-up of 2-10 times compared to traditionally used 2D point-and-click methods.

Additionally, it is essential to design user interfaces and interaction modalities by considering their specific needs and domains as well as the unique movements and conditions of the setups. To this end, Colley et al. [125] considered a vehicle setup and presented the SwiVR-Car-Seat to explore the impact of vehicle motion on VR interaction in automobiles. The authors designed an experiment using a low-cost rotating seat to observe the impact of vehicle rotation on touch, gesture, gaze, or speech interactions. The results revealed important insights for user interface design, indicating that vehicle motion had a detrimental effect on gaze- and gesture-based interaction methods. In contrast, touch and speech interactions were found to be more resilient in such environments. Additionally, Araujo et al. [122] proposed gaze-based control interfaces such as an overlay control interface, continuous-control interface, and waypoint navigation. In a wheelchair control setup, the semi-autonomous waypoint gaze interface yielded the fastest task completion time for each trial. The authors imply that this provides a more favorable user experience compared to other interaction methods.

b) Gaze-based Target Selection: Gaze-based user interface design requires efficient object and region selection to enable a smooth interaction experience. To this end, Sidenmark and Gellersen introduced Eye&Head [130], an approach that separately evaluates only gaze- and head-supported gaze actions for hands-free target selection. They found that eye-tracking data supported with head motion provides greater freedom to users during the selection process, as the user intent is more evident in such a setup. In another work, Sidenmark et al. proposed Outline Pursuits [127], which utilizes gaze to solve object selection problems in occluded virtual scenes. The authors assigned different motions to candidate objects and analyzed the correlation between gaze and object movements. The experimental results indicated that using Outline Pursuits, the selection process was completed with less effort in a shorter time than traditional ray-casting methods, also providing slightly better accuracy in highly occluded environments. Sidenmark et al. proposed another selection technique called Radi-Eye [128], to enable hand-free interaction in virtual smart home applications. Similar to their previous work, Radi-Eye uses gaze data to perform discrete or continuous target and object selections, while head movement is primarily used to confirm or modify the choice. As Radi-Eye provided a more precise and time-efficient interaction compared to previous works, it also provides important implications and design insights, especially for immersive virtual spaces.

Mardanbegi et al. presented EyeSeeThrough [129], a novel eye-tracking-based interaction technique that integrates both confirmation and selection processes through the line-of-sight direction. Results from user studies demonstrated that EyeSeeThrough surpasses the performance of two-stage selection methods in terms of time and comfort. Along the same lines, Pfeuffer et al. [131] evaluated the effectiveness of gaze-based interaction in virtual handheld menus. The standard selection methods, including using dwell time, gaze button, and cursor, were integrated with eye-tracking data and compared to a pointer-based selection method. The user studies on color selection and line drawing tasks indicated that gaze-based selection reduces physical effort compared to traditional selection methods. Minakata et al. [132] conducted a study to evaluate the performance of gaze, foot, head, and mouse pointing methods in selection tasks. The results indicated that head input was superior to gaze in terms of ease of calibration, effective target width, and throughput [164], which is the primary performance measure that takes into account both speed and accuracy of the users' responses. On the other hand, gaze input performed similarly to foot input.

Mutasim et al. [133] conducted another study to examine dwell and pinch movements as alternative versions of click-in devices supporting visual attention-based interaction. Their study indicated that pinch gestures could serve as a viable alternative to conventional button click-based selection methods. In another work, Mutasim et al. [134] evaluated the performance of a saccade-based selection and confirmation procedure in comparison with more traditional approaches such as dwell and button press. According to the authors, the saccade-based selection was the most time-efficient option but also the most error-prone. Recently, Sidenmark et al. [135] introduced a weighted pointer, an error-aware gaze-based interaction technique that is designed to maintain stability in the presence of eye-tracker sensor errors through the integration of fallback modalities. The authors demonstrated that a weighted pointer allowing automatic switching of modalities is more effective and favorable than techniques that require manual switching. In another recent work by Meng et al. [138], the authors investigated the efficiency of hands-free text selection techniques, including selection using the dwelling, blink, and voice. They found that target selection using blinks outperforms the other mechanisms in terms of time, accuracy, effort, and user preference. Additionally, to gain insights into the practical use of gaze-based selection techniques, Masopust et al. [136] analyzed the fatigue occurrence providing a comparison of selection techniques based on eye gaze and controller input. Their findings revealed that prolonged use of eye gaze could lead to fatigue, making it less suitable for interactions for an extended period of time compared to controller-based techniques.

c) Gaze-based Text Entry: Similar to the aforementioned eye-based target selection techniques, another way of interaction is text input. However, the nature of the task in text entry and its requirements are different; therefore, the interface design and interaction modalities should take those differences into account.

To this end, He et al. proposed TapGazer [139], which utilizes finger tapping to type on a virtual keyboard by incorporating gaze assistance for word selection. Lu et al. [140] presented BlinkType and NeckType text entry techniques, utilizing blink and neck movements in character selection, respectively. The authors found that using a blink signal is more favorable and reaches the highest word-per-minute rate compared to NeckType and dwell-based text-entry techniques. Another approach introduced by Ma et al. [141], combined brain-computer interfaces (BCIs) and eye-tracking data, which allowed VR users to compose ten words per minute and achieved 270 bit per second information transfer rate. Rajanna and Hansen [142] conducted a comparative investigation on the application of dwell-time and click inputs in conjunction with gaze information as a confirmation mechanism. The authors also analyzed various keyboard visualization options. According to their user studies, the most convenient combination was the use of click actions with visualization of the entire keyboard. In addition to these works, Mardanbegi and Thies [165] presented an eye tracking-based interaction toolkit EyeMRTK supporting text entry in Unity.

d) Gaze-based Disambiguation: Gaze-based interaction methods may suffer from inaccuracy and stability issues due to the imprecise target object detection caused by factors such as the accuracy of eye tracking, the complexity of virtual environments, or the nature of the task. In the following, we discuss eye gaze-based disambiguation techniques addressing these challenges.

Interaction methods that rely solely on the intersection of the gaze ray may cause less accurate detection of targeted objects, especially in scenes with objects at multiple depth levels. To address this issue, depth estimation methods based on the eye have been proposed to improve the interaction quality. Most existing approaches to this problem relied on the relationship between target depth and the vestibulo-ocular reflex (VOR) to fix gaze by moving the eye in the opposite direction of the head. Mardanbegi et al. [143] proposed a VOR-based depth estimation approach using VOR gain obtained by observing pupil centers, as a measure of in-depth computation, rather than the more conventional vergence measure. The VOR gain-based model, which only requires the use of one eye, was evaluated on eye-tracking data collected from 10 participants. Despite the limited number of participants, the findings of this study revealed that the performance level is comparable with the vergence measure conventionally used in in-depth estimation. Mardanbegi et al. [145] introduced a VOR-based technique to overcome disambiguation problems during gaze interaction by jointly utilizing eye and head movements to estimate gaze depth, leading to a better performance compared to conventional vergence-based methods in gaze depth estimation for the virtual objects that are located in deep in the scene. In another work, Weier et al. [144] built a regression model to improve gaze depth estimation by combining vergence measures with other features acquired through conventional ray-tracing techniques based on the user's point of regard. The model yielded a remarkable level of accuracy in terms of the average deviation from the benchmark depth over a 6-meter range while vergence measures offer accurate prediction only up to 1 meter. Similarly, disambiguation is also a significant challenge in hands-free interaction techniques. Chen et al. [146] evaluated disambiguation techniques such as head gaze, speech, and foot tap in different timing scenarios and found that head gaze outperformed other methods in eliminating disambiguation.

The utilization of user intent can also be employed to address disambiguation issues that arise in VR interaction based on eye tracking. By predicting user intent, it is possible to identify forthcoming movements of users and objects with which users will interact and the gaze point can be evaluated for the given user intent, which facilitates gaze-based interaction in ambiguous situations. To this end, the precise prediction of user intent is crucial and several works utilized eye-tracking-based approaches. For example, Keshava et al. [147] demonstrated the capability of point-of-regard (POR) regions to predict four different alignment tasks with cubes, achieving F_1-score of 0.51. David-John et al. [148] employed eye-tracking features to train a logistic regression model and revealed that characteristics of focal attention, gaze velocity, fixation, and saccade dynamics are highly relevant for predicting user intention. Hu et al. [166] proposed EHTask, a novel task classification method utilizing eye and head movement features, and showed that eye-tracking features differ based on the tasks in VR. In addition, EHTask performed better than state-of-the-art in task classification in 2D viewing. Similarly, for intent prediction, Alghofaili et al. [149] introduced an approach for the prediction of navigation assistance needs based on eye gaze and LSTMs, and the approach of the authors achieved a prediction accuracy of over 80%.

e) Gaze-based Hand Redirection: Hand redirection methods are essential for addressing the discrepancies between hand movements in virtual and real environments, as perfect alignment between the two cannot always be achieved. Discrepancies can lead to difficulties during hand interactions, such as the inability to interact with the intended objects. Eye-tracking-based methods can be utilized to assist in identifying the user's intent and direct hand to the right location.

Considering these, Cheng et al. [150] utilized a gaze-based hand redirection method to bridge the gap between real and virtual worlds with their proposed framework called Sparse Haptic Proxy. This framework consists of geometric primitives that mimic the haptic feedback of virtual objects. The authors employed the haptic re-targeting technique [167] that utilizes users' eye and hand behaviors to predict their intention and redirect their hand to a physical proxy. User studies showed that the proposed method achieved an accuracy of 97.5% for gaze-based user intention prediction and reported the maximal acceptable hand redirection angle as $40°$. Blink-Suppressed Hand Redirection (BSHR) was proposed by Zenner et al. [152] as a hand redirection approach based on body wrapping algorithm [150]. Unnoticeable instant changes are performed exploiting the natural blink motion of the eye, and real and virtual hand offset is adjusted with slight modifications when eyes are opened.

REACH+ proposed by Gonzalez et al. [151] combines eye-tracking data and hand motion to determine user intention in redirection frameworks, and it is designed to overcome challenges through physical interaction in virtual spaces. REACH+ redirects the users' hands to the intended target within arm-reachable range, improving the feeling of realism. In recent work, Sendhilnathan et al. [154] accomplished error detection in gesture input with the help of gaze dynamics. The authors analyzed the gaze patterns following gesture input and classified the input into three types: true input, input recognition errors, and user errors. Utilizing only gaze features such as fixation duration, saccade amplitude, and gaze velocity, the researchers successfully achieved a classification area under the curve of receiver-operator characteristic curve for one-vs-rest score (AUC-ROC-OVR) of 0.78 with a temporal convolutional network (TCN) and examined the consistency of such classification across tasks. More in terms of object manipulation tasks, Yu et al. [153] designed an interaction technique combining hand gestures with gaze information for object manipulation tasks in VR. They evaluated four alternative combination techniques while varying the object distance from the agent in the virtual environment. According to user studies, the authors showed that the combination of hand and gaze improved the usability and efficiency of object manipulation tasks in large environments where objects are not located within the arm's reach distance.

f) Gaze-based Walking Redirection: Another challenge in VR interaction, especially during mobile activities such as walking and running, is the limited space in the real world. Nguyen and Kunz [155] proposed a discrete rotation technique to minimize the required area in a real workspace. The authors first assessed the threshold for the discrete rotation in virtual scenes and obtained $9.1°$ and $2.4°$ for closed and open eyes, respectively. Then, they applied discrete scene rotation while the agent was walking and reduced the required area by 20% for these tasks. An alternative method to facilitate virtual walking activity in constrained physical space proposed by Sun et al. [168]. The authors employed a path-planning algorithm that considered positions of pre-existing static and dynamic obstacles, including walls and furniture, as well as multiple VR users within a designated room. Furthermore, the algorithm made use of the saccadic suppression periods to facilitate efficient path planning. Furthermore, the authors presented a subtle gaze direction (SGD) technique to increase the number of saccades, considerably enhancing the redirection gain.

Predicting future physical positions is a beneficial way to optimize the use of limited physical area [156], [157]. Stein et

TABLE III
A NON-EXHAUSTIVE OVERVIEW OF PAPERS DISCUSSED IN SECTION IV-B2.

Contribution	Characteristics
Gaze prediction	[159] DGaze: CNN model using object-, head-, and saliency-based features
	[169] SGaze: Data-driven model for real-time gaze prediction
	[170] FixationNet: Eye&Head tracking and task-related data
	[171] In-game variables and gaze features
	[172] Deep Future Gaze (DFG): GAN model
	[173] Head orientation prediction with linear SVM
Scanpath prediction	[174] SaltiNet: CNN model with saliency volumes
	[175] PathGAN: GAN model for scanpath prediction
	[176] ScanGAN360: Scanpath generation from graph
	[177] Utilization of clustering and graph-based algorithms
Saliency map estimation	[178] Benchmarking different methods
	[179] Multi-perspective 3D saliency map estimation
	[180] User-specific behaviors
	[181] LSTM model with head and saliency features
	[182] User attention in cinematographic VR movies
	[183] Salient object detection (SOD) & 360-SSOD dataset

al. [156] proposed to use future location predictions obtained through an LSTM model utilizing eye-tracking features. The authors demonstrated that the model, which can be applied in redirection approaches, is capable of predicting the position of users 2.5 seconds in advance, with an average error of 65 cm. A similar approach relying on the LSTM model was proposed by Bremer et al.[157], evaluating short-term (i.e., 50 ms) and long-term (i.e., 2.5s) predictions separately. The optimal performance in short-term predictions was obtained by utilizing position and orientation features. In contrast, the inclusion of eye-tracking features with these features resulted in the best performance for long-term predictions. The authors also stated that especially short-term predictions might be valuable to optimize computational sources for rendering or transmission bandwidth in streaming activities.

2) Predictive Gaze Analysis: Predicting the future gaze locations enhances the efficiency of gaze-based rendering techniques, particularly foveated rendering; therefore, methods combining gaze prediction and foveated rendering have gained strong momentum in recent years. This section discusses the related algorithms for gaze location prediction, scanpath prediction, and saliency map generation. We provide the overview of the papers discussed in this section in Table III.

a) Gaze Prediction: Foveated rendering methods depend on accurate gaze prediction, which directly impacts the performance and usability of foveated rendering applications. Therefore, researchers have been dedicating significant attention to exploring and implementing effective gaze prediction techniques, leveraging a range of methods from pure statistical ones to neural networks. For instance, Deep Future Gaze (DFG), proposed by Zhang et al. [172], is a GAN-based model that exploits spatial-temporal CNN as an encoder and anticipates upcoming gaze positions while generating future frames based on the current ones. A by-product of their work is an object search task (OST) dataset, considered one of the most extensive egocentric datasets. Hu et al. [169] proposed SGaze that utilizes eye-gaze information collected by an integrated eye tracker to forecast future gaze locations. SGaze was designed as a data-driven model based on eye-head coordination for real-time gaze prediction and is based on statistical models exploiting the relation of gaze and head movements, and it does not require any additional hardware or eye-tracker support. In this model, the latency in the movements of the eye and head was taken into account, and SGaze achieved better angular distance values compared to the baseline, which used the screen center as the gaze point. More recently, Hu et al. [159] also presented DGaze as a CNN-based gaze prediction algorithm incorporating dynamic object positions, head movements, and saliency features extracted with SAM-ResNet saliency predictor [184]. Object positions, head velocity data, and previous gaze location sequences are processed with a sequence encoder model to increase the precision. This technique was applied in real-time, and near-future gaze predictions were carried out corresponding to short-term predictions such as 200-1000 ms. DGaze achieved better angular distance values in both dynamic and stationary scenes compared to their prior method, SGaze [169].

FixationNet proposed by Hu et al. [170] is a neural network utilized to predict near-future gaze locations up to 600 ms in task-oriented virtual experiences utilizing eye-tracking data, head movements, saliency maps, and task-related information. Data from 27 individuals during the visual searching task in VR indicated a strong correlation between fixation positions and head movements, saliency maps, and content-related information. Besides estimating the gaze position, the proposed model outperforms state-of-art models in gaze prediction tasks for 150-600 ms in task-oriented and free-viewing experiments. Another gaze prediction method was introduced by Koulieris et al. [171], specifically for games that are heavily task-oriented. The proposed approach exploits the correlation between the game variables in their current states and gaze locations. Moreover, predicted gaze locations were used in the dynamic disparity manipulation method to improve the depth sense in the game scenes. Vielhaben et al. [173] predicted the user's future head orientation using past head movements and eye-tracking data by training a linear SVM model. Their model is able to predict future view-ports without utilizing any information related to the content in spherical videos; hence, it can be utilized to improve rendering performance as well as reduce the transmission load.

b) Scanpath Prediction: To provide a comprehensive overview of gaze prediction, we discuss the scanpath prediction referring to the sequence of gaze points in a given virtual environment. Unlike gaze prediction techniques that forecast individual samples, scanpath prediction involves predicting the sequence of gaze points as a whole, thereby allowing for the estimation of complete human attention on a given scene. While these two concepts are related, they serve different purposes and require different types of models and algorithms for prediction. To this end, Assens et al. [174] proposed SaltiNet, a deep CNN trained first to generate saliency volumes that capture temporal information, and then the scanpath is sampled from the saliency volumes. In another work, Assens et al. [175] designed a novel GAN model, called PathGAN, that employs a convolutional-recurrent architecture for both the encoder and decoder in order to predict scanpath. Recently, Martin et al. [176] introduced another GAN-based scanpath prediction model, ScanGAN360, which deploys a novel GAN objective function based on dynamic time warping. Unlike the ScanGAN360, Zhu et al. [177] suggested generating scanpaths by building a graph from a saliency map. The saliency map is first binarized and clustered into centers. Then, a weighted graph is created with the cluster centers as nodes, and a scanpath can be generated from the graph.

c) Saliency Map Estimation: A saliency map, also known as a heat map or attention map, is a visual representation of eye-gaze data on an image or on a scene. It serves as a visualization tool that highlights the regions that attract the most attention from human observers. An illustration of a saliency map, which shows the distribution of human attention over an image, is provided in Figure 5. The saliency map lays the foundation for numerous gaze-based VR applications and hence plays an important role. Different approaches have been proposed in the literature for saliency map estimation and generation. John et al. [178] benchmarked four different methods for saliency map generation in VR. The authors suggested utilizing the modified Gaussian kernel [185] with a scale of 5% for 360° saliency map generation due to ease of implementation. Pfeiffer and Memili [179] presented an approach that tackles challenges such as changing perspectives, dynamically moving objects, and depth of fixations in the generation of saliency maps for 3D environments. They aggregated the gaze data from multiple users and represented the distribution of visual attention at the object level using textures. Their method offers high-quality saliency maps for multi-perspective eye-tracking recordings. Sitzmann et al. [180] studied the difference in saliency patterns between desktop and stereoscopic vision among 169 participants and adapted traditional saliency prediction methods according to user-specific behaviors such as particular fixation biases. Nguyen et al. [181] conducted research on predicting head movement using an LSTM model in conjunction with saliency maps. Maranes et al. [182] exploited a dataset of 3259 users watching cinematographic VR movies and measured user attention via saliency maps. Closely related to visual saliency, Ma et al. [183] aimed at salient object detection (SOD) in 360° panorama images with the help of eye tracking and proposed a novel dataset, i.e., 360-SSOD, including manually annotated object-level saliency ground truth with balanced semantic distribution compared to existing datasets.

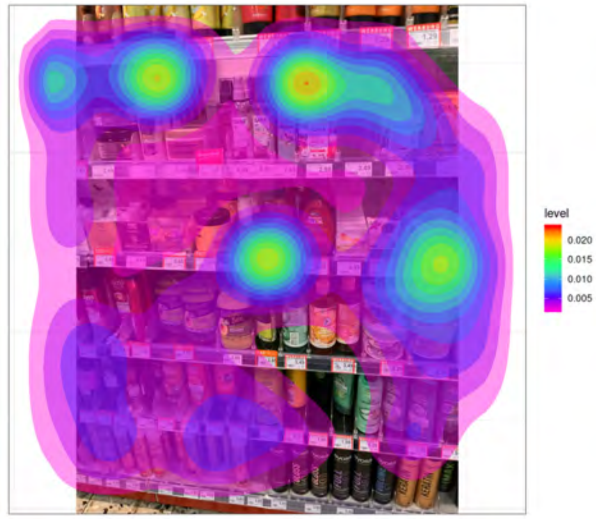


Fig. 5. A saliency map is a gaze visualization tool that highlights regions of an image with a color-coding scheme. An example of such a map is provided here.

3) Rendering Techniques: The quality of the 3D scene display is of paramount importance for VR devices, as it directly impacts the overall immersive experience for users. Therefore, it has been of great importance to render with high resolution and high refresh rates in VR environments. However, several challenges exist in this area regarding computational issues for real-time rendering, computational power, and realistic natural depth perception. To overcome those, different rendering techniques utilizing gaze information have been proposed in the literature. In this section, we discuss rendering techniques in three categories based on their objectives: ensuring computational load and power efficiency, optimization of transmission efficiency, and scene quality enhancement. By exploring these categories, we aim to provide a deeper understanding of the current state of the art in rendering in VR and the potential for future developments in this field. We provide an overview of the papers discussed in this section in Table IV.

a) Foveated Rendering: In this section, we first provide a comprehensive review of foveated rendering techniques. Subsequently, we discuss gaze-based methods with similar objectives. In a nutshell, foveated rendering approaches focus on graphical computations and rendering around gaze fixation in a high-quality manner while blurring other parts of the scene to reduce computation load without deteriorating the user experience. Figure 6 depicts illustrations of images with blurred peripheral regions.

