



Eye-tracking on virtual reality: a survey

Jesús Moreno-Arjonilla¹ · Alfonso López-Ruiz¹ · J. Roberto Jiménez-Pérez¹ · José E. Callejas-Aguilera² · Juan M. Jurado³

Received: 21 December 2022 / Accepted: 28 October 2023 / Published online: 5 February 2024
© The Author(s) 2024

Abstract

Virtual reality (VR) has evolved substantially beyond its initial remit of gaming and entertainment, catalyzed by advancements such as improved screen resolutions and more accessible devices. Among various interaction techniques introduced to VR, eye-tracking stands out as a pivotal development. It not only augments immersion but offers a nuanced insight into user behavior and attention. This precision in capturing gaze direction has made eye-tracking instrumental for applications far beyond mere interaction, influencing areas like medical diagnostics, neuroscientific research, educational interventions, and architectural design, to name a few. Though eye-tracking's integration into VR has been acknowledged in prior reviews, its true depth, spanning the intricacies of its deployment to its broader ramifications across diverse sectors, has been sparsely explored. This survey undertakes that endeavor, offering a comprehensive overview of eye-tracking's state of the art within the VR landscape. We delve into its technological nuances, its pivotal role in modern VR applications, and its transformative impact on domains ranging from medicine and neuroscience to marketing and education. Through this exploration, we aim to present a cohesive understanding of the current capabilities, challenges, and future potential of eye-tracking in VR, underscoring its significance and the novelty of our contribution.

Keywords Virtual reality · Eye-tracking · Perception · Attention

Jesús Moreno-Arjonilla, Alfonso López-Ruiz, J. Roberto Jiménez-Pérez, José E. Callejas-Aguilera and Juan M. Jurado have contributed equally to this work.

✉ Jesús Moreno-Arjonilla
jarjonil@ujaen.es

Alfonso López-Ruiz
allopezr@ujaen.es

J. Roberto Jiménez-Pérez
rjimenez@ujaen.es

José E. Callejas-Aguilera
jecalle@ujaen.es

Juan M. Jurado
jjurado@ujaen.es

¹ Computer Science Department, University of Jaén, 23071 Jaén, Spain

² Psychology Department, University of Jaén, 23071 Jaén, Spain

³ Software Engineering Department, University of Granada, 18071 Granada, Spain

1 Introduction

Virtual reality (VR) represents a significant advancement in technological paradigms, offering immersive digital environments that redefine conventional user interactions. The potential applications of VR, especially when integrated with eye-tracking, are vast and diverse. This survey endeavors to provide a comprehensive examination of the intersection between VR and eye-tracking, establishing a foundational platform for our forthcoming research trajectory in psychology.

Within the VR context, users are exposed to an array of stimuli that closely simulate real-world experiences. The democratization of VR technology, found in the increasing affordability and ubiquity of VR headsets, has catalyzed its adoption across a myriad of sectors. Complementing this is the integration of eye-tracking technology, which captures users' gaze patterns, providing a nuanced perspective into their cognitive and perceptual processes. While early iterations of this technology were constrained by the necessity for intricate equipment, recent innovations have facilitated

its seamless amalgamation with VR headsets, Augmented Reality (AR) interfaces, and conventional display systems.

The confluence of VR and eye-tracking has ushered in a plethora of applications, from entertainment to education and specialized research. Yet, this union is not without challenges. Considerations pertaining to individual optics, inherent ocular deviations, and issues of spatial accuracy underscore the necessity for continued research and refinement.

In this survey, we undertake a meticulous exploration of the VR landscape, delineating between commercial solutions and bespoke systems crafted for research-specific needs. Concurrently, we assess the broader ramifications of eye-tracking across an array of disciplines, emphasizing its quintessential role in procuring accurate attention metrics. Notably, while several VR systems deploy head-tracking methodologies to ascertain user orientation within virtual spaces, it is the granular accuracy of eye-tracking that proffers unparalleled insights into attentional dynamics.

Our primary objective is to traverse the existing literature, elucidating the nuances of attention tracking within VR. This spans a spectrum from pragmatic applications to the frontiers of hardware evolution. This synthesis seeks to disambiguate the complex relationship between VR and

eye-tracking, elucidating present capabilities, prospective developments, and existing challenges. For a contextual comparison, Table 1 juxtaposes our insights with extant reviews, underscoring the unique contribution of our work. It is with this foundational knowledge that we anticipate our foray into the domain of psychology, harnessing the synergistic potential of VR and eye-tracking.