To this end, Patney et al. [186] proposed a foveated rendering architecture to lessen computation power and reduce the number of shades in the scene. In their architecture, a perceptual target image is generated by adjusting the blur filter width according to retinal eccentricity to prevent temporal and spatial aliasing. The authors discovered that contrast enhancement could provide a larger tolerance for blurred region size through a user experience study, and as a result, contrast enhancement

TABLE IV
A NON-EXHAUSTIVE OVERVIEW OF PAPERS DISCUSSED IN SECTION IV-B3.

Contribution	Characteristics
Foveated rendering (FR)	[186] Blur Filter width based on retinal eccentricity [187] Eye dominance measured with Miles test [188] 4D-Light areas and gaze-based acceleration [189] Content aware foveated rendering using luminance contrast [190] Neural radiance fields (NeRF) [191] Foveation, depth of field, and longitudinal chromatic aberration [192] FR with a subsampling technique for ray tracing [193] RMFR: Rectangular mapping-based foveated rendering
Saccade-based FR	[194] Polynomial fitting for saccade landing position estimation [195] Recurrent neural network model [196] Data augmentation for neural networks
FR assessment	[197] Foveated rendering assessment
Power efficiency	[198] Gaze-based approach for colour discrimination [199] Gaze-based dynamic refocusing for gigapixel panoramas (GPP) [200] FocusVR: Gaze-based intelligent dimming technique
Streaming	[201] Log-rectilinear transformation [202] Codec supporting multi-resolution in a frame [203] Low resolutions 2D scene processing and mapping to 3D scenes
Depth sense enhancement	[204] Blurring based on focal points and gaze [205] Foveated rendering accounting for ocular parallax [206] Ocular parallax aware gaze-based stereo rendering [207] Gradual and stereoscopic depth adjustments
Vergence-accommodation conflict	[208] Phase-only spatial light modulator (SLM) [209] Decomposition method with gaze and head motions [210] Binocular disparities & screen distance [211] Accommodation-invariant display [212] Minimum eye-tracking accuracy for varifocal displays [213] Multi-focal and single-focal comparison

was applied as a post-processing step to the image. They demonstrated that the proposed technique achieved a similar level of temporal stability compared to temporally filtered non-foveated images and validated it through frequency analyses, which showed alignment with the perceptual target. Ye et al. [193] introduced a technique called rectangular mapping-based foveated rendering (RMFR), which renders scenes with non-uniform foveation levels based on eccentricity and scene complexity. RMFR offers superior visual quality compared to conventional foveated rendering methods while requiring minimal rendering cost. Meng et al. [188] proposed another foveated rendering approach, referred to as 3D-kernel foveated rendering (3D-KFR), which is combined with an eye-tracking-based acceleration algorithm. The proposed algorithm, designed specifically for visualizing high-resolution microscopic 4D-Light fields with depth cues, provides rendering acceleration up to a factor of 7.28 in HMDs.

Similar to the aforementioned works, Deng et al. [190] presented a gaze-contingent rendering approach with neural radiance fields (NeRF), which allows the rendering of 3D scenes with photo-realistic quality. However, this approach also requires heavy computation, which causes latency in the rendering process. To address this issue, the proposed technique involves encoding each retinal region, including the foveal, mid-periphery, and far-periphery regions, with varying levels of visual perception. According to their experiments, the authors indicated that the proposed method incorporating eye-gaze information reduces latency without compromising perceptual quality. Additionally, Meng et al. [187] took eye dominance into account and designed an approach providing more visual details to the dominant eye than the non-dominant eye to enhance the efficiency of foveated rendering. An image with a broader foveation area is rendered for the eye with higher ocular dominance. The authors conducted a study to estimate foveation level parameters individually for each eye, using Miles test [214]. They showed that the rendering performance of the devices was boosted with the proposed technique for the same level of perceived quality.

Most of the discussed foveated rendering algorithms consider the peripheral region and the sensitivity of human eyes to model foveated areas. In addition to those, Tursun et al. [189] introduced a content-aware foveated rendering method considering the luminance contrast of the exhibited content. The proposed model can forecast the resolution parameters as a function of luminance patterns by processing low-resolution frames before carrying out high-resolution rendering. Their user studies revealed that with their model, rendering performance increases while maintaining an imperceptible foveation layer. In a different study, Liu et al. [191] proposed a hybrid foveated rendering method that incorporates foveation, depth of field, and longitudinal chromatic aberration. Their methodology accounts for vergence and accommodation, visual acuity eccentricity, and color vision. The authors stated that the proposed model outperforms the state-of-the-art rendering techniques in frame rates while providing at least the same level of visual scene quality. In parallel to these works, Kim et al. [192] presented a promising foveated rendering approach to deal with the high computational demand of ray tracing, which can be considered as an advanced technique employed to achieve high visual quality and more realistic effects in scenes. The proposed foveated rendering approach includes a selective subsampling technique, which gradually decreases the rate in

Fig. 6. Foveated rendering provides a higher level of detail at the gaze location of the viewer while reducing the level of detail in the peripheral vision. The gaze locations are shown with blue circles in the figures.

the peripheral region while preserving high sampling rates in the foveal region. This allows the practical use of ray tracing in HMDs by reducing the high power requirement.

A handful of the rendering techniques rely on fixation prediction, and such techniques may suffer from latency which might result in the degradation of user experience because of the mismatch between actual gaze location and predicted location, especially during saccades. Therefore, forecasting the saccade landing points by exploiting saccadic suppression is essential to improve the quality of the foveated rendering [194], [195], [196]. The algorithm proposed by Arabadzhiyska et al. [194] forecasts the final position of saccades during the saccadic behavior and then renders the image considering the predicted landing location. To determine the final saccade location, the direction of the movements is predicted according to the previous gaze samples, while the saccade amplitude is predicted with a polynomial fitting approach. The authors indicated that this technique could be easily integrated into the existing foveated rendering systems to improve their performance. Similar to Arabadzhiyska et al. [194], Morales et al. [195] utilized recurrent neural network (RNN) models exploiting dynamic temporal relationships to predict the saccade landing point. Their LSTM-based model provides prediction at a lower error level compared to other state-of-the-art techniques. In RNN models such as LSTMs, the acquisition of suitable training data is vital for discerning temporal relationships. To address this challenge, Griffith and Komogortsev [196] proposed a data augmentation technique to improve the saccade landing point estimation with neural networks. Time-shifted imitations of the training data are used to boost the estimation reliability of the starting time of the saccadic movement. This augmentation method improves the median accuracy of the predicted final saccade point locations both for LSTMs and feed-forward neural network models.

Apart from the discussed foveated rendering techniques, qualitative or quantitative performance evaluation approaches are necessary to assess and compare the performance of existing rendering techniques properly. In addition, the robustness and effectiveness of the introduced assessment methods could vary in different test scenarios. Subjective assessment methods for foveated rendering were evaluated by Hsu et al. [197] in terms of efficiency and consistency. Efficiency is assessed based on the required time to reach perceptual ratio convergence. Meanwhile, the distribution of individual quality of experience (QoE) scores, which indicates the level of user satisfaction, expectation, and perception, is used for consistency. The authors stated that no subjective evaluation method could be defined as superior to any other, but they provide the research community with information on evaluation metrics so that researchers can decide on a metric that fits their specific needs.

While we present a comprehensive discussion of various foveated rendering techniques designed to alleviate VR devices' high power and GPU requirements and make them practically usable for real-world applications, there are several other techniques based on eye tracking that address similar challenges in different ways. These are mainly based on gaze-based color manipulation and dimming techniques [198], [200]. To this end, Duinkharjav et al. [198] presented a real-time applicable power reduction method based on gaze information and color discrimination. The authors achieved a display power requirement reduction of up to 24% with minimal degradation in perceptual quality. Similarly, Wee et al. [200] proposed a system called FocusVR to overcome the high power requirements of VR/AR devices with advanced high-resolution displays. Power consumption efficiency was enhanced by combining techniques like intelligent dimming, vignettes, and color mapping. The eye-tracking capability of these devices is utilized to execute more aggressive dimming techniques with less impact on user experience. The authors reported that FocusVR could reduce display and system power consumption up to 80% and 50%, respectively. In addition to these works, Lyu et al. [199] proposed an approach using eye-tracking data to deal with high GPU requirements and lags in gigapixel panorama (GPP) displays in HMDs. A rendering technique supporting panning, tilting, zooming, and dynamic refocusing based on gaze information was proposed to display GPP scenes in HMDs and the introduced system is capable of keeping the FPS rates above 50 without using high-end GPUs.

b) Streaming Techniques: Techniques that are used in foveated rendering can also be utilized or customized to reduce the bandwidth requirements of the high-resolution 360-degree video streaming, especially for VR devices [201], [202]. In such setups, eye-tracking data is processed, and content providers transfer the data based on gaze position using a similar methodology employed in foveated rendering. Although different approaches exist, the general framework for streaming solutions aiming to optimize bandwidth efficiency is given in Figure 7.

Li et al. [201] proposed a log-rectilinear transformation-based method using summed-area table filtering and off-the-shelf video codecs, that optimizes huge bandwidth requirements of the high-resolution live video streaming services for VR headsets. Log-rectilinear-based transformation provides about 10% efficiency in bandwidth usage compared to traditional log-polar transformation. It also reduced the flickering as measured by a metric suggested by Winkler et al. [215]. As an alternative, Lungaro et al. [202] presented a novel codec method, which intentionally causes some errors in the less-sensitive visual areas to support streaming video frames having regions with different qualities. The proposed solution offers an 83% improvement in the bandwidth requirement for high Quality of Service (QoS) levels over traditional solutions. Additionally, Chen et al. [203] proposed a gaze-contingent streaming approach that identifies the essential regions in 2D scenes and then maps them to 3D scenes to improve 3D rendering and streaming efficiency. The authors argued that the proposed system provides a better visual scene quality

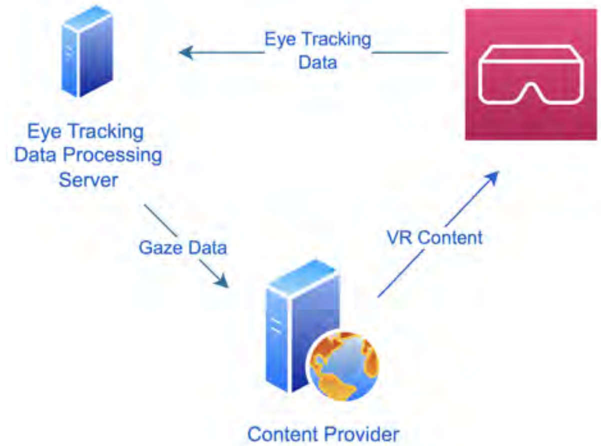


Fig. 7. A streaming framework example. Eye-tracking data is provided to a server that estimates future gaze locations and communicates with a content provider. The content provider streams VR content considering gaze position and optimizes the transmission process.

with temporal consistency compared to other systems.

c) Improving Scene Realism and Depth Perception: The aforementioned works, such as in the area of foveated or streaming techniques, offer solutions to deal with or mitigate the technical and computational limitations of VR devices, such as issues regarding power consumption, computational load, and bandwidth requirements. In addition to these, achieving high scene quality and high levels of realism are other challenging concepts in VR technology, which are related to the visual acuity of virtual scenes. The majority of research in this field focuses on depth perception, which is related to the degree of realism. Therefore in the following, we discuss the rendering techniques that have been introduced to enhance the realism sense focusing mainly on the users' depth perception. Additionally, we also explore the proposed solutions for enhancing depth perception, which addresses the issue of vergence-accommodation conflict (VAC) in HMDs.

Mauderer et al. [204] presented a gaze-contingent technique allowing the production of more realistic 3D images with improved depth perception by modifying the blurring rate of the objects in a range of distances considering gaze locations. The authors achieved higher levels of perceived depth and realism by employing focal points in combination with the gaze. Konrad et al. [205] proposed a rendering method based on ocular parallax exploiting small changes of images on the retina caused by eye rotation. The ocular parallax's visibility was evaluated with a series of experiments to measure perceived relative distance in the AR/VR scenes. According to the authors, ocular parallax rendering combined with gaze contingency offers a more authentic experience and improved depth perception in VR scenes. Another work accounting ocular parallax presented by Krajancih et al. [206] as a solution to the depth distortion in conventional rendering models. The introduced gaze-contingent stereo rendering method reduces the shape and disparity distortions and provides a more consistent depth sense while maintaining perceived relative distances between the rendered objects in a scene compared to conventional methods. Kellnhofer et al. [207] introduced another technique by using eye tracking as well as predicted gaze data to enhance depth perception while preserving the visual quality and comfort of the users. Their model employs a controller executing unperceived local depth adjustments for the stereoscopic content to optimize the trade-off between depth reproduction and visual comfort. The evaluations showed that their method provides a significant enhancement of depth perception while preserving the visual quality.

Vergence-accommodation conflict (VAC), which is referred to as the mismatch between the vergence and accommodation points of the eye, is another challenge that can affect the perceived image quality, the sense of depth, and realism in near-eye displays. This conflict occurs due to changes in the shape of the eye lens during eye movements. In some VR displays that are designed for a fixed distance, the degradation of perceived image quality can be attributed to the conflict between the vergence and accommodation of the eyes. To address this challenge, Matsuda et al. [208] proposed the use of a phase-only spatial light modulator (SLM) made from the liquid crystal on silicon (LCOS) as a lens with a dynamically adaptive shape that is positioned between the display screen and the viewing optics. The authors presented a rendering technique that decomposes the scene and generates images for each display plane while taking the gaze location into account. Mercier et al. [209] proposed a computationally plausible decomposition method that utilizes gaze and head motions to align the layers of the multifocal displays in HMDs. The experiments, which were conducted on a multifocal testbed designed for eye-tracking and accommodation measures, showed that eye tracking could be effectively utilized for plane alignment tasks. In addition, Konrad et al. [211] introduced an accommodation invariant display approach for near-eye displays, which makes scenes independent of the accommodation state of the eyes. This approach replaces traditional retinal blur with disparity and vergence to drive accommodation. Batmaz et al. [213] examined the effect of VAC on user performance during a hand-pointing task within arm's reach using multi-focal and single-focal VR displays. The authors concluded that multi-focal displays are better than single-focal displays in terms of both time and error rate, which corresponds to the percentage of missed targets during virtual hand interaction tasks. More recently, Aizenman et al. [210] systematically explored eye movement and binocular disparities when wearing an HTC Vive Pro Eye. The authors proposed an optimal screen distance to minimize the adverse effects of the VAC by analyzing the distribution of fixation distances. However, the accuracy of eye tracking can pose challenges for these solutions. To this end, Dunn [212] reported that the minimum eye-tracking accuracy required for the implementation of varifocal displays in VR is $1.444°$.

C. Understanding Human Cognition, Visual Attention, and Perception

Plenty of works in the literature focus on computer vision- and machine learning-based approaches to identify eye regions and detect gaze and concentrate on human-computer interaction, mostly in an online fashion for rendering and real-time user support. In addition to those, researchers often collect eye-tracking data from VR setups and process and analyze them in an offline fashion to understand human behavior from a perspective of cognitive and visual attention processes as well. Therefore, in this section, we give a comprehensive review of both theoretical and practical studies of eye-tracked VR with respect to understanding human behaviors. We first outline in Section IV-C1 how eye tracking can be used to understand human cognition in VR and during the use of HMDs. Then, we explore the works that investigate the relationship between human visual attention and gaze behavior in Section IV-C2. Later, in Section IV-C3, we provide the research related to human perception of VR spaces, considering perceived realism and cybersickness.

1) Human Cognition: Unlike the fields that are tangible, the realm of human cognition is characterized by abstraction and is hence more elusive. To close this gap, eye-tracked VR has shown great potential in the study of human cognition. Gaze features such as eye movement patterns, pupil diameters, and aggregated eye features such as fixation and saccade durations have been widely used as indicators for different human cognitive processes. For instance, Bozkir et al. [216] showed cognitive load differences in a driving simulation in VR with machine learning and were able to predict human cognitive load with an accuracy of over 80% by incorporating gaze features, particularly by using pupil diameters. Souchet et al. [217] conducted cognitive load prediction using features from different modalities, including gaze features, electrocardiogram, and electrodermal activity. The authors found that electrocardiogram features were the most relevant features correlated with self-reported subjective cognitive load levels compared to the other modalities in their setup. Duchowski et al. [218] pointed out the drawbacks of pupillometry in cognitive load assessment that is caused by ambient illumination and off-axis distortion and suggested the wavelet-based estimation of pupil diameter oscillation frequency as an alternative feature. The suggested metric showed great potential to classify task difficulty in a user study where 17 participants were asked to do math calculations.

Cognitive load prediction tasks were also utilized in aviation-related scenarios. Luong et al. [219] utilized features from various modalities, including pupil features such as pupil diameter, pupil diameter amplitude, and constriction and dilatation speed of pupil to evaluate the cognitive load in a VR flight simulation. Using a random forest estimator, the authors were able to classify four levels of cognitive load with an accuracy of up to 65%. Wilson et al. [220] also studied cognitive load in flight simulation with a deep learning model using various sensing modalities. They fed pupil dilation and blink features into their models and obtained 80% accuracy in the cognitive load estimation task.

Related to cognitive load, Bækgaard et al. [221] examined the correlation between pupil dilation and the difficulty index in Fitts' Law [222]. Fitts' Law describes the relationship between the time, distance, and accuracy of an individual's movement in reaching a given target while taking the level of cognitive load or task difficulty into account. It enables the assessment of task difficulty or cognitive load by defining the difficulty index as a quantitative measure. Findings of Bækgaard et al. indicate that tasks that demand higher cognitive load lead to a larger pupil diameter, whereas there is no significant relationship between pupil diameter and the Fitts' Law index, which is related to motor task complexity. One common challenge in measuring pupil dilation is its sensitivity to illumination. To overcome this challenge, John et al. [223] attempted to exclude the impact of the light source on pupil dilation in both 2D monitors and HMDs using three different models. Although a linear calibration model yielded prediction errors of 0.41 and 0.60 mm for pupil constriction and dilation, respectively, and in 2D, an exponential model dominates with an error of 0.27 mm in VR. More related to application use cases, in recent work, Gao et al. [224] compared the effects of five VR locomotion methods on cognitive response, including arm swinging, dash, grappling, joystick, as well as teleportation. To measure cognitive load, gaze features like blink rates, fixation durations, and saccade amplitudes were employed. The authors discovered that locomotion methods could significantly influence the cognitive state in VR, while joystick and teleportation resulted in lower cognitive load in participants.

Human emotions such as fear and frustration also play an important role in the human cognitive state, and an increasing amount of studies have been devoted to identifying the relationship between gaze behavior and human emotion in virtual spaces. For instance, Pan et al. [225] used pupil diameter changes to measure the influence of the fear of the sea on working memory. The authors designed n-back memory tests for 29 subjects and found that pupil diameter changes noticeably increased in undersea simulation in all tasks. Furthermore, Luong and Holz [226] investigated users' pupillometry responses along with pulmonary, electrodermal, and cardiac data to capture sensations of fear, frustration, and insight. Insight refers to the moment when a player suddenly understands the solution to a puzzle or clue. They achieved an F1 score of over 71% to predict these senses using classification models including logistic regression and support vector machine models.

Apart from the estimation of the cognitive load, inferences about human cognition have also been used as a supportive source of information in various application domains, especially to enhance the interaction experience in VR. Lindlbauer et al. [227] utilized cognitive load, which was estimated by the index of pupil activity metric based on pupil dilation change, to optimize and eliminate the manual adjustment process of the level of information displayed in mixed reality glasses. The authors showed that, with their approach, the amount of secondary task interactions was reduced by 36%. Kübler et al. [228] utilized cognitive load as an indicative factor to detect hazard situations in a driving simulator. SVM models trained with wavelet transformation components extracted from pupil diameter data were used in the detection of hazardous situations. The authors demonstrated that pupil dilation could be used as an indicator of hazard perception in driving simulations. However, they also noted that it cannot be relied upon as a stand-alone detection mechanism due to the number of false positives.

When applications of human cognition are considered, one of the most important domains is medicine, as human cognitive processes are related to neurological and psychological diseases. In addition, since it is possible to manipulate and control experimental conditions very precisely in VR, researchers often employed VR and eye tracking together in this domain. For instance, to understand the response of children suffering from autism spectrum disorder (ASD) to facial expressions, Bekele et al. [229] conducted a user study with 10 ASD patients and 10 typically developing (TD) peers in multiple emotion recognition tasks using a facial expression system in VR. By analyzing participants' gaze features such as region-of-interests (ROIs), fixation durations, blink rates, and pupil diameters, the latter being indicative of cognitive load, the authors found that the way how individuals with ASD identify emotional faces is significantly different from their TD peers. In another work by Bekele et al. [230], the authors introduced a novel multi-modal adaptive social interaction in a VR platform and suggested the use of gaze information for adaptive intervention. In particular, they used ROIs to adjust the face occlusion of virtual characters in real time. In their emotion recognition experiments with 6 ASD and 6 age-matched children, a performance increase of 3% associated with the novel gaze-sensitive mechanism was observed. Similarly, Kim et al. [231] presented a VR-based interactive social skill training system (VISTA) that uses eye-tracking data, including ROIs and pupil features, to understand the characteristics of people with ASD. In a user study of 20 participants, the authors found that the ASD group was highly engaged with VISTA while showing a dramatic difference from the neurotypical group in biological signals, such as a larger variation in pupil diameters, which in general indicates a higher cognitive load. The research in the medical domain that includes human cognition, eye tracking, and VR is not limited to ASD. For instance, Orlosky et al. [232] presented a remote diagnosis system in VR for neurodegenerative diseases like Parkinson's. The authors designed various tasks to elicit abnormal eye behaviors such as ocular tremors, square wave jerks, and abnormal pursuits, which could be indicators for neurodegenerative diseases and cognitive load. Based on the experimental results, the authors claimed that the VR interface successfully elicited five types of abnormal eye movements, while eye physicians could identify three out of four abnormal eye behaviors. The authors also implemented three different visualization techniques for eye movements, which can be helpful for doctors as small movements can hardly be identified by humans. While the diagnosis of such diseases could not only be done with the help of eye tracking and VR due to the fact that those diseases are far more complex for only machines and data-driven practices to identify, such systems could be used for supporting doctors in their regular workflows. In Table VI, we provide a brief overview of papers discussed in Section IV-C1, including their purposes, eye-tracking features, and the number of participants involved in their experiments, whereas Table V includes abbreviations for the features.

2) Visual Attention: In this section, we give a holistic view of prior works linking eye-tracked VR and human visual

TABLE V
FEATURE ABBREVIATIONS.

Name	Abbreviation	Name	Abbreviation
Pupil-based features	pup	Blinks & Eye openness	blnk
Fixation	fix	Saccade	sac
Interpupillary distance	ipd	Oscillation frequency	freq
Eye vergence angle	eva	Smooth pursuit	spur
Phoria	phr	Eye gaze	gaze
Convergence	cvgn		

TABLE VI
AN OVERVIEW OF THE PAPERS RELATED TO HUMAN COGNITION.

Paper	# Users	Purpose	Eye-tracking Features
[216]	16	Cognitive load prediction in driving simulation	pup
[217]	92	Cognitive load prediction	pup, sac, blnk
[218]	17	Cognitive load prediction	pup, freq
[219]	75	Cognitive load prediction in flight simulation	pup
[220]	40	Cognitive load prediction in flight simulation	pup, blnk
[221]	27	Cognitive load analysis in physiology	pup
[223]	24	Cognitive load analysis	pup
[224]	15	Cognitive response analysis for VR locomotion	blnk, fix, sac
[225]	29	Emotion recognition	pup
[226]	24	Emotion recognition	pup
[227]	12	Cognitive load prediction	pup
[228]	31	Cognitive load analysis for hazard detection	pup
[229]	20	Understanding ASD	gaze, fix, blnk, pup
[230]	12	Understanding ASD for adaptive intervention	gaze, pup, blnk, fix, sac
[231]	20	Understanding ASD for social skill training	gaze, pup
[232]	16	Parkinson's disease detection	gaze, pup

attention. Human visual attention is highly related to where people look at the scene mainly with their eye gaze and often correlates with cognitive and perceptual abilities. In addition, visual attention is affected by the 3D scene design and external stimulus-related factors in virtual spaces. To understand these, researchers have addressed various aspects of human visual attention. Schmitz et al. [233] explored how peripheral flicker and central arrow stimuli differently impacted visual attention guidance in virtual panoramic videos. Based on their analysis conducted with 25 users, the researchers reported that the participants preferred the central arrow stimuli over the peripheral flicker stimuli, perceiving the former as more rewarding. Additionally, the authors claimed that traditional attention mechanisms might not be fully applicable to panoramic videos.

The effect of endogenous and exogenous cues to orient visual attention was investigated by Soret et al. [234], with eye-tracking analysis. The results showed that such cue types reduce the reaction time of the user as initially expected. Liu et al. [235] performed unordered tasks more efficiently over time by employing dynamic visual cues to guide users' visual attention towards specific substeps of the performed task. The proposed method allows the user to determine the sequence of task completion, which subsequently influences the presented set of cues which are dynamically updated based on hand proximity and eye gaze. The authors stated that visual cues based on eye gaze improve task completion times more significant than the cues based on hand proximity.

Lange et al. introduced HiveFive [236] as an attention guidance technique designed to direct the users' attention to context-relevant points using swarm motion. The proposed technique achieves lower response latency, resulting in a less adverse impact on immersion. Visual attention guidance was also proposed in VR-based driving simulators as well. For instance, Bozkir et al. [237] provided gaze-aware warning cues for pedestrian crossing scenarios in a low-cost VR setup to enhance driver attention. The experimental results, which include metrics such as minimal distances to pedestrians, participant pupil diameters, and drivers' inputs on accelerator and brake, showed that eye-tracked VR might be an effective tool for driver training in safety-critical situations due to the effectiveness of the visual cues.

Notifications are often considered a form of external stimulus that may interrupt the user during virtual experiences. Gaze-based visual attention analysis could help to understand the impact of notifications and potentially provide a less disruptive VR experience. To this end, Hsieh et al. [238] explored the perception of notifications by VR users when they received them in a message form from the real world. The authors provided design suggestions to eliminate the disruptive effect of these notifications, such as by measuring users' level of engagement using eye-tracking data to find the convenient times for notifications and locating the message according to the user's peripheral area. Similarly, Chen et al. [239] proposed a technique using a deep learning model by processing time series sensor data, including gaze angles and gaze shift rates, to predict the optimal time for notification. Their deep learning model was able to predict a favorable time for notifications with 71% precision, and further improvements were observed when the information related to user activities and engagement was considered. Similarly to the aforementioned works about notifications, the assessment of visual attention is considered useful for the identification of appropriate time and region to insert unperceived modifications within a virtual scene. Mise-Unseen [240] by Marwecki et al. is an approach that injects scene changes imperceptibly into the users' field of view. Eye-tracking data were used to assess the user's attention, intent, and spatial memory, which were employed to adjust the injection time. The authors demonstrated that the use of gaze data in combination with masking techniques enables the insertion of modifications without being perceived by the user.

a) Attention-based Virtual Space Design: Analysis of visual attention is also useful during the design phase of virtual spaces. By analyzing and understanding the visual attention of users, 3D virtual spaces could be designed in a way that supports the primary task objective. Alghofaili et al. [241] presented an attention-based approach to optimize the placement of visual elements and design a 3D environment taking predefined goals into account. The authors employed a regression model to predict gaze duration and then showed the efficiency of the optimization, combining cost functions for regularization, number of elements, and primary design objective related to the visual elements that wanted to be placed. Hillaire et al. [242] introduced an algorithm for exploring virtual scenes by simulating human visual attention. The proposed algorithm, which can also be applied during first-person exploration that includes walking and turning movements, uses a surface element-based representation instead of a mesh-based representation. The authors stated that the proposed algorithm and model can be used to optimize the quality of the scene to accelerate the rendering process.

b) Joint Attention: With VR, there is an additional ease to assess the joint attention of users as with each HMD and user, it is possible to get precise gaze and head orientations in the same virtual space. As the information on visual attention in VR also promotes interaction and collaboration, joint attention has been a focal point for some. For instance, Špakov et al. [243] analyzed the effect of eye-gaze- and head-gaze-based visual attention sharing in cooperative games. In such VR games, sharing eye-gaze led to a higher level of subjective ratings of teamwork and shorter game duration compared to sharing head-gaze. Similarly, Kasahara et al. [244] evaluated the effect of the shared egocentric videos with gaze locations in perspective sharing. The behavior changes and decisions of the participants with experience sharing are analyzed while sharing parallel views. During the drawing activity, the researchers observed that individuals can develop behaviors to complement their partner's memory and decision. Furthermore, even in complex cases, individuals can create mechanisms to comprehend their physical embodiment and spatial relationship with their pairs.

In the realm of medicine, impairments in joint attention can serve as an early indicator of ASD. Training and enhancing the joint attention mechanisms are essential in interventions for individuals for instance with ASD. Mei et al. [245] attempted to address skills training for the joint attention of ASD patients using customizable virtual humans (CVH). The researchers designed an educational drum-playing scenario in an eye-tracked virtual environment for 10 ASD patients and measured visual attention with the help of gaze data. Their user study showed that CVH improved participants' visual focus on relevant regions on the scene at the cost of an increase in reaction time.

c) Visual Attention in Virtual Learning Spaces: Eye tracking has also found its applications in the realm of pedagogy, particularly in virtual learning spaces. Especially with the impact of the COVID-19 pandemic, remote and immersive learning spaces have become popular, especially in the form of virtual classrooms. In such simulations and setups, the visual attention of students and teachers plays an important role in the learning processes. To facilitate this, Ahuja et al. [246] proposed a novel 3D digital twin classroom simulation that could be accessed through both VR headsets and web interface and incorporated a computer vision-based head and eye tracking system. Two cameras were deployed for the system to track students and teachers, respectively, and gaze data were visualized in diverse forms, such as students' saliency map on the board and intersections of instructor's gaze on student planes. A controlled study showed that this novel non-HMD classroom gaze system almost halved the gaze prediction error of its predecessors, whereas a reliability of 92.54% was reported in an in-the-wild assessment. Focusing more on the virtual simulations, Gao et al. [247] utilized several gaze features such as fixation durations, saccade durations, and pupil diameters to investigate the influence of various classroom manipulations on learners in a VR classroom, including different sitting positions, virtual avatar styles, and peer hand-raising behaviors. The authors found that such manipulations affect user attention and cognition differently, which could lead to different learning and engagement outcomes in the long term. The same manipulations were also studied by Bozkir et al. [248] in an eye-tracked VR classroom. In this work, the authors focused on virtual objects and attention times on the objects related to learning and engagement in the classroom. The authors showed that learners sitting in the front part of the virtual classroom paid more attention to the teacher and board, whereas others who sit in the back visually focused more on their peers. They also discovered that cartoon- and realistic-styled avatars grabbed different amounts of attention during the classroom discourse. Later, Gao et al. [249] explored various gaze features in a virtual lecture to scrutinize the effects of social interactions. The authors considered peer-learner hand-raisings as a cue of social interaction in their virtual classroom and found that such hand-raising animations and the number of peer-learners that raise their hands affect students' attention in different ways. Attention and cognition analyses toward such cues in virtual learning spaces are especially important as student self-concepts could be influenced positively or negatively with constantly the same amount of peers raising their hands. Using the same VR classroom space, Gao et al. [250] also showed that boys and girls visually behave

TABLE VII
OVERVIEW OF THE PAPERS RELATED TO HUMAN VISUAL ATTENTION.

Paper	# Users	Purpose	Eye-tracking Features	Visual Stimuli
[233]	25	Attention analysis in panaromic videos	gaze	Cues
[234]	20	Attention guidance in sandwich preparation	gaze	Cues
[235]	15	Attention guidance in unordered tasks	gaze	Dynamic cues
[236]	20	Attention guidance	gaze	Swarm motion
[237]	16	Attention guidance in driving simulation	pup	Cues
[238]	40	Attention analysis	gaze	Notifications
[239]	20	Attention analysis	gaze	Notifications
[240]	15	Attention analysis for scene changes	gaze	Scene changes
[241]	23	Attention analysis for environment design	gaze	N/A
[242]	12	Attention analysis for environment design	gaze	N/A
[243]	40	Joint attention analysis in collaboration	gaze	Shared gaze
[244]	40	Joint attention analysis in collaboration	gaze	Shared gaze
[245]	10	Joint attention training for patients with ASD	gaze	N/A
[246]	13	Attention analysis in education	saliency, gaze	N/A
[247]	288	Attention analysis in education	fix, sac, pup	N/A
[248]	280	Attention analysis in education	gaze	N/A
[249]	280	Attention analysis in education	pup, fix, sac, gaze	N/A
[250]	280	Gender analysis in education	pup, fix, sac, gaze	N/A
[251]	26	Gaze visualization in education	gaze	Visualized gaze

differently with their eye gaze in the classroom discourse and indicated that utilization of this information could be of help for personalized learning support. Their explainable machine learning model resulted in a gender classification accuracy of over 70%. Complementary to the aforementioned works, Rahman et al. [251] focused on the perspective of the teacher during a lecture in VR and proposed six different gaze visualization techniques from the teachers' view to identify distracted students. These visualization techniques covered gaze ring, gaze disk, gaze arrow, gaze trail, gaze trail with arrows, and gaze heatmap. The user study results showed that gaze trail visualization outperformed the others in terms of popularity and the application of 3D gaze heatmaps was found to be problematic. Similar to papers about human cognition, we summarize the papers that are related to visual attention in Table VII by also including the stimulus types associated with each paper.

3) Perception: Human visual attention and cognition provide a lot of information on how humans experience virtual spaces; however, the research in these fields does not directly answer the question of how humans perceive these spaces. Considering perception engineering and its combining aspect of engineering and human physiology [252], many researchers have focused on the perception aspect of VR using eye movements as an important source of information. While some of the findings on human perception overlap with the findings on cognition and visual attention, it is important to address the findings individually and tackle the issues altogether to design more immersive and usable virtual spaces.

Depth perception is a well-studied topic when human perception is considered in the VR domain. For instance, Katzakis et al. [253] investigated human depth perception during a 3D painting task in both foveal and peripheral regions. The study's findings revealed that the accuracy of the task was dependent on the location of the target rather than the starting point, and highly accurate 3D painting tasks can be accomplished in virtual spaces when targets are within arm-reachable range. Nonetheless, it was observed that overestimation of depth occurred in all cases, but it was more significant in the cases starting from the periphery and terminating in the foveal region. In another work, Arefin et al. [254] revealed the relationship of perceptual depth level with eye vergence angle and interpupillary distance in VR. The authors stated that changes in perceptual depth could be inferred by using these features.

Research on human perception also goes beyond depth perception. Serrano et al. [255] systematically investigated the human perception of continuity in VR movies, which have different cinematographic requirements from conventional 2D movies. The authors used gaze information and eye-tracking data along with the videos and revealed new metrics such as scanpath errors describing visual attention and offered design insights for VR content creation. Additionally, MacQuarrie and Steed [256] examined HMD users' perception capacity to comprehend the eye-gaze direction of virtual volumetric characters in varying display resolutions, virtual character positions, head rotations, and gaze directions. The authors also indicated that perceptual accuracy is considerably position-dependent and influenced by the direction of view.

a) Understanding Perceptual Limits: Identifying the perceptual limits of users is essential for various VR applications that consist of reorientation and repositioning, which often require discrete or continuous scene changes. However, it is essential to insert the scene changes imperceptibly during the VR experience. Optimizing the amount and timing of these changes according to users' perceptual limits ensures that users' virtual experience is not disturbed by the inserted changes. Researchers have examined the perceptual limits and abilities of VR users to have a detailed understanding of this and provide them with tailored VR solutions. For example, Shin et al. [257] examined the effect of unobtrusive and prompt virtual object movements on users' body posture. The authors provided design insights while also assessing users' perceptual limits for unnoticeable position changes of the virtual objects during eye blinks. Langbehn et al. [258] investigated conscious or unconscious blinks that suppress human visual perception which can be exploited to design motion algorithms in VR. The authors examined the upper limits of unnoticed changes during eye blinks and demonstrated that users could not perceive translational changes of 4-9 cm and rotations of $2° - 5°$. Furthermore, they also stated that a 50% improvement can be observed in curvature gain, which is defined as $\frac{1}{r}$ where r corresponds to the radius of a circular path in the real world while moving in a straight line in VR.

In addition to the blinks, Keyvanara and Allison [259] evaluated the human limits during saccadic behaviors and what effects and manipulations are unperceivable during saccadic suppressions that occur during constant horizontal and vertical camera movements by utilizing a Bayesian procedure. The authors reported that additionally applied sudden horizontal camera movements to the continuously moving virtual camera on the vertical axis are less observable. In another work, Keyvanara and Allison [260] examined the visual sensitivity of VR users against changes in a 3D scene during saccadic suppression. The findings of the authors revealed that certain image transformations, such as rotations along the roll axis, are more noticeable by users during horizontal saccades. Similarly, Bolte and Lappe [261] proposed the use of saccadic suppression to include rotations and translations into the scene. The authors modified two eye movement classification algorithms, which were originally proposed by Behrens et al. [262] and Niemenlehto [263] to meet the time and reliability requirements of online saccade detection. They investigated the limitations of human perception during detected saccadic periods and found that users were unable to perceive positional changes up to 50 cm in the direction of their eye gaze and rotational changes up to $5°$.

b) Perceived Realism of Virtual Spaces and Perceptual Considerations for Virtual vs Real Worlds: Sense of realism is related to the visual characteristics of the 3D scenes we explore, the ecological validity of the simulations, and how we perceive those 3D spaces physiologically. As human visual attention drives eyes and we perceive the scenes accordingly, eye tracking-based analysis paves the way for the understanding and estimation of the realism level of virtual environments. This also allows the possibility of developing techniques for enhancing the perceived realism of VR environments taking the perceptual limits of the human visual system.

To this end, the effects of light sources on human vision, such as temporal eye adaptation, perceptual glare, visual acuity reduction, and scotopic color vision, can be used to enhance the realism level in virtual scenes. Luidolt et al. [264] utilized gaze direction and pupil size as means of lighting effect adjustment in order to produce more realistic low-light scenes that are tailored to the users' vision. The authors also revealed that the effect of the light in virtual scenes is highly subjective. Perception and reaction differences between virtual- and real-world tasks also provide insights to create more realistic virtual environments. In this respect, eye-gaze activity is a crucial indicator for understanding human behavior in real and virtual worlds [265], [266]. To achieve the goal of designing realistic crowd simulations, Berton et al. [265] investigated gaze behavior during crowd walking activity. The differences in gazing activity in real and virtual simulations were analyzed and virtual walking simulations were performed at different crowd levels. As the crowd level increases, participants' eye movements become more restricted, scanning a narrower portion of the street. Furthermore, the authors noted that users tend to direct their attention to the individuals who are in front of them. They also provided some design recommendations, such as using a constant number of neighboring agents and positioning them by considering collision risk in virtual crowd simulations. In a different study, Berton et al. [266] exploited gaze data to compare collision avoidance behaviors in the real world and several VR environments. Their findings revealed that collision avoidance behaviors are similar across different settings. A work investigating the existence of the stare-in-the-crowd effect, which refers to the tendency of individuals to detect and observe the gaze directed to them, was designed by Raimbaud et al. [267]. The occurrence of the stare-in-the-crowd effect was revealed through the identification of differences in gaze characteristics. Additionally, the authors observed a negative correlation between the dwell time associated with this effect and social anxiety scores while providing insights to achieve realistic interactive environments with virtual avatars. Another research on the similarities and differences between the real and virtual worlds was conducted by Gupta et al. [268] in which the authors examined the autobiographical memory (AM) that is often considered as one of the vital factors in human perception. The effect of autobiographical memory was revealed on diverse physiological cues like eye gaze, pupil diameter, and electrodermal activity (EDA). The authors presented evidence of the potential efficacy of the features in assessing autobiographical memory within virtual environments.

Avatars are essential components of virtual spaces and reality due to being a proxy for social interactions within virtual environments and they also have different impacts on perceived realism. To this end, eye-tracking methods to generate realistic avatars raised some attention from the research community. For example, Borland et al. [269] augmented a virtual self-avatar with eye movement animations to achieve a more comprehensive virtual embodiment and improve self-recognition. The authors compared representative animations of real eye movements with simulated eye movements without requiring eye-tracking hardware and reported that the use of eye movement animations results in an increased subjective sense of self-identification. Bergström et al. [270] evaluated the plausibility of virtual musical performance, with changing environmental conditions and gaze-based attributes of audiences and musicians. It was found that the gaze behavior of the virtual agents and distractions that do not comply with the nature of the environment had substantial effects. More recently, Ma and Pan [271] presented a technical framework to generate self-avatars with facial expressions and assessed the psychological effects of realistic avatars. In their study, participants found the facial expressions of cartoon-like avatars to be more controllable than realistic-looking avatars. Additionally, the authors observed that participants had a higher sense of body ownership in the first trial, regardless of the type of avatar used.

Eyes, faces, and heads are important components of virtual avatars and a handful of works focused specifically on the simulations of these components. For instance, Le et al. [272] introduced a method to automatically simulate head movement, eye gaze, and eyelid movement based on speech input. Gaussian Mixture Models and Nonlinear Dynamic Canonical Correlation Analysis were employed to simulate head movement and eye gaze behavior, respectively. Non-negative linear regression was applied for intentional blinks, while unconscious blinks were generated using a log-normal distribution. The authors reported superior performance compared to state-of-the-art algorithms in the generation of head and eye motions. Thies et al. [273] introduced FaceVR, an image-based face synthesis approach that includes an eye-tracking algorithm based on monocular videos, aiming to render more realistic outputs on stereo displays. Ladwig et al. [274] proposed a low-cost solution to construct the face of a person wearing VR glasses utilizing GANs to produce face images. Additionally, the authors created an eye-tracking method that gives cues for iris position and gaze direction. Similarly, Song et al. [275] presented a CNN-based 3D face-eye reconstruction technique to construct a 3D personalized avatar with eye movements. The proposed algorithm is able to construct personalized avatars with facial expressions and eye animations of the VR user. Another CNN-based technique introduced by Olszewski et al. [276] supports speech animation, as well as emotional expressions using mouth and eye-tracking cameras for HMDs. The proposed method outperforms existing state-of-art techniques in terms of the fidelity of animations without requiring individualized calibration. Zhao et al. [277] proposed a face reconstruction algorithm using a personalized 3D head model along with a colorization algorithm for near-infrared eye images without causing the red-eye effect, which is defined as color distortion in the iris of the eyes.

c) Cybersickness: While virtual spaces and setups provide users with a lot of exciting opportunities, the use of VR HMDs may lead to several types of physical discomfort, which are often called VR sickness or cybersickness and these are closely related to human perception. Oftentimes, eye tracking has been used to understand the causes and severity of cybersickness as well as to develop methods to mitigate or prevent cybersickness. For example, Islam et al. [278] proposed a cybersickness severity prediction algorithm utilizing built-in HMD sensors including eye and head trackers. This algorithm relies on deep fusion network models and it was evaluated on the data collected from 30 participants during a VR video game. The authors demonstrated the capability of their models by incorporating eye- and head-tracking features with video stimuli and they achieved an accuracy of up to 87.77%. In another work by Islam et al. [279], the authors introduced a deep fusion network based on Deep Temporal Convolutional Networks (DeepTCN) fusing physiological, head-tracking, and eye-tracking data. DeepTCN is able to predict cybersickness 60 seconds in advance with a 0.49 mean-squared error (MSE) on a scale between 0-10. Among different types of fused data, eye-tracking data incorporated with heart rate and galvanic skin response achieved the best performance. Similarly, Lee et al. [280] employed eye movement features to predict cybersickness severity along with disparity and optical flow maps that respectively represent depth and movement in the image. Their 3D CNN model achieved better precision rates with the inclusion of eye-tracking features compared to Padmanaban et al. [281].

Cybersickness can occur in all types of virtual experiences; however, virtual experiences that require motion are more likely to cause this problem due to the mismatch between the movements in real and virtual worlds. For instance, virtual locomotion, which allows the exploration of virtual environments, is one of the leading causes of VR sicknesses like nausea and dizziness due to one-sided movement in the virtual world. Zayer et al. [282] provided a survey of virtual locomotion techniques that enhance the overall VR experience by mitigating motion sickness under the following categories; walking-based, steering-based, selection-based, and manipulation-based methods. The authors evaluated the strengths and weaknesses of these techniques to provide comprehensive guidance for the community. It is stated that motion sickness occurring during locomotion is usually caused by visual-vestibular conflict and field of view (FOV) constraints are deployed to mitigate VR discomfort but this also restricts users' FOV. Adhanom et al. [283] presented a gaze position-based foveated FOV restriction method to improve existing techniques that rely on head gaze to mitigate cybersickness. These existing techniques are limited to cases where the head and eye gaze are aligned. The authors showed that their proposed method provides users with the flexibility to experience a wider visual scan field while preserving the same level of VR sickness and noticeability.

d) Perceptual Expertise Level: As people at different levels of expertise and skill tend to have different visual perception patterns, the applications of eye-tracked VR have been developed with leaps and bounds in the realm of expertise training and assessment. Driving training is particularly of interest among all VR training applications. For instance, Lang et al. [284] studied the improper driving habits of 50 users in a driving simulation with eye-tracking-enabled VR headsets and helped them improve their driving skills by designing synthesized personalized training routes considering their perceptual habits. Their user study conveyed that the 10 participants trained in a customized VR setup outperformed the others who were trained by conventional methods on average with respect to response time in emergency situations, training persistence, and an evaluation score based on inappropriate driving actions (e.g., not signaling before a turn).

Expertise assessment and its applications also go beyond driving training. For example, Hosp et al. [285] studied the gaze behavior of 35 football goalkeepers in VR when 360° videos captured from goalkeepers' viewpoint were displayed to them. By analyzing their eye movements, the authors were able to classify them into three levels of expertise: elite youth player, regional league player, and novice player, with an accuracy of up to 78%. The possibility of identification of differences in the gazing behaviors of humans with different expertise levels can help train novices by providing them with visual cues that are similar to the visual behaviors of experts.

In another domain, Orlosky et al. [286] applied eye tracking to measure and classify the English language understanding of 16 users in VR using features like pupil diameters and eye movements. The authors reached a prediction accuracy of 75% for words that are considered easy and medium with SVMs. When they extended their analyses to include words that are categorized as hard, the authors were able to obtain a prediction accuracy of up to 62%.

The applications in the industry are also important and especially these applications have the potential to be used in everyday life. To this end, Burova et al. [287] developed an elevator maintenance simulation in eye-tracked VR to facilitate industrial AR prototypes, which are challenging to build due to safety concerns in the real world. The authors recorded several types of behavioral data from users including eye behaviors and enabled gaze data visualization in the form of scanpaths and heatmaps during training playback in VR. Based on their survey including 12 elevator maintenance experts, the authors indicated that the domain experts hold constructive opinions regarding the utilization of eye tracking and gaze visualizations for industrial training. Additionally, Gisler et al. [288] examined the relationship between training success and human behaviors in sanitary apprenticeship tasks. The authors combined gaze positions, head movements, and attention durations on the focused objects by using statistical summaries to predict the training success of users. Experiments including 48 sanitary apprentices in an industrial training task achieved a 10-20% improvement in predicting users' training success compared to a baseline model. In addition, skill training for searching and tracking in virtual police rooms has been examined by Harris et al. [289]. They designed virtual experiments in which officers are supposed to search and strive to gather evidence for an investigation while also taking various criminal activities into account. The perceptual-cognitive skills of 54 participants were analyzed in multiple training scenarios by taking saccadic eye movements (i.e., saccade sizes) into account. The results showed that visual search expertise can successfully be trained in VR.

e) Vision Impairment, Ocular Examination, and Medical Perception: Eye tracking and VR have also been utilized to study vision impairment and ocular examination, as people with vision problems perceive the world differently. Just like other subfields discussed in this paper, in VR, it is possible to have very controlled programmed experiments compared to in-the-wild settings to study such phenomena, which is why researchers have attempted to design ecologically valid simulations and study vision impairment issues in VR. For example, to better understand the perception of people with cataracts, Krösl et al. [290] successfully simulated cataract patients' vision in a virtual environment. As cataracts lead to special visual effects at the lens center and periphery, the authors used eye tracking to account for these gaze-dependent symptoms. More related to ocular examination using eye tracking in VR, Kim et al. [291] transplanted the Developmental Eye Movement test to VR HMDs, which is a clinical eye test widely used to determine abnormalities in visual function and to assess ocular motor skills. In subsequent research, Kim et al. [292] designed the King-Devick test, which is a standard measurement for the assessment of saccadic eye movement and dynamic visual acuity, using HMDs both in VR and AR. Hotta et al. [293] also designed a gaze-based ocular examination for visual field defects in a virtual environment using fixation and saccadic features. Compared to conventional tests that take over 30 minutes, the proposed method can be conducted in 5 minutes while retaining sufficient accuracy and yet improving reliability. In addition to the aforementioned studies, medical VR applications can actually go beyond the setups that require HMDs. For instance, Kübler et al. [294] analyzed the eye and head movements of patients with homonymous visual field defects (HVFD) in a virtual driving simulation. The virtual environment was used with a VR cabin, while participants' eye movements were tracked by a head-mounted eye tracker. The findings of the authors confirmed their hypothesis that a certain share of HVFD patients can strengthen their viewing behaviors with an increased amount of visual scanning to improve their driving skills. Similar to earlier sections, we present an overview of the works that are related to human perception in Table VIII.

V. Security and Privacy in Eye Tracking and Implications for Virtual Reality

As it is possible to extract and infer a lot of useful information about users utilizing eye trackers in VR, eye tracking is considered a powerful sensing modality based on the aforementioned studies. In those studies, we mostly considered works that utilize or are directly applicable to VR setups. However, virtual environments are relatively new for people and possibilities with eye movements are not evaluated extensively compared to real-world settings. Despite this, it is known that human eye movements depend on the context and stimulus and we first argue that as long as similar stimuli are presented to the users in virtual environments, similar inference possibilities will be possible in these environments like in the real world. Liebling and Preibusch [3] stated previously that apart from person identification, it is possible to link eye movements with many attributes such as gender, sexual preferences, body mass index, health status, or tasks mostly in real-world settings. These would presumably apply to virtual settings as well, as some of them such as gender and task predictions can already be carried out accurately in virtual settings [24], [295], [250], which is in line with the literature and we foresee that many other inference possibilities will follow going forward when such virtual environments and setups are more prevalent in daily life. To this end, Xu et al. [296] stated that while the existence of eye-tracking data is vital for utility tasks such as improving the efficiency of VR rendering, it also gives opportunities, especially for companies for targeted advertising based on such unique characteristics of users, which may be thought as privacy intrusion. Considering all of these, as identification of aforementioned attributes (e.g., personal identification) might be of help for users for authentication and personalization purposes while their use of VR setups, we first discuss these possibilities in Section V-A. Then in Section V-B, we mainly discuss the methodological works that provide privacy-preserving eye tracking in VR.

TABLE VIII
OVERVIEW OF THE PAPERS RELATED TO HUMAN PERCEPTION.

Paper	# Users	Purpose	Eye-tracking Features
[253]	18	Depth perception analysis in 3D painting	gaze
[254]	24	Depth perception analysis	ipd, eva
[255]	49	Continuity analysis in VR movies	gaze
[256]	37	Perceptual limit analysis	gaze
[257]	27	Perceptual limit analysis	blnk
[258]	32	Perceptual limit analysis for redirected walking	blnk
[259]	36	Perceptual limit analysis	sac
[260]	10	Perceptual limit analysis	sac
[261]	13	Perceptual limit analysis	sac
[264]	5	Perceptual realism analysis in low-light 3D scenes	gaze, pup
[265]	21	Perception analysis in crowd walking	gaze, fix
[266]	17	Perception analysis in collision avoidance task	gaze, fix
[267]	30	Perception analysis in psychology	fix
[268]	20	Perception analysis	gaze, pup, blink
[269]	12	Perceived realism analysis for virtual avatars	gaze
[270]	20	Perceived realism analysis	gaze
[271]	18	Perceived realism analysis for virtual avatars	gaze
[272]	20	Perceived realism analysis for virtual avatars	gaze
[273]	18	Face & eye synthesis for virtual avatars	N/A
[274]	5	Face & eye synthesis for virtual avatars	N/A
[275]	7	Face & eye synthesis for virtual avatars	N/A
[276]	310	Face & eye synthesis for virtual avatars	N/A
[277]	3	Face & eye synthesis for virtual avatars	N/A
[278]	30	VR sickness prediction	gaze, pup, cvgn
[279]	30	VR sickness prediction	gaze, pup, cvgn, blnk
[282]	N/A	VR sickness mitigation	N/A
[280]	96	VR sickness prediction	saliency
[283]	22	VR sickness mitigation	gaze
[284]	50	Skill assessment and training for driving	gaze
[285]	35	Skill assessment and training in soccer playing	gaze, sac, fix, spur
[286]	16	Skill assessment and training in education	pup, sac, fix
[287]	12	Skill assessment and training for elevator maintenance	fix, scanpath, saliency
[288]	48	Skill assessment and training for sanitary apprentice	gaze
[289]	54	Skill assessment and training for police room search task	gaze, sac, fix, entropy
[290]	21	Vision simulation for patients with cataract	gaze, pup
[291]	39	Visual acuity assessment for ocular examination	eye dominance, ipd, cvgn, phr
[292]	30	Visual acuity assessment for ocular examination	sac
[293]	2	Visual field analysis for ocular examination	sac, fix
[294]	14	Visual field analysis in driving	sac, fix, gaze

A. Eye-based Authentication for VR

Visual interaction and user assistance are two of the most prominent uses of behavioral data including eye movements, especially in VR applications that work in real-time. Apart from those uses and related applications, such behavioral data have been utilized for authentication purposes as well. One big challenge for behavioral authentication is that accuracies of the methods are not very high which can cause an unpleasant user experience. For instance, Pfeuffer et al. [297] studied authentication in VR by utilizing body motion data including head, hand, and eye movement data for different tasks such as pointing, grabbing, walking, and typing. The authors found that while the best-achieved accuracies for identification are in the vicinity of 60% and they drop when user group size is increased. The authors also showed that body motions can be of help for adaptive settings in VR. However, they did not use any gaze features apart from the gaze ray obtained from the eye tracker. In another study, Liebers et al. [298] studied user authentication based on spatial movements in VR and reached accuracies up to 90%. In addition, in another study, Liebers et al. [299] also showed that gazing behaviors along with head orientation can authenticate users in VR with an almost perfect level of accuracy. However, the limitation of those studies is the small sample size and it is an open question of how utilized features perform when a wide range of users are available and how usable the proposed authentication methods would be. That is why when authentication is considered, more distinguishing and fine-grained features are needed. This could be achieved by using eye-tracking data obtained from high-frequency eye trackers that are available nowadays, and that will potentially be available in the near future.

When eye-based authentication is taken into account, there are two prominent directions: iris-based and eye movement-based authentication. On the one hand, while iris textures are like visual fingerprints and help authenticate users with very high accuracies [300], [301] and iris-based authentication has also been used in real-world scenarios such as in airports and border-crossings [302], especially due to privacy reasons, many of the commercial HMDs that integrate eye trackers do not provide raw eye images to the users or applications (e.g., HTC Vive Pro Eye and Microsoft HoloLens 2). This can mitigate the problem to some extent; however, in the case of iris-textures utilized in VR, data protection issues should be handled carefully possibly by using encryption schemes [301]. On the other hand, even if the iris textures are not captured or not accessible with these devices, eye movements will still be output by the eye trackers and it is also plausible to carry out authentication using eye movements in the background over a certain period of time depending on the eye tracker frequencies and granularity of extracted eye movement features to perform this task. While we do not carry out an extensive review for eye-based authentication, we provide essential studies that are related to implicit biometric authentication especially utilizing eye movements in VR, and discuss why they are related to privacy aspects of such data. For more detailed information on eye-based authentication and authentication with different modalities including eye tracking, we refer the reader to the survey papers by Katsini et al. [26] and Stephenson et al. [303], respectively.

While not necessarily in VR, biometric identification was carried out successfully using oculomotor plant models and eye movements [304], [305]. Task-independent authentication is also possible using eye movements [306]. More related to VR, Eberz et al. [307] argued that specific eye movement features could be used for biometric authentication in cheap consumer-level devices during everyday tasks such as reading, writing, and web browsing. While the authors did not use a VR setup, their sampling rate was 50 Hz, which is even lower than the sampling rates of the eye trackers of today's consumer-grade HMDs. In follow-up work, Eberz et al. [308] presented a continuous authentication system based on eye movement biometrics and stated that for eye movement-based authentication, a precise calibration should be done and effects of light sensitivity and task dependence of eye movements should be considered when designing the authentication systems. In another work, Zhang et al. [309] proposed a continuous authentication scheme based on eye movements for VR headsets and showed that they could continuously authenticate the wearer of the HMD in the background, which shows the potential for personalized use of HMDs. Zhu et al. [310] used blinking patterns and pupil sizes for user authentication in VR headsets. While all of these works show the plausibility of biometric authentication based on eye movements, most of them suffer from issues in terms of usability or privacy in the context of VR, which is similar to behavioral authentication using spatial movements. From a usability point of view, authentication models based on eye and gaze movements do not work with as high accuracies as iris authentication, which could irritate HMD users if errors occur constantly. Furthermore, such models mostly depend on the temporal movement of the gaze regardless of explicit or implicit authentication, which leads to longer authentication times compared to using single iris images. Similar implications were also stressed and partly confirmed by the work of Lohr and Komogortsev [311] that biometric authentication performance using eye movements is comparable to 4-digit PIN entry when eye tracking signal quality is extremely high along with very high sampling frequencies (i.e., 1kHz). The authors stated that 5 seconds of eye movement data are needed to achieve such performance. Niitsu and Nakayama [312] also measured measurement time and presentation size on biometric authentication and they found that using 3 seconds maximized the biometric authentication success rates. While they did not use a VR setup for this, the results could be transferred to VR as well, as long as the quality of the eye-tracking data and presentation sizes are comparable. Despite these attempts at biometric authentication based on eye movements, Friedman et al. [313] recently showed it is not possible to identify users very accurately based on their eye movements if the considered number of identities is comparably large, which is in line with findings from research using behavioral data for authentication [297], [298].

These findings imply that while implicit eye-based authentication might not work similarly to PIN-based authentication in terms of usability and security, eye movements include patterns that help identify individuals to a certain extent. These data may thus be more useful for personalization where each user's data are accumulated over time during their use of VR HMDs and utilized for authentication-related settings such as Two-Factor Authentication (2FA). Since such an authentication process should run continuously in the background, longer waiting times for authenticating would not irritate the users as they would not be actively aware of the ongoing process. In addition, this way of authentication will potentially make the overall VR experience more personalized as each user's unique viewing behaviors would be captured by their devices over time. Several works could support this argumentation. For instance, Lohr et al. [314] proposed a real-time capable architecture for eye movement-based authentication in VR and argued that eye-tracking-based biometrics will become a standard way of authentication for VR. Luo et al. [315] explored the human visual system (HVS) as a novel authentication method for VR HMDs and proposed OcuLock which is resistant against impersonation and statistical attacks while maintaining a stable performance over a 2-month period. Furthermore, Friedman and Komogortsev [316] assessed the effectiveness of biometric feature normalization techniques, including real-world eye movement features along with synthetic ones. They found that the effectiveness of different biometric normalization techniques on real-world data depends on the inter-correlation of the features. Mathis et al. [317] used VR to evaluate the usability and security of real-world authentication systems, showing the potential of VR as a testbed for authentication purposes.

Despite several disadvantages such as the need for consistent tasks, high-frequency and high-quality eye-movement data, and a decrease in authentication performance when the number of involved individuals is increased, previous work consistently showed that eye movement patterns are representative of personal identities. This is useful if the goal is to authenticate the user in VR or personalize the virtual environments. However, if users do not want to get the advantage of these possibilities and yet want to use gaze-based interactions and obtain assistive support, the linkage of eye movements to user identity and characteristics implies privacy risks that should be handled in a methodological way. This means that personal patterns that are associated with eye movements should be either hidden in the data or handled in a privacy-preserving manner, while the interaction utility and user experience during the virtual experience should be kept high.

B. Privacy-preserving Eye Tracking for VR

Research in eye-based authentication [26], privacy considerations for eye tracking [3], [22], and privacy risks of data collection in XR [318, pp. 12-14] demonstrate that when authentication is not preferred, it is advisable to preserve the privacy of the individuals and specific attributes of the individuals in the data, in addition to the good data hygiene practices. Recently, what inferences could be carried out by using the pupil dilation and gaze in the context of XR is also highlighted by Future of Privacy Forum [319], by mentioning concerns on users' sexual orientation, gender, race, and health. In addition, Garrido et al. [320] highlighted data privacy issues in VR considering threat and defense models and included eye-tracking sensors and data as discussion points.

To provide users with privacy-preserving eye-tracking solutions, researchers have explored a few different approaches applied at various stages of the eye-tracking pipeline. Privacy-preserving methods have been applied to sensor data in the form of eye images, aggregated data in the form of saliency maps and feature vectors, and sample-level gaze position data. Each method transforms or modifies data in a variety of ways to create a trade-off between data privacy and utility that varies by method and intended application of the data.

Recently, formal methods such as differential privacy [321] have attracted the attention of VR and eye-tracking communities. Differential privacy is indeed an established approach that has been applied across multiple domains, most prominently in the context of databases and aggregate survey data. The overall aim of differential privacy is to hide the information on whether an individual was included in a particular database or not. This is achieved by adding randomly generated noise according to an ϵ parameter to query outputs such that the inclusion of a specific participant does not change the queried function outputs significantly. Since several works focus on differential privacy in the context of eye tracking, we provide the formal definition of the ϵ-differential privacy as follows.

Definition 1: ϵ-Differential Privacy (ϵ-DP) [321], [322]. A mechanism M is considered ϵ-differentially private for all databases D_1 and D_2 that differ at most in one element for all $S \subseteq Range(M)$ with;

$$\Pr[M(D_1) \in S] \leq e^{\epsilon} \Pr[M(D_2) \in S]. \quad (1)$$

The eye-tracking community initially utilized standard differential privacy mechanisms, namely Gaussian and Exponential mechanisms, on saliency heatmaps [323] and on aggregated eye movement features that were collected from VR reading tasks [24]. These initial works showed the potential of formal methods for the VR domain; however, they suffer from the correlations in the data that could jeopardize privacy, especially from temporal correlations as independent noise sampling from the standard mechanisms allows adversaries to reconstruct signals that are very close to the original ones. Bozkir et al. [295] further iterated these works and addressed the temporal correlation issue by translating the data into difference signals and using the frequency domain to add the randomly generated noise. The authors noted that apart from the correlation challenge, privacy-utility trade-off becomes very important as eye-tracking data already consists of a certain amount of noise due to the limitations of sensors and image processing approaches. Furthermore, the privacy noise needed to establish a strong differential privacy guarantee is known to hurt utility across other domains and ends up masking valuable insights from the processed data [324].

Based on these prior results, researchers have also pursued alternative privacy guarantees to differential privacy to better understand the range of privacy-utility trade-offs for eye-tracking data. Specifically, David-John et al. explored k-anonymity and k, γ-plausible deniability [325], [326]. These alternative guarantees specifically target the privacy risk of re-identification, and their mathematical formulation is related to the ability to match released data to the original identity in a dataset. Alternative guarantees allow dataset owners to preserve privacy by reducing the risk of linking data to identities and thus mitigating harms related to the inferences that can be made from eye tracking, while also retaining data utility across different applications.

First, k-anonymity is a definition of privacy in the context of re-identification from a dataset and it was proposed by Samarati and Sweeney [327] as follows.

Definition 2: k-anonymity. Given a person-specific dataset D_1, a de-identified dataset D_2 is k-anonymized by privacy process $\mathcal{P} : D_1 \mapsto D_2$ if all released features $\Gamma_d = \mathcal{P}(\Gamma) \in D_2$ cannot be recognized as Γ with probability $> \frac{1}{k}$.

The privacy guarantee is interpreted with a lone privacy parameter k linked to the upper bound on re-identification. Utilizing k-anonymity on eye-tracking feature data, David-John et al. [325] showed that person re-identification accuracies drop to chance levels while the utility of a model trained on privacy-preserving data is kept within reasonable ranges when the document-type classification was the utility task. However, the methods used to achieve k-anonymity typically depend on the duplication or generalization of data which is not ideal when releasing datasets for research or statistical purposes. This led David-John et al. [326], [325] to introduce the guarantee of k, γ-plausible deniability to eye-tracking data as well. Plausible deniability differs from k-anonymity in that it specifically applies to synthetic data generated by a model. All synthetic data generated by the model is tested to ensure that it meets the following privacy guarantee prior to release.

Definition 3: Plausible Deniability. For any dataset D where $|D| \geq k$, and any record y generated by a probabilistic generative model $\mathbf{M}$ such that $y = \mathbf{M}(d_1)$ for $d_i \in D$, it is stated that y is releasable with (k,γ)-plausible deniability if there exist at least $k-1$ unique records $d_2, ..., d_k \in D \setminus \{d_1\}$, such that

$$\gamma^{-1} \leq \frac{Pr\{y = \mathbf{M}(d_i)\}}{Pr\{y = \mathbf{M}(d_j)\}} \leq \gamma$$

where $k \geq 1$ is an integer and $\gamma \geq 1$ is a real number.

When considering datasets of gaze samples, the researchers found that different formal methods produced practical privacy-utility trade-offs across different applications. Namely, practical trade-offs were achieved for the utility task of activity type recognition for both plausible deniability and a sample-based differential privacy method; while for a gaze prediction task only k-anonymity produced a dataset that had minimal errors when training a deep neural network model. The takeaway from these works is that the recommended method often depends on the target application and whether privacy protections require differential privacy or guarantees specific only to re-identification.

In another work, Li et al. [328] adapted a differential privacy approach for providing real-time privacy control of ϵ for eye-tracking data based on location indistinguishability [329]. The key difference between this approach and the other methods described earlier is the inclusion of spatial privacy parameter r. The authors proposed tuning the value of r to objects currently in view of the user, which requires an object detection model to be run in parallel to the privacy noise method. The method was applied in real-time to an interactive gaze-controlled action game using webcam-based eye tracking and offline to a VR dataset of video viewing. The authors reported that according to subjective feedback they received from their subjects, they enjoyed the gaming experience with real-time privacy protection. However, their real-time interaction utility evaluation was limited to eleven participants. In addition, while the suggested method is effective in protecting against re-identification, the VR dataset application was limited to the utility of predicting the visual correction prescription of users. In Table IX, we summarize the existing formal privacy methods for eye-tracking data, their corresponding guarantee, data type, and whether the trade-off was considered practical or not.

As most of the works that provide formal privacy guarantees add a considerable amount of noise to the data, this affects the performance of the utility tasks negatively. When real-time interaction is not needed and the utility task is limited to data mining, privacy protection from formal methods by adding noise is reasonable; however, especially in applications that require a real-time working capability, the amount of noise necessary for a privacy guarantee can significantly deteriorate the user experience. Taking this into consideration, researchers proposed solutions by underlining the importance of real-time and practical use by trying to limit the amount of noise introduced to real-time gaze data while still reducing the risk of re-identification. David-John et al. [330] presented a privacy protection method by utilizing spatial and temporal downsampling of gaze positions along with additive Gaussian noise when streaming the data. The findings of the authors show that when proposed methods are applied, re-identification rates drop significantly, without a formal guarantee, while the performance of gaze prediction as a utility task is minimally affected. In another work, Fuhl et al. [331] proposed training reinforcement learning agents by maximizing the rewards for utility tasks (e.g., expertise prediction, document-type classification) and by minimizing them for privacy tasks (e.g., gender detection, person identification). The authors' approach outperforms privacy protection by using generative adversarial networks (GANs) and differentially private manipulation; however, their approach works in a probabilistic way meaning that it does not guarantee privacy in a formal manner, and is most appropriate when specific risks and adversary models are known and not expected to change.

In addition to the aforementioned formal and probabilistic approaches, researchers also studied function-specific systems focused on gaze estimation [332], [333]. Bozkir et al. [332] proposed a function-specific cryptography-based method utilizing a randomized encoding-based framework in a three-party setup, where one party is identified as a server (e.g., a cloud instance) that trains machine learning models using sensitive eye movement data and the other two parties provide the sensitive data in a masked way. While raw and sensitive data from input parties are not visible to any party except the data owner, the inference time allows a real-time interaction as long as efficient communication between the parties is established. While this work is suitable for real-time interaction for any VR application, the utilized privacy framework is limited to two data-provider parties. More recently, Elfares et al. [333] proposed a federated learning approach for appearance-based gaze estimation in the wild using pseudo-gradient optimization. In federated learning [334], the machine learning models are trained in a decentralized manner, implying that sensitive data are not distributed around but kept locally, preserving data privacy. The authors showed that both in person-independent setup and in the majority of the person-specific setups, their adaptive federated learning approach outperformed vanilla federated averaging [334] for the gaze estimation task. While neither of the works used data collected from VR to train and evaluate their models, Bozkir et al. [332] used synthetic images that are generated from UnityEyes framework [86] that are very comparable to eye images obtained from VR HMDs. In contrast, Elfares et al. [333] used a real-world dataset (i.e., MPIIGaze dataset [113]). However, since VR HMDs can already be considered as personal devices, federated learning frameworks and processing data in a decentralized way fit well and their approach is also directly applicable in VR as well. The decentralization concept was recently also proposed in an anonymous eye-tracking data collection protocol for VR by eliminating the third parties for data processing and manipulation purposes [335]. While the authors eliminated third-party platforms or entities by using blockchains and smart contacts, in their protocol, if the cryptocurrency wallets to validate data authenticity are somehow linked to a wallet associated with know-your-customer (KYC) validation, an adversary could potentially identify the user to a certain extent.

Apart from the research that focused on protecting the privacy of eye movements and gaze vectors over time, a considerable amount of work focused on degrading iris authentication [336], [337], obfuscation [338], and protecting the personal identities [339] when iris textures are involved. To this end, John et al. [336] proposed an approach by utilizing an

TABLE IX
SUMMARY OF FORMAL PRIVACY METHODS APPLIED TO EYE MOVEMENT DATA. THE CHECKMARKS INDICATE WHETHER THERE IS A PRACTICAL TRADE-OFF FOR THE UTILITY TASK AND APPLICATION.

Mechanism	Guarantee	Data type	Utility task	Practical trade-off
Laplacian-DP [323]	ϵ-DP	Fixation map	Saliency map generation	×
Gaussian-DP [323]	ϵ, δ-DP	Fixation map	Saliency map generation	✓
k-same-select sequence [325]	k-anonymity	Features	Document type classification	✓
Marginals [325]	k,γ-PD	Features	Document type classification	×
Exponential-DP [24]	ϵ-DP	Features	Document type classification	×
DCFPA [295]	ϵ-DP	Features	Document type classification	✓
CFPA [295]	ϵ-DP	Features	Document type classification	✓
k-same-synth [326]	k-anonymity	Samples	Activity type classification	×
Event-synth-PD [326]	k,γ-PD	Samples	Activity type classification	✓
Kalϵido [328]	ϵ-DP	Samples	Activity type classification	✓
k-same-synth [326]	k-anonymity	Samples	Gaze prediction	✓
Event-synth-PD [326]	k,γ-PD	Samples	Gaze prediction	×
Kalϵido [328]	ϵ-DP	Samples	Gaze prediction	×
Kalϵido [328]	ϵ-DP	Samples	Gaze-based web game	✓

optical defocus in an eye-tracker setup. The authors found that such defocus causes errors in the range of calibration errors of typical eye trackers. In a later work, John et al. [340] analyzed security-utility trade-off for iris authentication using the optical defocus in a chin-rest setup and found similar results for degradation of iris authentication. Evaluating a similar setup in an immersive VR environment was proposed as a future work by the authors. In another work, John et al. [337] proposed adding pixel noise to break the iris signature to protect users from spoofing attacks. The authors argued that it is possible to replace up to 50% of the pixels in the eye image while keeping the gaze estimation error less than $2.5°$. In further studies, Eskildsen and Hansen found that an optimal method to remove the iris signature without impacting gaze estimation combines an edge-preserving filter with additive noise [338]. While these works did not directly use eye-tracking data collected from VR setups, as VR with HMDs provides a more controlled environment for eye-tracking data collection, we foresee that similar results would be obtained from VR setups as long as the underlying gaze estimation approach is the same. In the context of VR eye tracking, Chaudhary and Pelz [339] proposed replacing the iris texture regions with synthetic iris templates using a Rubber Sheet Model on the OpenEDS dataset [46], collected using VR HMDs. The authors found that such video manipulations do not degrade the semantic segmentation and pupil detection performance, which is similar to the findings of John et al. [337].

In summary, according to the current literature, privacy-preserving eye-tracking methods can be clustered into two groups, one concerning the protection of sensitive gaze and eye movement information over time, including the features that are aggregated from the gaze movements such as fixation durations or saccade rates, and the other concerning the iris obfuscation and degradation. Ultimately, the recommended privacy-preserving method depends on the target application and whether the system is running in real-time or being applied offline to a dataset. We note that real-time methods should be tuned to the privacy and utility context of a given application, and it is often challenging to find a good trade-off if formal guarantees are necessary. We also recommend that user experience should be explicitly evaluated when new privacy approaches are developed for eye-tracking data in immersive VR setups, as user experience and usability are essential factors that cannot be ignored for the sake of achieving high levels of privacy or security.

VI. DISCUSSION AND FUTURE DIRECTIONS

Taking the wide range of possibilities with eye-tracking data in VR, state-of-the-art methods to preserve the privacy of such data, and suggested best practices by the VR and privacy communities into account, detailed considerations of privacy are needed for eye tracking [341], [319]. We identify three main directions for the eye-tracking community to follow particularly for privacy and discuss these in the following subsections.

A. *Privacy in Social VR and Metaverse*

Eye tracking is heavily integrated into discussions regarding privacy in Social VR due to the important role of eye movements as a non-verbal cue during social interaction. Social interactions present special considerations for privacy based on context, as expectations vary between interactions in public spaces and interactions with friends or family in private spaces [342]. Potential solutions to establish privacy in social interactions include modulating the degree of accuracy or resolution of eye movements being mapped to a virtual avatar. On the other hand, researchers have found that non-verbal communication, through eye movements and gestures, can help preserve privacy for users who wish to avoid vocal communication in social VR [343]; suggesting that users desire control over VR sensor data streams to preserve privacy. Analogous methods exist for modulating privacy in telecommunications today, as users commonly turn off their webcam feed during Zoom or video calls to preserve the privacy of not paying attention or their background environment. Generating synthetic face and eye animations directly from audio has been proposed [344], which can help preserve privacy from raw eye movements, but is still prone to animation artifacts or incorrectly relaying social cues of the user. Open areas where advancements in understanding privacy for social VR would be critical are environments geared towards children [345], individuals with behavioral conditions such as ASD [346], and communities with privacy norms that vary from the general population [347].

Recent usage of the term “Metaverse” involves a broad vision of the future of immersive reality. While exact definitions vary, the broad futuristic visions include a connected and integrated usage of VR that spans many aspects of our daily life, superseding the use of the Internet globally today. The privacy concerns introduced by such a deeply connected and immersive VR are yet to be realized, but the long-term implications of capturing eye-tracking data of users in such environments are a critical line of future research. While unexplored in current VR platforms that include functionalities akin to a ‘Metaverse,’ an example of potential privacy concerns for the future is captured by Tadayoshi Kohno’s short story “The Schuhmacher” [348]. The story represents a fictional reality where persistent tracking of behavior, for the purpose of advertising to customers, is a two-way street that both benefits a shopkeeper’s business and has negative impacts on his relationship and social status within his town. Privacy concerns due to eye tracking on a societal scale are largely unexplored. To further establish the privacy risks that arise from eye tracking in naturalistic real-world scenarios, eye-tracking datasets such as the Virtual Experiences Database[18] can be explored. Such databases capture large-scale gaze data from typical daily activities, as opposed to prescribed experimental tasks that comprise the vast majority of research datasets. Characterizing how frequently privacy-sensitive scenarios arise due to eye tracking, and the magnitude of the privacy risk from the perspective of users can lead to an initial understanding of how to anticipate long-term risks introduced by an eye-tracked “Metaverse”.

B. Privacy vs Utility and Usability

Computational methods that attempt to preserve the privacy of the users mostly achieve privacy by adding a certain amount of noise as was the case for the works utilize differential privacy [24], [323], [337], [295]. However, as also reported by these previous works, there is a utility-privacy trade-off that should be taken care of. In differential privacy, it is especially challenging to find a good spot in terms of utility-privacy trade-off when eye-tracking datasets are considered because the longer the signals are, the more noise should be added due to higher sensitivities that contribute to the noise. Considering eye-tracking data are already a noisy source of information, especially when utilized in real time, adding additional noise to preserve privacy may ruin the user experience. Li et al. [328] addressed this issue by obtaining information on how much users enjoyed the experience when different utility-privacy levels were provided during gameplay in a desktop setup. With the immersion provided by the VR setups, the user experience and enjoyment levels already change without any privacy provided in the first place, so it is an open question of how the privacy-utility trade-off transfers to immersive setups. In addition, even if the privacy-preserving solutions are intended to be used in an offline way such as private data mining, as high amounts of noise may destroy the patterns in the data, it is important for the eye-tracking community to find private data representations that have little effect on the utility tasks while preserving the privacy.

Privacy-preserving computations aside, it is equally important to provide such solutions in an adaptive and user-centric way. For instance, one user might be purely interested in a better user experience and utility, while the other one might prefer a privacy-enhanced experience during the use of VR environments. To achieve this, user behaviors should be known in advance and virtual environments can be adapted accordingly. In fact, such privacy aspects such as preferences and attitudes along with usability have been well studied for a wide range of devices including smartphones [349], [350], IoT devices [351], [352], lifelogging cameras [353], [354], smart glasses [355], [354], [356], or augmented reality glasses [357], [23]. Despite this, these devices utilize real environments with a negligible amount to no vision augmentation. We argue that studying similar user behaviors for VR environments and devices with a focus on eye-tracking data will enable deploying user adaptive privacy schemes for virtual spaces. Such adaptions should also go hand in hand with privacy regulations as well since regulations differ based on the countries or regions [358], especially when data protection is taken into consideration.

C. Stimulus and Environmental Aspects for Privacy-preserving Eye Tracking in VR

Privacy concerns from eye-tracking data primarily depend on what the user is looking at and the context of their environment when data is being recorded. The stimulus being viewed determines what private information is at risk. For example, eye movement behavior during a VR driving simulator may reveal the driver’s age or their medical conditions such as visual field loss [359]. However, a driving task would have an extremely low chance of revealing sexual orientation, of which accurate classification requires viewing and revealing erotic stimuli [360]. Beyond just stimulus, we also must consider the context of the user, such as whether they are in the comfort of their own home viewing VR content alone, playing VR games online with friends, or using VR at work for remote collaboration in a professional setting. Prior research

[18] http://visualexperiencedatabase.org/, Last access 01/19/2023.

has demonstrated how users feel about sharing gaze data with some environments, for example, indicating that users are more likely to agree to share gaze data with medical government agencies, but are not comfortable sharing with their employer for internal use [24]. Thus, practical privacy risks and user expectations for eye tracking depend on both stimulus and environmental context.

We identify the need for further research into understanding which types of stimuli or experiences can reveal certain types of private information (age, gender, ethnicity, sexual orientation, emotion, and identity), how much data is necessary for accurate classification, and how frequently those stimuli appear in typical VR use. There is a pressing need to further quantify privacy risk across environments and bridge the gap between real-world usage of VR and findings from laboratory studies in ideal conditions. Additionally, understanding user expectations and anticipated societal norms across environments is critical to producing effective privacy-preserving systems specific to different domains. One future research direction is to view eye-tracking data through the lens of contextual integrity (CI) [361]. CI is a theoretical framework of privacy that introduces three main concepts: context, informational norms, and contextual purposes or values. Contexts capture distinct social spheres that arise naturally as part of society, including politics, religion, healthcare, or education. Informational norms are best characterized as a flow of data that society deems appropriate, and consists of a sender, recipient, data subject, data type, and finally a transmission principle that provides the logic for when and how a data flow can occur. Contextual purposes capture the social value of a context, for example, by sharing gaze data with a medical provider, a doctor may be able to diagnose a condition that otherwise would have gone unnoticed, which has inherent value to the data subject and is understood within society. Focused research is required to transfer the theoretical components of CI into a practical privacy mechanism for eye tracking. Similar efforts have been attempted in other context-based computer science research [362]. Establishing models and methods that define and enforce societal norms on eye tracking is a grand challenge for the future socio-technical landscape of VR and eye-tracking technology.

VII. Conclusion

In this paper, we systematically focused on covering the works in eye tracking in VR and the security and privacy implications including authentication schemes and methods that provide privacy-preserving eye-tracking data manipulations. To this end, we scanned papers that were published between 2012 and 2022 in major venues for VR, eye tracking, and privacy. In addition to the extensive literature review, we further provided and discussed three main research directions, especially by keeping privacy as the main focus for the research community and we indicated the importance of not only privacy but also utility and usability.

VIII. Acknowledgments

EB and EK acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC number 2064/1 - Project number 390727645.

References

[1] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, p. 1151–1160.

[2] J. Durbin, "Nvidia estimates vr is 20 years away from resolutions that match the human eye," https://uploadvr.com/nvidia-estimates-20-years-away-vr-eye-quality-resolution/, may 2017.

[3] D. J. Liebling and S. Preibusch, "Privacy considerations for a pervasive eye tracking world," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 1169–1177.

[4] O. Lappi, "Eye tracking in the wild: the good, the bad and the ugly," *Journal of Eye Movement Research*, vol. 8, no. 5, 2015.

[5] A. T. Duchowski, "Eye-based interaction in graphical systems: 20 years later gaze applications, analytics, & interaction," in *ACM SIGGRAPH 2020 Courses*. ACM, 2020.

[6] A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz, "The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality," *ACM Computing Surveys*, vol. 55, no. 3, 2022.

[7] N. Silva, T. Blascheck, R. Jianu, N. Rodrigues, D. Weiskopf, M. Raubal, and T. Schreck, "Eye tracking support for visual analytics systems: Foundations, current applications, and research challenges," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[8] B. Ens, B. Bach, M. Cordeil, U. Engelke, M. Serrano, W. Willett, A. Prouzeau, C. Anthes, W. Büschel, C. Dunne, T. Dwyer, J. Grubert, J. H. Haga, N. Kirshenbaum, D. Kobayashi, T. Lin, M. Olaosebikan, F. Pointecker, D. Saffo, N. Saquib, D. Schmalstieg, D. A. Szafir, M. Whitlock, and Y. Yang, "Grand challenges in immersive analytics," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2021.

[9] L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf, "Evaluating mixed and augmented reality: A systematic literature review (2009-2019)," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 438–451.

[10] J. Orlosky, T. Toyama, K. Kiyokawa, and D. Sonntag, "Modular: Eye-controlled vision augmentations for head mounted displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1259–1268, 2015.

[11] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs, "Focusar: Autofocus augmented reality eyeglasses for both real world and virtual imagery," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2906–2916, 2018.

[12] J. Grubert, Y. Itoh, K. Moser, and J. E. Swan, "A survey of calibration methods for optical see-through head-mounted displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 9, pp. 2649–2662, 2018.

[13] J. Spjut, B. Boudaoud, J. Kim, T. Greer, R. Albert, M. Stengel, K. Akşit, and D. Luebke, "Toward standardized classification of foveated displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 2126–2134, 2020.

[14] J. Kim, Y. Jeong, M. Stengel, K. Akşit, R. Albert, B. Boudaoud, T. Greer, J. Kim, W. Lopes, Z. Majercik, P. Shirley, J. Spjut, M. McGuire, and D. Luebke, "Foveated ar: Dynamically-foveated augmented reality display," *ACM Trans. Graph.*, vol. 38, no. 4, 2019.

[15] V. Clay, P. König, and S. König, "Eye tracking in virtual reality," *Journal of Eye Movement Research*, 2019.

[16] G. Kramida, "Resolving the vergence-accommodation conflict in head-mounted displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 7, pp. 1912–1931, 2016.

[17] X. Xia, Y. Guan, A. State, P. Chakravarthula, K. Rathinavel, T.-J. Cham, and H. Fuchs, "Towards a switchable ar/vr near-eye display with accommodation-vergence and eyeglass prescription support," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 11, pp. 3114–3124, 2019.

[18] I. B. Adhanom, P. MacNeilage, and E. Folmer, "Eye-tracking in virtual reality: a broad review of applications and challenges," *Virtual Reality*, 2023.

[19] C. Gressel, R. Overdorf, I. Hagenstedt, M. Karaboga, H. Lurtz, M. Raschke, and A. Bulling, "Privacy-aware eye tracking: Challenges and future directions," *IEEE Pervasive Computing*, vol. 22, no. 1, pp. 95–102, 2023.

[20] I. Chernyak, G. Chernyak, J. K. S. Bland, and P. D. P. Rahier, "Important considerations of data collection and curation for reliable benchmarking of end-user eye-tracking systems," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2021.

[21] I. B. Adhanom, S. C. Lee, E. Folmer, and P. MacNeilage, "Gazemetrics: An open-source tool for measuring the data quality of hmd-based eye trackers," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[22] J. L. Kröger, O. H.-M. Lutz, and F. Müller, *What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking*. Cham: Springer International Publishing, 2020, pp. 226–241.

[23] K. Lebeck, K. Ruth, T. Kohno, and F. Roesner, "Towards security and privacy for multi-user augmented reality: Foundations with end users," in *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018, pp. 392–408.

[24] J. Steil, I. Hagestedt, M. X. Huang, and A. Bulling, "Privacy-aware eye tracking using differential privacy," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019, pp. 27:1–27:9.

[25] D. Roth, G. Bente, P. Kullmann, D. Mal, C. F. Purps, K. Vogeley, and M. E. Latoschik, "Technologies for social augmentations in user-embodied virtual reality," in *25th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2019.

[26] C. Katsini, Y. Abdrabou, G. E. Raptis, M. Khamis, and F. Alt, "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*. ACM, 2020, p. 1–21.

[27] Y. Cheng, H. Wang, Y. Bao, and F. Lu, "Appearance-based gaze estimation with deep learning: A review and benchmark," 2021.

[28] A. A. Akinyelu and P. Blignaut, "Convolutional neural network-based methods for eye gaze estimation: A survey," *IEEE Access*, vol. 8, pp. 142 581–142 605, 2020.

[29] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16 495–16 519, 2017.

[30] I. B. Adhanom, S. C. Lee, E. Folmer, and P. MacNeilage, "Gazemetrics: An open-source tool for measuring the data quality of hmd-based eye trackers," in *ACM symposium on eye tracking research and applications*. ACM, 2020, pp. 1–5.

[31] I. Schuetz and K. Fiehler, "Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability," *Journal of Eye Movement Research*, vol. 15, no. 3, 2022.

[32] D. J. Lohr, L. Friedman, and O. V. Komogortsev, "Evaluating the data quality of eye tracking signals from a virtual reality system: Case study using smi's eye-tracking htc vive," *arXiv preprint arXiv:1912.02083*, 2019.

[33] T. Roth, M. Weier, A. Hinkenjann, Y. Li, and P. Slusallek, "A quality-centered analysis of eye tracking data in foveated rendering," *Journal of Eye Movement Research*, vol. 10, no. 5, 2017.

[34] C. Maraj, J. Hurter, and J. Pruitt, "Using head-mounted displays for virtual reality: Investigating subjective reactions to eye-tracking scenarios," in *International Conference on Human-Computer Interaction*. Springer, 2021, pp. 381–394.

[35] N. Stein, D. C. Niehorster, T. Watson, F. Steinicke, K. Rifai, S. Wahl, and M. Lappe, "A comparison of eye tracking latencies among several commercial head-mounted displays," *i-Perception*, vol. 12, no. 1, 2021.

[36] S. Aziz and O. Komogortsev, "An assessment of the eye tracking signal quality captured in the hololens 2," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–6.

[37] M. Stengel, S. Grogorick, M. Eisemann, E. Eisemann, and M. A. Magnor, "An affordable solution for binocular eye tracking and calibration in head-mounted displays," in *Proceedings of the 23rd ACM International Conference on Multimedia*. ACM, 2015, p. 15–24.

[38] S. W. Greenwald, L. Loreti, M. Funk, R. Zilberman, and P. Maes, "Eye gaze tracking with google cardboard using purkinje images," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 2016, p. 19–22.

[39] K. Ahuja, R. Islam, V. Parashar, K. Dey, C. Harrison, and M. Goel, "Eyespyvr: Interactive eye sensing using off-the-shelf, smartphone-based vr headsets," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 2, 2018.

[40] T. C. Stratmann, U. Gruenefeld, and S. Boll, "Eyemr: Low-cost eye-tracking for rapid-prototyping in head-mounted mixed reality," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, 2018.

[41] R. Li, E. Whitmire, M. Stengel, B. Boudaoud, J. Kautz, D. Luebke, S. Patel, and K. Akşit, "Optical gaze tracking with spatially-sparse single-pixel detectors," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 117–126.

[42] D. Katrychuk, H. K. Griffith, and O. V. Komogortsev, "Power-efficient and shift-robust eye-tracking sensor for portable vr headsets," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[43] J. Shimizu and G. Chernyshov, "Eye movement interactions in google cardboard using a low cost eog setup," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 2016, pp. 1773–1776.

[44] G. Bernal, N. Hidalgo, C. Russomanno, and P. Maes, "Galea: A physiological sensing system for behavioral research in virtual environments," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 66–76.

[45] E. Whitmire, L. Trutoiu, R. Cavin, D. Perek, B. Scally, J. Phillips, and S. Patel, "Eyecontact: scleral coil eye tracking for virtual reality," in *Proceedings of the 2016 ACM International Symposium on Wearable Computers*. ACM, 2016, pp. 184–191.

[46] S. J. Garbin, Y. Shen, I. Schuetz, R. Cavin, G. Hughes, and S. S. Talathi, "OpenEDS: Open Eye Dataset," 2019.

[47] S. J. Garbin, O. Komogortsev, R. Cavin, G. Hughes, Y. Shen, I. Schuetz, and S. S. Talathi, "Dataset for eye tracking on a virtual reality platform," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[48] J. Kim, M. Stengel, A. Majercik, S. De Mello, D. Dunn, S. Laine, M. McGuire, and D. Luebke, "Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 1–12.

[49] M. Jiang, S. Chen, J. Yang, and Q. Zhao, "Fantastic answers and where to find them: Immersive question-directed visual attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 2980–2989.

[50] I. Agtzidis, M. Startsev, and M. Dorr, "360-degree video gaze behaviour: A ground-truth data set and a classification algorithm for eye movements," in *Proceedings of the 27th ACM International Conference on Multimedia*. ACM, 2019, p. 1007–1015.

[51] W. Fuhl, G. Kasneci, and E. Kasneci, "Teyed: Over 20 million real-world eye images with pupil, eyelid, and iris 2d and 3d segmentations, 2d and 3d landmarks, 3d eyeball, gaze vector, and eye movement types," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 367–375.

[52] S. Aziz, D. J. Lohr, and O. Komogortsev, "Synchroneyes: A novel, paired data set of eye movements recorded simultaneously with remote and wearable eye-tracking devices," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–6.

[53] Y. Jin, J. Liu, F. Wang, and S. Cui, "Where are you looking? a large-scale dataset of head and gaze behavior for 360-degree videos and a pilot study," in *Proceedings of the 30th ACM International Conference on Multimedia*. ACM, 2022, pp. 1025–1034.

[54] K. J. Emery, M. Zannoli, J. Warren, L. Xiao, and S. S. Talathi, "Openneeds: A dataset of gaze, head, hand, and scene signals during exploration in open-ended vr environments," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2021.

[55] L. Tabbaa, R. Searle, S. M. Bafti, M. M. Hossain, J. Intarasisrisawat, M. Glancy, and C. S. Ang, "Vreed: Virtual reality emotion recognition dataset using eye tracking & physiological measures," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2022.

[56] S. Zhang, Q. Ma, Y. Zhang, Z. Qian, T. Kwon, M. Pollefeys, F. Bogo, and S. Tang, "Egobody: Human body shape and motion of interacting people from head-mounted devices," in *European Conference on Computer Vision*. Springer, 2022, pp. 180–200.

[57] L. Chen, Y. Li, X. Bai, X. Wang, Y. Hu, M. Song, L. Xie, Y. Yan, and E. Yin, "Real-time gaze tracking with head-eye coordination for head-mounted displays," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 82–91.

[58] T. Xue, A. El Ali, T. Zhang, G. Ding, and P. Cesar, "Rcea-360vr: Real-time, continuous emotion annotation in 360° vr videos for collecting precise viewport-dependent ground truth labels," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[59] P. Kansal and S. Devanathan, "Eyenet: Attention based convolutional encoder-decoder network for eye region segmentation," in *2019*

IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 3688–3693.

[60] W. Fuhl, D. Geisler, W. Rosenstiel, and E. Kasneci, "The applicability of cycle gans for pupil and eyelid segmentation, data generation and image refinement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*. IEEE, 2019.

[61] R. S. Kothari, A. K. Chaudhary, R. J. Bailey, J. B. Pelz, and G. J. Diaz, "Ellseg: An ellipse segmentation framework for robust gaze tracking," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2757–2767, 2021.

[62] Y. Shen, O. Komogortsev, and S. S. Talathi, "Domain adaptation for eye segmentation," in *Computer Vision – ECCV 2020 Workshops*. Cham: Springer International Publishing, 2020, pp. 555–569.

[63] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Eye-mms: Miniature multi-scale segmentation network of key eye-regions in embedded applications," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*. IEEE, 2019.

[64] J. Perry and A. S. Fernandez, "Eyeseg: Fast and efficient few-shot semantic segmentation," in *Computer Vision – ECCV 2020 Workshops*. Cham: Springer International Publishing, 2020, pp. 570–582.

[65] A. K. Chaudhary, R. Kothari, M. Acharya, S. Dangi, N. Nair, R. Bailey, C. Kanan, G. Diaz, and J. B. Pelz, "Ritnet: Real-time semantic segmentation of the eye for gaze tracking," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3698–3702.

[66] N. Damer, F. Boutros, F. Kirchbuchner, and A. Kuijper, "D-id-net: Two-stage domain and identity learning for identity-preserving image generation from semantic segmentation," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3677–3682.

[67] M. Buehler, S. Park, S. De Mello, X. Zhang, and O. Hilliges, "Content-consistent generation of realistic eyes with style," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 1–5.

[68] C. Lu, Q. Zhang, K. Krishnakumar, J. Chen, H. Fuchs, S. Talathi, and K. Liu, "Geometry-aware eye image-to-image translation," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–7.

[69] Y. Feng, N. Goulding-Hotta, A. Khan, H. Reyserhove, and Y. Zhu, "Real-time gaze tracking with event-driven eye segmentation," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 399–408.

[70] A. N. Angelopoulos, J. N. Martel, A. P. Kohli, J. Conradt, and G. Wetzstein, "Event-based near-eye gaze tracking beyond 10,000 hz," *IEEE Transactions on Visualization & Computer Graphics*, vol. 27, no. 5, pp. 2577–2586, 2021.

[71] C. Lu, P. Chakravarthula, K. Liu, X. Liu, S. Li, and H. Fuchs, "Neural 3d gaze: 3d pupil localization and gaze tracking based on anatomical eye model and neural refraction correction," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 375–383.

[72] R. Ranjan, S. De Mello, and J. Kautz, "Light-weight head pose invariant gaze tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. IEEE, 2018, p. 2156–2164.

[73] Y. Yu and J.-M. Odobez, "Unsupervised representation learning for gaze estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020, pp. 7314–7324.

[74] C. Palmero Cantarino, O. V. Komogortsev, and S. S. Talathi, "Benefits of temporal information for appearance-based gaze estimation," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[75] Y. Cheng, F. Lu, and X. Zhang, "Appearance-based gaze estimation via evaluation-guided asymmetric regression," in *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2018.

[76] S. Stojanov, S. S. Talathi, and A. Sharma, "The benefits of depth information for head-mounted gaze estimation," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–7.

[77] L. Sidenmark and A. Lundström, "Gaze behaviour on interacted objects during hand interaction in virtual reality for eye tracking calibration," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[78] I. Agtzidis and M. Dorr, "Getting (more) real: Bringing eye movement classification to hmd experiments with equirectangular stimuli," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[79] T. Rolff, F. Steinicke, and S. Frintrop, "When do saccades begin? prediction of saccades as a time-to-event problem," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–7.

[80] A. T. Duchowski and A. T. Duchowski, *Eye tracking methodology: Theory and practice*. Springer, 2017.

[81] I. Rigas, H. Raffle, and O. V. Komogortsev, "Photosensor oculography: Survey and parametric analysis of designs using model-based simulation," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 6, pp. 670–681, 2018.

[82] J. Shenoy, J. Fong, J. Tan, A. Roorda, and R. Ng, "R-slam: Optimizing eye tracking from rolling shutter video of the retina," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 2021, pp. 4852–4861.

[83] S. Thrun, "Simultaneous localization and mapping," *Robotics and cognitive approaches to spatial mapping*, pp. 13–41, 2008.

[84] Y. Sugano, Y. Matsushita, and Y. Sato, "Learning-by-synthesis for appearance-based 3d gaze estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2014, pp. 1821–1828.

[85] E. Wood, T. Baltrusaitis, X. Zhang, Y. Sugano, P. Robinson, and A. Bulling, "Rendering of eyes for eye-shape registration and gaze estimation," in *Proceedings of the IEEE international conference on computer vision*. IEEE, 2015, pp. 3756–3764.

[86] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and A. Bulling, "Learning an appearance-based gaze estimator from one million synthesised images," in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 2016, pp. 131–138.

[87] K. Wang, R. Zhao, and Q. Ji, "A hierarchical generative model for eye image synthesis and eye gaze estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018, pp. 440–448.

[88] Y. Yu, G. Liu, and J.-M. Odobez, "Improving few-shot user-specific gaze adaptation via gaze redirection synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, pp. 11 937–11 946.

[89] T. Xue, S. Ghosh, G. Ding, A. El Ali, and P. Cesar, "Designing real-time, continuous emotion annotation techniques for 360 vr videos," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020, pp. 1–9.

[90] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.

[91] W. Sankowski, K. Grabowski, M. Napieralska, M. Zubert, and A. Napieralski, "Reliable algorithm for iris segmentation in eye image," *Image and vision computing*, vol. 28, no. 2, pp. 231–237, 2010.

[92] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2016, pp. 770–778.

[93] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*. Springer, 2018, pp. 3–19.

[94] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*. IEEE, 2017, pp. 2223–2232.

[95] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2017, pp. 240–248.

[96] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[97] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2017, pp. 4700–4708.

[98] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *Journal of Vision*, vol. 16, no. 326, 2016.

[99] S. Park, A. Spurr, and O. Hilliges, "Deep pictorial gaze estimation," in *Proceedings of the European conference on computer vision (ECCV)*. Springer, 2018, pp. 721–738.

[100] S. Park, X. Zhang, A. Bulling, and O. Hilliges, "Learning to find eye region landmarks for remote gaze estimation in unconstrained settings," in *Proceedings of the 2018 ACM symposium on eye tracking research & applications*. ACM, 2018, pp. 1–10.

[101] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 478–500, 2009.

[102] X. Zhang, Y. Sugano, and A. Bulling, "Evaluation of appearance-based methods and implications for gaze-based applications," in *Proceedings of the 2019 CHI conference on human factors in computing systems*. ACM, 2019, pp. 1–13.

[103] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2015, pp. 4511–4520.

[104] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proceedings of the 2006 symposium on Eye tracking research & applications*. ACM, 2006, pp. 87–94.

[105] Z. Zhu, Q. Ji, and K. P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1. IEEE, 2006, pp. 1132–1135.

[106] J.-G. Wang, E. Sung, and R. Venkateswarlu, "Eye gaze estimation from a single image of one eye," in *Proceedings Ninth IEEE International Conference on Computer Vision*. IEEE, 2003, pp. 136–143.

[107] E. Wood and A. Bulling, "Eyetab: Model-based gaze estimation on unmodified tablet computers," in *Proceedings of the symposium on eye tracking research and applications*. ACM, 2014, pp. 207–210.

[108] A. A. Akinyelu and P. Blignaut, "Convolutional neural network-based methods for eye gaze estimation: A survey," *IEEE Access*, vol. 8, pp. 142 581–142 605, 2020.

[109] K. A. F. Mora and J.-M. Odobez, "Person independent 3d gaze estimation from remote rgb-d cameras," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 2787–2791.

[110] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2016, pp. 2176–2184.

[111] Y. Wang, T. Shen, G. Yuan, J. Bian, and X. Fu, "Appearance-based gaze estimation using deep features and random forest regression," *Knowledge-Based Systems*, vol. 110, pp. 293–301, 2016.

[112] H. Su, C. R. Qi, Y. Li, and L. J. Guibas, "Render for cnn: Viewpoint estimation in images using cnns trained with rendered 3d model views," in *Proceedings of the IEEE international conference on computer vision*. IEEE, 2015, pp. 2686–2694.

[113] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "MPIIGaze: Real-world dataset and deep appearance-based gaze estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 162–175, 2019.

[114] Z. Zhang, D. Crandall, M. Proulx, S. Talathi, and A. Sharma, "Can gaze inform egocentric action recognition?" in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–7.

[115] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 2000, pp. 71–78.

[116] O. V. Komogortsev, D. V. Gobert, S. Jayarathna, S. M. Gowda *et al.*, "Standardization of automated analyses of oculomotor fixation and saccadic behaviors," *IEEE Transactions on biomedical engineering*, vol. 57, no. 11, pp. 2635–2645, 2010.

[117] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes," *Journal of vision*, vol. 10, no. 10, pp. 28–28, 2010.

[118] K. Reiter, K. Pfeuffer, A. Esteves, T. Mittermeier, and F. Alt, "Look & turn: One-handed and expressive menu interaction by gaze and arm turns in vr," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–7.

[119] M. Choi, D. Sakamoto, and T. Ono, "Kuiper belt: Utilizing the "out-of-natural angle" region in the eye-gaze interaction for virtual reality," in *CHI Conference on Human Factors in Computing Systems*. ACM, 2022, pp. 1–17.

[120] T. Kim, A. Ham, S. Ahn, and G. Lee, "Lattice menu: A low-error gaze-based marking menu utilizing target-assisted gaze gestures on a lattice of visual anchors," in *CHI Conference on Human Factors in Computing Systems*. ACM, 2022, pp. 1–12.

[121] S. Ahn, S. Santosa, M. Parent, D. Wigdor, T. Grossman, and M. Giordano, "Stickypie: A gaze-based, scale-invariant marking menu optimized for ar/vr," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[122] J. M. Araujo, G. Zhang, J. P. P. Hansen, and S. Puthusserypady, "Exploring eye-gaze wheelchair control," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[123] X. Yi, Y. Lu, Z. Cai, Z. Wu, Y. Wang, and Y. Shi, "Gazedock: Gaze-only menu selection in virtual reality using auto-triggering peripheral menu," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 832–842.

[124] G. Lee, J. Healey, and D. Manocha, "Vrdoc: Gaze-based interactions for vr reading experience," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 787–796.

[125] M. Colley, P. Jansen, E. Rukzio, and J. Gugenheimer, "Swivr-car-seat: Exploring vehicle motion effects on interaction quality in virtual reality automated driving using a motorized swivel seat," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2022.

[126] U. Günther, K. I. S. Harrington, R. Dachselt, and I. F. Sbalzarini, "Bionic tracking: Using eye tracking to track biological cells in virtual reality," in *Computer Vision – ECCV 2020 Workshops*. Cham: Springer International Publishing, 2020, pp. 280–297.

[127] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen, "Outline pursuits: Gaze-assisted selection of occluded objects in virtual reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020, p. 1–13.

[128] L. Sidenmark, D. Potts, B. Bapisch, and H. Gellersen, "Radi-eye: Hands-free radial interfaces for 3d interaction using gaze-activated head-crossing," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[129] D. Mardanbegi, B. Mayer, K. Pfeuffer, S. Jalaliniya, H. Gellersen, and A. Perzl, "Eyeseethrough: Unifying tool selection and application in virtual environments," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 474–483.

[130] L. Sidenmark and H. Gellersen, "Eye&head: Synergetic eye and head movement for gaze pointing and selection," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 2019, p. 1161–1174.

[131] K. Pfeuffer, L. Mecke, S. Delgado Rodriguez, M. Hassib, H. Maier, and F. Alt, "Empirical evaluation of gaze-enhanced menus in virtual reality," in *26th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2020.

[132] K. Minakata, J. P. Hansen, I. S. MacKenzie, P. Bækgaard, and V. Rajanna, "Pointing by gaze, head, and foot in a head-mounted display," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[133] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger, "Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2021.

[134] A. Mutasim, A. U. Batmaz, M. Hudhud Mughrabi, and W. Stuerzlinger, "Performance analysis of saccades for primary and confirmatory target selection," in *28th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2022, pp. 1–12.

[135] L. Sidenmark, M. Parent, C.-H. Wu, J. Chan, M. Glueck, D. Wigdor, T. Grossman, and M. Giordano, "Weighted pointer: Error-aware gaze-based interaction through fallback modalities," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3585–3595, 2022.

[136] L. M. Masopust, D. Bauer, S. Yao, and K.-L. Ma, "A comparison of the fatigue progression of eye-tracked and motion-controlled interaction in immersive space," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 460–469.

[137] X. Wei, X. Shi, and L. Wang, "Label guidance based object locating in virtual reality," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 441–449.

[138] X. Meng, W. Xu, and H.-N. Liang, "An exploration of hands-free text selection for virtual reality head-mounted displays," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 74–81.

[139] Z. He, C. Lutteroth, and K. Perlin, "Tapgazer: Text entry with finger tapping and gaze-directed word selection," in *CHI Conference on Human Factors in Computing Systems*. ACM, 2022, pp. 1–16.

[140] X. Lu, D. Yu, H.-N. Liang, W. Xu, Y. Chen, X. Li, and K. Hasan, "Exploration of hands-free text entry techniques for virtual reality," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 344–349.

[141] X. Ma, Z. Yao, Y. Wang, W. Pei, and H. Chen, "Combining brain-computer interface and eye tracking for high-speed text entry in virtual reality," in *23rd International Conference on Intelligent User Interfaces*. ACM, 2018, p. 263–267.

[142] V. Rajanna and J. P. Hansen, "Gaze typing in virtual reality: Impact of keyboard design, selection method, and motion," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, 2018.

[143] D. Mardanbegi, C. Clarke, and H. Gellersen, "Monocular gaze depth estimation using the vestibulo-ocular reflex," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[144] M. Weier, T. Roth, A. Hinkenjann, and P. Slusallek, "Predicting the gaze depth in head-mounted displays using multiple feature regression," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, 2018.

[145] D. Mardanbegi, T. Langlotz, and H. Gellersen, "Resolving target ambiguity in 3d gaze interaction through vor depth estimation," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 1–12.

[146] D. L. Chen, R. Balakrishnan, and T. Grossman, "Disambiguation techniques for freehand object manipulations in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 285–292.

[147] A. Keshava, A. Aumeistere, K. Izdebski, and P. Konig, "Decoding task from oculomotor behavior in virtual reality," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[148] B. David-John, C. Peacock, T. Zhang, T. S. Murdison, H. Benko, and T. R. Jonker, "Towards gaze-based prediction of the intent to interact in virtual reality," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2021.

[149] R. Alghofaili, Y. Sawahata, H. Huang, H.-C. Wang, T. Shiratori, and L.-F. Yu, "Lost in style: Gaze-driven adaptive aid for vr navigation," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 1–12.

[150] L.-P. Cheng, E. Ofek, C. Holz, H. Benko, and A. D. Wilson, "Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, p. 3718–3728.

[151] E. J. Gonzalez, P. Abtahi, and S. Follmer, "Reach+: Extending the reachability of encountered-type haptics devices through dynamic redirection in vr," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. ACM, 2020, p. 236–248.

[152] A. Zenner, K. P. Regitz, and A. Krüger, "Blink-suppressed hand redirection," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2021, pp. 75–84.

[153] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves, "Gaze-supported 3d object manipulation in virtual reality," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2021.

[154] N. Sendhilnathan, T. Zhang, B. Lafreniere, T. Grossman, and T. R. Jonker, "Detecting input recognition errors and user errors using gaze dynamics in virtual reality," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, 2022, pp. 1–19.

[155] A. Nguyen and A. Kunz, "Discrete scene rotation during blinks and its effect on redirected walking algorithms," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2018.

[156] N. Stein, G. Bremer, and M. Lappe, "Eye tracking-based lstm for locomotion prediction in vr," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 493–503.

[157] G. Bremer, N. Stein, and M. Lappe, "Predicting future position from natural walking and eye movements with machine learning," in *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 2021, pp. 19–28.

[158] R. J. Jacob, "What you look at is what you get: eye movement-based interaction techniques," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1990, pp. 11–18.

[159] Z. Hu, S. Li, C. Zhang, K. Yi, G. Wang, and D. Manocha, "DGaze: CNN-based gaze prediction in dynamic scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1902–1911, 2020.

[160] T. Foulsham, E. Walker, and A. Kingstone, "The where, what and when of gaze allocation in the lab and the natural environment," *Vision research*, vol. 51, no. 17, pp. 1920–1931, 2011.

[161] J. S. Stahl, "Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity," *Experimental brain research*, vol. 136, no. 2, pp. 200–210, 2001.

[162] ——, "Amplitude of human head movements associated with horizontal saccades," *Experimental brain research*, vol. 126, no. 1, pp. 41–54, 1999.

[163] M. H. Urbina, M. Lorenz, and A. Huckauf, "Pies with eyes: the limits of hierarchical pie menus in gaze control," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 2010, pp. 93–96.

[164] I. S. . E. of Human-System Interaction (Subcommittee), *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs).: Guidance on Usability*. International Organization for Standardization, 1998.

[165] D. Mardanbegi and T. Pfeiffer, "Eyemrtk: A toolkit for developing eye gaze interactive applications in virtual and augmented reality," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[166] Z. Hu, A. Bulling, S. Li, and G. Wang, "Ehtask: Recognizing user tasks from eye and head movements in immersive virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021.

[167] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson, "Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences," in *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 2016, pp. 1968–1979.

[168] Q. Sun, A. Patney, L.-Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. Mcguire, D. Luebke, and A. Kaufman, "Towards virtual reality infinite walking: Dynamic saccadic redirection," *ACM Trans. Graph.*, vol. 37, no. 4, 2018.

[169] Z. Hu, C. Zhang, S. Li, G. Wang, and D. Manocha, "Sgaze: A data-driven eye-head coordination model for realtime gaze prediction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 2002–2010, 2019.

[170] Z. Hu, A. Bulling, S. Li, and G. Wang, "Fixationnet: Forecasting eye fixations in task-oriented virtual environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2681–2690, 2021.

[171] G. A. Koulieris, G. Drettakis, D. Cunningham, and K. Mania, "Gaze prediction using machine learning for dynamic stereo manipulation in games," in *2016 IEEE Virtual Reality (VR)*. IEEE, 2016, pp. 113–120.

[172] M. Zhang, K. Teck Ma, J. Hwee Lim, Q. Zhao, and J. Feng, "Deep future gaze: Gaze anticipation on egocentric videos using adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017, pp. 4372–4381.

[173] J. Vielhaben, H. Camalan, W. Samek, and M. Wenzel, "Viewport forecasting in 360° virtual reality videos with machine learning," in *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 2019.

[174] M. Assens Reina, X. Giro-i Nieto, K. McGuinness, and N. E. O'Connor, "Saltinet: Scan-path prediction on 360 degree images using saliency volumes," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*. IEEE, 2017, pp. 2331–2338.

[175] M. Assens, X. Giro-i Nieto, K. McGuinness, and N. E. O'Connor, "Pathgan: Visual scanpath prediction with generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Springer, 2018.

[176] D. Martin, A. Serrano, A. W. Bergman, G. Wetzstein, and B. Masia, "Scangan360: A generative model of realistic scanpaths for 360° images," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2003–2013, 2022.

[177] Y. Zhu, G. Zhai, and X. Min, "The prediction of head and eye movement for 360 degree images," *Signal Processing: Image Communication*, vol. 69, pp. 15–25, 2018.

[178] B. John, P. Raiturkar, O. Le Meur, and E. Jain, "A benchmark of four methods for generating 360° saliency maps from eye tracking data," in *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 2018, pp. 136–139.

[179] T. Pfeiffer and C. Memili, "Model-based real-time visualization of realistic three-dimensional heat maps for mobile eye tracking and eye tracking in virtual reality," in *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ACM, 2016, p. 95–102.

[180] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, "Saliency in vr: How do people explore virtual environments?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1633–1642, 2018.

[181] A. Nguyen, Z. Yan, and K. Nahrstedt, "Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head

movement prediction," in *Proceedings of the 26th ACM International Conference on Multimedia*. ACM, 2018, p. 1190–1198.

[182] C. Marañes, D. Gutierrez, and A. Serrano, "Exploring the impact of 360° movie cuts in users' attention," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 73–82.

[183] G. Ma, S. Li, C. Chen, A. Hao, and H. Qin, "Stage-wise salient object detection in 360° omnidirectional image via object-level semantical saliency ranking," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 12, pp. 3535–3545, 2020.

[184] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "Predicting human eye fixations via an LSTM-based saliency attentive model," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5142–5154, 2018.

[185] E. Upenik and T. Ebrahimi, "A simple method to obtain visual attention data in head mounted virtual reality," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2017, pp. 73–78.

[186] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn, "Towards foveated rendering for gaze-tracked virtual reality," *ACM Trans. Graph.*, vol. 35, no. 6, 2016.

[187] X. Meng, R. Du, and A. Varshney, "Eye-dominance-guided foveated rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1972–1980, 2020.

[188] X. Meng, R. Du, J. F. JaJa, and A. Varshney, "3d-kernel foveated rendering for light fields," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020.

[189] O. T. Tursun, E. Arabadzhiyska-Koleva, M. Wernikowski, R. Mantiuk, H.-P. Seidel, K. Myszkowski, and P. Didyk, "Luminance-contrast-aware foveated rendering," *ACM Trans. Graph.*, vol. 38, no. 4, 2019.

[190] N. Deng, Z. He, J. Ye, B. Duinkharjav, P. Chakravarthula, X. Yang, and Q. Sun, "Fov-nerf: Foveated neural radiance fields for virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3854–3864, 2022.

[191] J. Liu, C. Mantel, and S. Forchhammer, "Perception-driven hybrid foveated depth of field rendering for head-mounted displays," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 1–10.

[192] Y. Kim, Y. Ko, and I. Ihm, "Selective foveated ray tracing for head-mounted displays," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 413–421.

[193] J. Ye, A. Xie, S. Jabbireddy, Y. Li, X. Yang, and X. Meng, "Rectangular mapping-based foveated rendering," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 756–764.

[194] E. Arabadzhiyska, O. T. Tursun, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Saccade landing position prediction for gaze-contingent rendering," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 36, no. 4, 2017.

[195] A. Morales, F. M. Costela, R. Tolosana, and R. L. Woods, "Saccade landing point prediction: A novel approach based on recurrent neural networks," in *Proceedings of the 2018 International Conference on Machine Learning Technologies*. ACM, 2018, pp. 1–5.

[196] H. Griffith and O. Komogortsev, "A shift-based data augmentation strategy for improving saccade landing point prediction," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[197] C.-F. Hsu, A. Chen, C.-H. Hsu, C.-Y. Huang, C.-L. Lei, and K.-T. Chen, "Is foveated rendering perceivable in virtual reality? exploring the efficiency and consistency of quality assessment methods," in *Proceedings of the 25th ACM International Conference on Multimedia*. ACM, 2017, p. 55–63.

[198] B. Duinkharjav, K. Chen, A. Tyagi, J. He, Y. Zhu, and Q. Sun, "Color-perception-guided display power reduction for virtual reality," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 6, pp. 1–16, 2022.

[199] W. Lyu, P. Ding, Y. Zhang, A. Chen, M. Wu, S. Yin, and J. Yu, "Refocusable gigapixel panoramas for immersive vr experiences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2028–2040, 2021.

[200] T. K. Wee, E. Cuervo, and R. Balan, "Focusvr: Effective 8 usable vr display power management," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, 2018.

[201] D. Li, R. Du, A. Babu, C. D. Brumar, and A. Varshney, "A log-rectilinear transformation for foveated 360-degree video streaming," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2638–2647, 2021.

[202] P. Lungaro, R. Sjöberg, A. J. F. Valero, A. Mittal, and K. Tollmar, "Gaze-aware streaming solutions for the next generation of mobile vr experiences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1535–1544, 2018.

[203] S. Chen, B. Duinkharjav, X. Sun, L.-Y. Wei, S. Petrangeli, J. Echevarria, C. Silva, and Q. Sun, "Instant reality: Gaze-contingent perceptual optimization for 3d virtual reality streaming," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2157–2167, 2022.

[204] M. Mauderer, S. Conte, M. A. Nacenta, and D. Vishwanath, "Depth perception with gaze-contingent depth of field," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, p. 217–226.

[205] R. Konrad, A. Angelopoulos, and G. Wetzstein, "Gaze-contingent ocular parallax rendering for virtual reality," *ACM Trans. Graph.*, vol. 39, no. 2, 2020.

[206] B. Krajancich, P. Kellnhofer, and G. Wetzstein, "Optimizing depth perception in virtual and augmented reality through gaze-contingent stereo rendering," *ACM Trans. Graph.*, vol. 39, no. 6, 2020.

[207] P. Kellnhofer, P. Didyk, K. Myszkowski, M. M. Hefeeda, H.-P. Seidel, and W. Matusik, "Gazestereo3d: Seamless disparity manipulations," *ACM Trans. Graph.*, vol. 35, no. 4, 2016.

[208] N. Matsuda, A. Fix, and D. Lanman, "Focal surface displays," *ACM Trans. Graph.*, vol. 36, no. 4, 2017.

[209] O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman, "Fast gaze-contingent optimal decompositions for multifocal displays," *ACM Trans. Graph.*, vol. 36, no. 6, 2017.

[210] A. M. Aizenman, G. A. Koulieris, A. Gibaldi, V. Sehgal, D. M. Levi, and M. S. Banks, "The statistics of eye movements and binocular disparities during vr gaming: Implications for headset design," *ACM Transactions on Graphics (TOG)*, 2022.

[211] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein, "Accommodation-invariant computational near-eye displays," *ACM Trans. Graph.*, vol. 36, no. 4, 2017.

[212] D. Dunn, "Required accuracy of gaze tracking for varifocal displays," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 1838–1842.

[213] A. U. Batmaz, M. D. Barrera Machuca, J. Sun, and W. Stuerzlinger, "The effect of the vergence-accommodation conflict on virtual hand pointing in immersive displays," in *CHI Conference on Human Factors in Computing Systems*. ACM, 2022, pp. 1–15.

[214] H. L. Roth, A. N. Lora, and K. M. Heilman, "Effects of monocular viewing and eye dominance on spatial attention," *Brain*, vol. 125, no. 9, pp. 2023–2035, 2002.

[215] S. Winkler, E. D. Gelasca, and T. Ebrahimi, "Toward perceptual metrics for video watermark evaluation," in *Applications of Digital Image Processing XXVI*, vol. 5203. SPIE, 2003, pp. 371–378.

[216] E. Bozkir, D. Geisler, and E. Kasneci, "Person independent, privacy preserving, and real time assessment of cognitive load using eye tracking in a virtual reality setup," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 1834–1837.

[217] A. D. Souchet, M. L. Diallo, and D. Lourdeaux, "Cognitive load classification with a stroop task in virtual reality based on physiological data," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 656–666.

[218] A. T. Duchowski, K. Krejtz, I. Krejtz, C. Biele, A. Niedzielska, P. Kiefer, M. Raubal, and I. Giannopoulos, "The index of pupillary activity: Measuring cognitive load ¡i¿vis-à-vis¡/i¿ task difficulty with pupil oscillation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 1–13.

[219] T. Luong, N. Martin, A. Raison, F. Argelaguet, J.-M. Diverrez, and A. Lécuyer, "Towards real-time recognition of users mental workload using integrated physiological sensors into a vr hmd," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2020, pp. 425–437.

[220] J. C. Wilson, S. Nair, S. Scielzo, and E. C. Larson, "Objective measures of cognitive load using deep multi-modal learning: A use-case in aviation," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 1, 2021.

[221] P. Bækgaard, J. P. Hansen, K. Minakata, and I. S. MacKenzie, "A fitts' law study of pupil dilations in a head-mounted display," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[222] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement." *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.

[223] B. John, P. Raiturkar, A. Banerjee, and E. Jain, "An evaluation of pupillary light response models for 2d screens and vr hmds," in

Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. ACM, 2018.

[224] H. Gao, L. Frommelt, and E. Kasneci, "The evaluation of gait-free locomotion methods with eye movement in virtual reality," in *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2022, pp. 530–535.

[225] D. Pan, Q. Xu, S. Ma, and K. Zhang, "The impact of fear of the sea on working memory performance: A research based on virtual reality," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2018.

[226] T. Luong and C. Holz, "Characterizing physiological responses to fear, frustration, and insight in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3917–3927, 2022.

[227] D. Lindlbauer, A. M. Feit, and O. Hilliges, "Context-aware online adaptation of mixed reality interfaces," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 2019, p. 147–160.

[228] T. C. Kübler, E. Kasneci, and F. Vintila, "Pupil response as an indicator of hazard perception during simulator driving," *Journal of Eye Movement Research*, vol. 10, no. 4, 2017.

[229] E. Bekele, Z. Zheng, A. Swanson, J. Crittendon, Z. Warren, and N. Sarkar, "Understanding how adolescents with autism respond to facial expressions in virtual reality environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 711–720, 2013.

[230] E. Bekele, J. Wade, D. Bian, J. Fan, A. Swanson, Z. Warren, and N. Sarkar, "Multimodal adaptive social interaction in virtual environment (masi-vr) for children with autism spectrum disorders (asd)," in *2016 IEEE Virtual Reality (VR)*. IEEE, 2016, pp. 121–130.

[231] B. Kim, D. Jeong, M. Jeong, T. Noh, S.-I. Kim, T. Kim, S.-Y. Jang, H. J. Yoo, J. Kim, H. Hong *et al.*, "Vista: User-centered vr training system for effectively deriving characteristics of people with autism spectrum disorder," in *28th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2022, pp. 1–12.

[232] J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan, and H. Devos, "Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1302–1311, 2017.

[233] A. Schmitz, A. MacQuarrie, S. Julier, N. Binetti, and A. Steed, "Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 63–72.

[234] R. Soret, P. Charras, C. Hurter, and V. Peysakhovich, "Attentional orienting in virtual reality using endogenous and exogenous cues in auditory and visual modalities," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[235] J.-S. Liu, P. Wang, B. Tversky, and S. Feiner, "Adaptive visual cues for guiding a bimanual unordered task in virtual reality," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 431–440.

[236] D. Lange, T. C. Stratmann, U. Gruenefeld, and S. Boll, "Hivefive: Immersion preserving attention guidance in virtual reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020, p. 1–13.

[237] E. Bozkir, D. Geisler, and E. Kasneci, "Assessment of driver attention during a safety critical situation in VR to generate VR-based training," in *ACM Symposium on Applied Perception 2019*. ACM, 2019.

[238] C.-Y. Hsieh, Y.-S. Chiang, H.-Y. Chiu, and Y.-J. Chang, "Bridging the virtual and real worlds: A preliminary study of messaging notifications in virtual reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020, p. 1–14.

[239] K.-W. Chen, Y.-J. Chang, and L. Chan, "Predicting opportune moments to deliver notifications in virtual reality," in *CHI Conference on Human Factors in Computing Systems*. ACM, 2022, pp. 1–18.

[240] S. Marwecki, A. D. Wilson, E. Ofek, M. Gonzalez Franco, and C. Holz, "Mise-unseen: Using eye tracking to hide virtual reality scene changes in plain sight," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 2019, p. 777–789.

[241] R. Alghofaili, M. S. Solah, H. Huang, Y. Sawahata, M. Pomplun, and L.-F. Yu, "Optimizing visual element placement via visual attention analysis," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 464–473.

[242] S. Hillaire, A. Lecuyer, T. Regia-Corte, R. Cozot, J. Royan, and G. Breton, "Design and application of real-time visual attention model for the exploration of 3d virtual environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 3, pp. 356–368, 2012.

[243] O. Špakov, H. Istance, K.-J. Räihä, T. Viitanen, and H. Siirtola, "Eye gaze and head gaze in collaborative games," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[244] S. Kasahara, M. Ando, K. Suganuma, and J. Rekimoto, "Parallel eyes: Exploring human capability and behaviors with paralleled first person view sharing," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, p. 1561–1572.

[245] C. Mei, B. T. Zahed, L. Mason, and J. Ouarles, "Towards joint attention training for children with asd - a vr game approach and eye gaze exploration," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2018, pp. 289–296.

[246] K. Ahuja, D. Shah, S. Pareddy, F. Xhakaj, A. Ogan, Y. Agarwal, and C. Harrison, "Classroom digital twins with instrumentation-free gaze tracking," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[247] H. Gao, E. Bozkir, L. Hasenbein, J.-U. Hahn, R. Göllner, and E. Kasneci, "Digital transformations of classrooms in virtual reality," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[248] E. Bozkir, P. Stark, H. Gao, L. Hasenbein, J.-U. Hahn, E. Kasneci, and R. Göllner, "Exploiting object-of-interest information to understand attention in vr classrooms," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2021, pp. 597–605.

[249] H. Gao, L. Hasenbein, E. Bozkir, R. Göllner, and E. Kasneci, "Evaluating the effects of virtual human animation on students in an immersive vr classroom using eye movements," in *28th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2022, pp. 1–11.

[250] ——, "Exploring gender differences in computational thinking learning in a VR classroom: Developing machine learning models using eye-tracking data and explaining the models," *International Journal of Artificial Intelligence in Education*, 2022.

[251] Y. Rahman, S. M. Asish, N. P. Fisher, E. C. Bruce, A. K. Kulshreshth, and C. W. Borst, "Exploring eye gaze visualization techniques for identifying distracted students in educational vr," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 868–877.

[252] S. M. LaValle, *Virtual Reality*. Cambridge, United Kingdom: Cambridge University Press, 2020. [Online]. Available: http://lavalle.pl/vr/

[253] N. Katzakis, L. Chen, O. Ariza, R. J. Teather, and F. Steinicke, "Evaluation of 3d pointing accuracy in the fovea and periphery in immersive head-mounted display environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 1929–1936, 2021.

[254] M. S. Arefin, J. E. Swan II, R. A. Cohen Hoffing, and S. M. Thurman, "Estimating perceptual depth changes with eye vergence and interpupillary distance using an eye tracker in virtual reality," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022, pp. 1–7.

[255] A. Serrano, V. Sitzmann, J. Ruiz-Borau, G. Wetzstein, D. Gutierrez, and B. Masia, "Movie editing and cognitive event segmentation in virtual reality video," *ACM Trans. Graph.*, vol. 36, no. 4, 2017.

[256] A. MacQuarrie and A. Steed, "Perception of volumetric characters' eye-gaze direction in head-mounted displays," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 645–654.

[257] J. G. Shin, D. Kim, C. So, and D. Saakes, "Body follows eye: Unobtrusive posture manipulation through a dynamic content position in virtual reality." ACM, 2020, p. 1–14.

[258] E. Langbehn, F. Steinicke, M. Lappe, G. F. Welch, and G. Bruder, "In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality," *ACM Trans. Graph.*, vol. 37, no. 4, 2018.

[259] M. Keyvanara and R. Allison, "Effect of a constant camera rotation on the visibility of transsaccadic camera shifts," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[260] M. Keyvanara and R. S. Allison, "Sensitivity to natural 3d image transformations during eye movements," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, 2018.

[261] B. Bolte and M. Lappe, "Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 4, pp. 545–552, 2015.

[262] F. Behrens, M. MacKeben, and W. Schröder-Preikschat, "An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters," *Behavior research methods*, vol. 42, no. 3, pp. 701–708, 2010.

[263] P.-H. Niemenlehto, "Constant false alarm rate detection of saccadic eye movements in electro-oculography," *Computer methods and programs in biomedicine*, vol. 96, no. 2, pp. 158–171, 2009.

[264] L. R. Luidolt, M. Wimmer, and K. Krösl, "Gaze-dependent simulation of light perception in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 12, pp. 3557–3567, 2020.

[265] F. Berton, L. Hoyet, A.-H. Olivier, J. Bruneau, O. Le Meur, and J. Pettre, "Eye-gaze activity in crowds: Impact of virtual reality and density," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 322–331.

[266] F. Berton, A.-H. Olivier, J. Bruneau, L. Hoyet, and J. Pettre, "Studying gaze behaviour during collision avoidance with a virtual walker: Influence of the virtual reality setup," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 717–725.

[267] P. Raimbaud, A. Jovane, K. Zibrek, C. Pacchierotti, M. Christie, L. Hoyet, J. Pettre, and A.-H. Olivier, "The stare-in-the-crowd effect in virtual reality," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 281–290.

[268] K. Gupta, S. W. Chan, Y. S. Pai, N. Strachan, J. Su, A. Sumich, S. Nanayakkara, and M. Billinghurst, "Total vrecall: using biosignals to recognize emotional autobiographical memory in virtual reality," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–21, 2022.

[269] D. Borland, T. Peck, and M. Slater, "An evaluation of self-avatar eye movement for virtual embodiment," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 4, pp. 591–596, 2013.

[270] I. Bergström, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater, "The plausibility of a string quartet performance in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1352–1359, 2017.

[271] F. Ma and X. Pan, "Visual fidelity effects on expressive self-avatar in virtual reality: First impressions matter," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2022, pp. 57–65.

[272] B. H. Le, X. Ma, and Z. Deng, "Live speech driven head-and-eye motion generators," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 11, pp. 1902–1914, 2012.

[273] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Facevr: Real-time gaze-aware facial reenactment in virtual reality," *ACM Trans. Graph.*, vol. 37, no. 2, 2018.

[274] P. Ladwig, A. Pech, R. Dörner, and C. Geiger, "Unmasking communication partners: A low-cost ai solution for digitally removing head-mounted displays in vr-based telepresence," in *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2020, pp. 82–90.

[275] G. Song, J. Cai, T.-J. Cham, J. Zheng, J. Zhang, and H. Fuchs, "Real-time 3d face-eye performance capture of a person wearing vr headset," in *Proceedings of the 26th ACM International Conference on Multimedia*. ACM, 2018, p. 923–931.

[276] K. Olszewski, J. J. Lim, S. Saito, and H. Li, "High-fidelity facial and speech animation for vr hmds," *ACM Trans. Graph.*, vol. 35, no. 6, 2016.

[277] Y. Zhao, Q. Xu, W. Chen, C. Du, J. Xing, X. Huang, and R. Yang, "Mask-off: Synthesizing face images in the presence of head-mounted displays," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 267–276.

[278] R. Islam, K. Desai, and J. Quarles, "Cybersickness prediction from integrated hmd's sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2021, pp. 31–40.

[279] ——, "Towards forecasting the onset of cybersickness by fusing physiological, head-tracking and eye-tracking with multimodal deep fusion network," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2022, pp. 121–130.

[280] T. M. Lee, J.-C. Yoon, and I.-K. Lee, "Motion sickness prediction in stereoscopic videos using 3d convolutional neural networks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1919–1927, 2019.

[281] N. Padmanaban, T. Ruban, V. Sitzmann, A. M. Norcia, and G. Wetzstein, "Towards a machine-learning approach for sickness prediction in 360 stereoscopic videos," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1594–1603, 2018.

[282] M. Al Zayer, P. MacNeilage, and E. Folmer, "Virtual locomotion: A survey," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 6, pp. 2315–2334, 2020.

[283] I. B. Adhanom, N. Navarro Griffin, P. MacNeilage, and E. Folmer, "The effect of a foveated field-of-view restrictor on vr sickness," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020, pp. 645–652.

[284] Y. Lang, L. Wei, F. Xu, Y. Zhao, and L.-F. Yu, "Synthesizing personalized training programs for improving driving habits via virtual reality," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2018, pp. 297–304.

[285] B. W. Hosp, F. Schultz, O. Höner, and E. Kasneci, "Soccer goalkeeper expertise identification based on eye movements," *PLOS ONE*, vol. 16, no. 5, pp. 1–22, 05 2021.

[286] J. Orlosky, B. Huynh, and T. Hollerer, "Using eye tracked virtual reality to classify understanding of vocabulary in recall tasks," in *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 2019, pp. 66–667.

[287] A. Burova, J. Mäkelä, J. Hakulinen, T. Keskinen, H. Heinonen, S. Siltanen, and M. Turunen, "Utilizing vr and gaze tracking to develop ar solutions for industrial maintenance," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020, p. 1–13.

[288] J. Gisler, J. Schneider, J. Handali, V. Holzwarth, C. Hirt, W. Fuhl, J. Vom Brocke, and A. Kunz, "Indicators of training success in virtual reality using head and eye movements," in *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2021, pp. 280–285.

[289] D. J. Harris, K. J. Hardcastle, M. R. Wilson, and S. J. Vine, "Assessing the learning and transfer of gaze behaviours in immersive virtual reality," *Virtual Reality*, 2021.

[290] K. Krösl, C. Elvezio, M. Hürbe, S. Karst, M. Wimmer, and S. Feiner, "Icthroughvr: Illuminating cataracts through virtual reality," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 655–663.

[291] J.-H. Kim, H.-J. Son, S.-J. Lee, D.-Y. Yun, S.-C. Kwon, and S.-H. Lee, "Effectiveness of a virtual reality head-mounted display system-based developmental eye movement test," *Journal of Eye Movement Research*, vol. 9, no. 6, 2016.

[292] J.-H. Kim, H.-J. Son, S.-H. Lee, and S.-C. Kwon, "Vr/ar head-mounted display system based measurement and evaluation of dynamic visual acuity," *Journal of Eye Movement Research*, vol. 12, no. 8, 2019.

[293] K. Hotta, O. D. A. Prima, T. Imabuchi, and H. Ito, "Vr-hmd eye tracker in active visual field testing," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 1843–1847.

[294] T. C. Kübler, E. Kasneci, W. Rosenstiel, K. Aehling, M. Heister, K. Nagel, U. Schiefer, and E. Papageorgiou, "Driving with homonymous visual field defects: Driving performance and compensatory gaze movements," *Journal of Eye Movement Research*, vol. 8, no. 5, 2015.

[295] E. Bozkir, O. Günlü, W. Fuhl, R. F. Schaefer, and E. Kasneci, "Differential privacy for eye tracking with temporal correlations," *Plos one*, vol. 16, no. 8, p. e0255979, 2021.

[296] M. Xu, W. C. Ng, W. Y. B. Lim, J. Kang, Z. Xiong, D. Niyato, Q. Yang, X. S. Shen, and C. Miao, "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Communications Surveys & Tutorials*, pp. 1–1, 2022.

[297] K. Pfeuffer, M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt, "Behavioural biometrics in vr: Identifying people from body motion and relations in virtual reality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 1–12.

[298] J. Liebers, M. Abdelaziz, L. Mecke, A. Saad, J. Auda, U. Gruenefeld, F. Alt, and S. Schneegass, "Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[299] J. Liebers, P. Horn, C. Burschik, U. Gruenefeld, and S. Schneegass, "Using gaze behavior and head orientation for implicit identification in virtual reality," in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2021, pp. 22:1–22:9.

[300] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," *Pattern Recognition*, vol. 43, no. 3, pp. 1016–1026, 2010.

[301] X. Song, Z. Chen, and D. Sun, "Iris ciphertext authentication system based on fully homomorphic encryption," *Journal of Information Processing Systems*, vol. 16, no. 3, pp. 599–611, 2020.

[302] J. Daugman, "Iris recognition at airports and border-crossings," in *Encyclopedia of Biometrics*. Springer, 2009, pp. 819–825.

[303] S. Stephenson, B. Pal, S. Fan, E. Fernandes, Y. Zhao, and R. Chatterjee, "Sok: Authentication in augmented and virtual reality," in *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022, pp. 1552–1552.

[304] O. V. Komogortsev and C. D. Holland, "Biometric authentication via complex oculomotor behavior," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2013, pp. 1–8.

[305] O. V. Komogortsev, S. Jayarathna, C. R. Aragon, and M. Mahmoud, "Biometric identification via an oculomotor plant mathematical model," in *ACM Symposium on Eye-Tracking Research & Applications*. ACM, 2010, pp. 57–60.

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

[307] S. Eberz, K. B. Rasmussen, V. Lenders, and I. Martinovic, "Looks like eve: Exposing insider threats using eye movement biometrics," *ACM Trans. Priv. Secur.*, vol. 19, no. 1, pp. 1:1–1:31, 2016.

[308] S. Eberz, G. Lovisotto, K. B. Rasmussen, V. Lenders, and I. Martinovic, "28 blinks later: Tackling practical challenges of eye movement biometrics," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2019, p. 1187–1199.

[309] Y. Zhang, W. Hu, W. Xu, C. T. Chou, and J. Hu, "Continuous authentication using eye movement response of implicit visual stimuli," *ACM Interact. Mob. Wearable Ubiquitous Technology*, vol. 1, no. 4, pp. 177:1–177:22, 2018.

[310] H. Zhu, W. Jin, M. Xiao, S. Murali, and M. Li, "Blinkey: A two-factor user authentication method for virtual reality devices," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 4, 2020.

[311] D. Lohr and O. V. Komogortsev, "Eye know you too: Toward viable end-to-end eye movement biometrics for user authentication," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3151–3164, 2022.

[312] Y. Niitsu and M. Nakayama, "Effects of measurement time and presentation size conditions on biometric identification using eye movements," in *ACM Symposium on Eye Tracking Research and Applications*. New York, NY, USA: ACM, 2021.

[313] L. Friedman, H. Stern, V. Prokopenko, S. Djanian, H. Griffith, and O. Komogortsev, "Biometric performance as a function of gallery size," *Applied Sciences*, vol. 12, no. 21, 2022.

[314] D. Lohr, S.-H. Berndt, and O. Komogortsev, "An implementation of eye movement-driven biometrics in virtual reality," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, 2018.

[315] S. Luo, A. Nguyen, C. Song, F. Lin, W. Xu, and Z. Yan, "Oculock: Exploring human visual system for authentication in virtual reality head-mounted display," in *Network and Distributed Systems Security (NDSS) Symposium*, 2020.

[316] L. Friedman and O. V. Komogortsev, "Assessment of the effectiveness of seven biometric feature normalization techniques," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2528–2536, 2019.

[317] F. Mathis, K. Vaniea, and M. Khamis, "Replicueauth: Validating the use of a lab-based virtual reality setup for evaluating authentication systems," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[318] A. H. Mhaidli and F. Schaub, "Identifying manipulative advertising techniques in xr through scenario construction," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021.

[319] Future of Privacy Forum, "New infographic highlights xr technology data flows and privacy risks," https://fpf.org/blog/new-infographic-highlights-xr-technology-data-flows-and-privacy-risks/, Oct 2022, accessed: 2022-11-18.

[320] G. M. Garrido, V. Nair, and D. Song, "Sok: Data privacy in virtual reality," 2023.

[321] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, S. Halevi and T. Rabin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–284.

[322] C. Dwork, "Differential privacy," in *Automata, Languages and Programming*. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 2006, pp. 1–12.

[323] A. Liu, L. Xia, A. Duchowski, R. Bailey, K. Holmqvist, and E. Jain, "Differential privacy for eye-tracking data," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. New York, NY, USA: ACM, 2019.

[324] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*. ACM, 2011, pp. 193–204.

[325] B. David-John, K. Butler, and E. Jain, "For your eyes only: Privacy-preserving eye-tracking datasets," in *2022 Symposium on Eye Tracking Research and Applications*. ACM, 2022.

[326] ——, "Privacy-preserving datasets of eye-tracking samples with applications in xr," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2774–2784, 2023.

[327] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," 1998.

[328] J. Li, A. R. Chowdhury, K. Fawaz, and Y. Kim, "Kalεido: Real-time privacy control for eye-tracking systems," in *30th USENIX Security Symposium (USENIX Security 21)*. Vancouver, B.C.: USENIX Association, 2021.

[329] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 901–914.

[330] B. David-John, D. Hosfelt, K. Butler, and E. Jain, "A privacy-preserving approach to streaming eye-tracking data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2555–2565, 2021.

[331] W. Fuhl, E. Bozkir, and E. Kasneci, "Reinforcement learning for the privacy preservation and manipulation of eye tracking data," in *Artificial Neural Networks and Machine Learning – ICANN 2021*. Springer International Publishing, 2021, pp. 595–607.

[332] E. Bozkir, A. B. Ünal, M. Akgün, E. Kasneci, and N. Pfeifer, "Privacy preserving gaze estimation using synthetic images via a randomized encoding based framework," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020, pp. 21:1–21:5.

[333] M. Elfares, Z. Hu, P. Reisert, A. Bulling, and R. Küsters, "Federated learning for appearance-based gaze estimation in the wild," in *Proceedings of Machine Learning Research, NeurIPS 2022 Gaze Meets ML Workshop*, 2022, pp. 1–17.

[334] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54. PMLR, 2017, pp. 1273–1282.

[335] E. Bozkir, S. Eivazi, M. Akgün, and E. Kasneci, "Eye tracking data collection protocol for vr for remotely located subjects using blockchain and smart contracts," in *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 2020, pp. 397–401.

[336] B. John, S. Koppal, and E. Jain, "EyeVEIL: Degrading iris authentication in eye tracking headsets," in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019.

[337] B. John, A. Liu, L. Xia, S. Koppal, and E. Jain, "Let it snow: Adding pixel noise to protect the user's identity," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[338] A. M. Eskildsen and D. Witzner Hansen, "Analysis of iris obfuscation: Generalising eye information processes for privacy studies in eye tracking." in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2021.

[339] A. K. Chaudhary and J. B. Pelz, "Privacy-preserving eye videos using rubber sheet model," in *ACM Symposium on Eye Tracking Research and Applications*. ACM, 2020.

[340] B. John, S. Jörg, S. Koppal, and E. Jain, "The security-utility trade-off for iris authentication and eye animation for social virtual avatars," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1880–1890, 2020.

[341] B. David-John, D. Hosfelt, K. Butler, and E. Jain, "Let's SOUP up XR: Collected thoughts from an IEEE VR workshop on privacy in mixed reality," *VR4Sec: Security for VR and VR for Security, SOUPS 2021 Workshop*, 2021. [Online]. Available: https://par.nsf.gov/biblio/10354958

[342] D. Maloney, S. Zamanifard, and G. Freeman, "Anonymity vs. familiarity: Self-disclosure and privacy in social virtual reality," in *26th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2020.

[343] D. Maloney, G. Freeman, and D. Y. Wohn, ""talking without a voice" understanding non-verbal communication in social virtual reality," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–25, 2020.

[344] M. Villanueva Aylagas, H. Anadon Leon, M. Teye, and K. Tollmar, "Voice2Face: Audio-driven Facial and Tongue Rig Animations with cVAEs," *Computer Graphics Forum*, 2022.

[345] D. Maloney, G. Freeman, and A. Robb, "A virtual space for all: Exploring children's experience in social virtual reality," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, 2020, pp. 472–483.

[346] J. G. Kim, T. Kim, S.-I. Kim, S.-y. Jang, E. B. Lee, H. Yoo, K. Han, and H. Hong, "The workplace playbook vr: Exploring the design space of virtual reality to foster understanding of and support for autistic people," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, pp. 1–24, 2022.

[347] D. Maloney, G. Freeman, and A. Robb, "Social virtual reality: ethical considerations and future directions for an emerging research space," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2021, pp. 271–277.

[348] T. Kohno, "The schuhmacher," *IEEE Security & Privacy*, vol. 20, no. 05, pp. 14–15, 2022.

[349] A. P. Felt, S. Egelman, and D. Wagner, "I've got 99 problems, but vibration ain't one: A survey of smartphone users' concerns," in *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 2012, p. 33–44.

[350] R. Balebako, A. Marsh, J. Lin, J. Hong, and L. F. Cranor, "The privacy and security behaviors of smartphone app developers," in *NDSS Workshop on Usable Security (USEC)*, 2014, pp. 1–10.

[351] P. E. Naeini, S. Bhagavatula, H. Habib, M. Degeling, L. Bauer, L. F. Cranor, and N. Sadeh, "Privacy expectations and preferences in an {IoT} world," in *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, 2017, pp. 399–412.

[352] P. Emami Naeini, M. Degeling, L. Bauer, R. Chow, L. F. Cranor, M. R. Haghighat, and H. Patterson, "The influence of friends and experts on privacy decision making in iot scenarios," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, 2018.

[353] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia, "Privacy behaviors of lifeloggers using wearable cameras," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, p. 571–582.

[354] R. Hoyle, R. Templeman, D. Anthony, D. Crandall, and A. Kapadia, "Sensitive lifelogs: A privacy analysis of photos from wearable cameras," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, p. 1645–1648.

[355] L. Lee, J. Lee, S. Egelman, and D. Wagner, "Information disclosure concerns in the age of wearable computing," in *NDSS Workshop on Usable Security (USEC)*, vol. 1, 2016, pp. 1–10.

[356] M. Koelle, M. Kranz, and A. Möller, "Don't look at me that way! understanding user attitudes towards data glasses usage," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2015, p. 362–372.

[357] T. Denning, Z. Dehlawi, and T. Kohno, "In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, p. 2377–2386.

[358] F. Lucarini, "The differences between the california consumer privacy act and the gdpr," https://advisera.com/articles/gdpr-vs-ccpa-what-are-the-main-differences/, last accessed 30/12/2022.

[359] E. Kasneci, K. Sippel, K. Aehling, M. Heister, W. Rosenstiel, U. Schiefer, and E. Papageorgiou, "Driving with binocular visual field loss? a study on a supervised on-road parcours with simultaneous eye and head tracking," *PloS one*, vol. 9, no. 2, p. e87470, 2014.

[360] G. Rieger and R. C. Savin-Williams, "The eyes have it: Sex and sexual orientation differences in pupil dilation patterns," *PloS one*, vol. 7, no. 8, p. e40256, 2012.

[361] H. Nissenbaum, "Privacy as contextual integrity," *Wash. L. Rev.*, vol. 79, p. 119, 2004.

[362] S. Benthall, S. Gürses, H. Nissenbaum *et al.*, *Contextual integrity through the lens of computer science*. Now Publishers, 2017.