This paper is organized as follows: first, the methodology to find previous work is presented; then, the basis of eye functioning is summarized to understand the movements that are frequently monitored in eye-tracking and the metrics that are commonly used in studies that use this type of hardware. Commercial and custom hardware are following analyzed by evaluating work solely based on end-user applications or focused on building an eye-tracking system. Following, the eye-tracking procedure, together with the calibration, is detailed on the basis of works implementing their own system or pipeline, instead of delving into commercial software. Optimizations concerning VR rendering are then presented to show current and future techniques aimed at fusing the understanding of how the eye works and efficient rendering. An in-depth review of the applicability of eye-tracking

Table 1 Comparison between this survey and previous VR studies

Main topic	Surveys	Brief description
Visualization of eye-tracking data	Blascheck et al. (2017)	Study of different techniques for the presentation of data generated with eye-tracking technologies
Gaze estimation	Kar and Corcoran (2017)	Overview of systems and methods for estimating the gaze vector
Experimentation	Clay et al (2019)	Exhaustive exploration of methods and tools which can be applied in the implementation of VR experiments using eye-tracking
Perception theory, display engineering and tracking technologies	Koulieris et al (2019)	Investigation about the theory of perception and vision as well as the latest advancements in tracking technologies and display engineering
Application and usability	Li et al (2021)	Literature review from 2000 to 2019 classifying results according to eye-tracking methods, application measures and assessing the usability of eye-tracking data in VR
Eye tracking measures and emergent patterns	Rappa et al. (2019)	Describe the ways in which eye-tracking devices have been used in VR and MRI environments and identify emerging patterns in the results to inform recommendations for future research
Trends	Li and Barmaki (2019)	Latest research progress in ACM Symposium on Eye Tracking Research and Applications 2019 and recent representative works
Multimodalities in VR	Martin et al. (2022)	Multimodality in VR and its role and benefits in the user experience, along with different applications that leverage multimodality across many disciplines
Rendering optimizations	Matthews et al. (2020)	Brief overview of VR and eye-tracking history, as well as the main fields of rendering optimization: traditional and perception-based approaches
	Mohanto et al (2022)	Overview of rendering work for optimizing the performance of foveated rendering. From geometrical approaches to adaptive solutions, including raster, ray, and hardware-oriented solutions
Eye-tracking in VR	Our survey	While other surveys have examined eye-tracking systems, they typically focus on specific applications such as rendering optimizations or eye-tracking data analysis. Our survey makes a significant contribution by providing an extensive review of the use, integration, and implementation of eye-tracking into VR systems

is finally introduced and organized in several expertise fields, in order to showcase the large number of articles using this technology. This survey ends with a discussion that remarks the benefits and drawbacks of this technology, and the conclusions, where the maturity level and future trends are analyzed.

1.1 Eye-tracking benchmarks

VR encompasses a wide spectrum of research fields, among which the role of computer graphics in content generation and visualization stands out. Recently, some studies have shown the benefits of using eye tracking in VR in different areas such as interaction and attention tracking. Regarding interaction, Luro and Sundstedt (2019) made a comparison between gaze aiming in VR and traditional controllers in a point-and-shoot task. They studied different target trajectories and speeds from the collected data using the system usability scale (SUS) and cognitive load questionnaires (NASA TLX). Results indicated that gaze can be used as a replacement for aiming in VR without negatively affecting task performance and comfort. Participants showed less physical demand using gaze tracking and the SUS reported similar results for both methods. Other studies such as the one conducted by Joo and Jeong (2020), proposed a user interface (UI) based on eye tracking and demonstrated that it reduces time spent on simple operations, thus avoiding dedicated controllers and their usage time. In addition, Clay et al (2019) demonstrated the usefulness of eye-tracking combined with VR exploring methods and tools that can be used in experimentation.

Regarding the study of user attention, other investigations compare the use of head tracking with eye tracking in recording user attention to different areas of interest. For example, Llanes-Jurado et al (2021) made a comparison between eye tracking and head tracking by studying multiple areas in a virtual environment. They showed that there is a high similitude between head and eye horizontal gazes and suggested a new threshold for areas of interest in virtual environments in order to compare both technologies. In another study, they presented rule-based criteria to calibrate fixation identification focused on different features (Llanes-Jurado et al 2020). Gaze tracking was also compared against controller tracking to show that the first is better suited for aiming at fast-moving targets, with faster reaction times and less physical effort while following the target path (Luro and Sundstedt 2019). In addition, Blattgerste et al (2018) compared eye-tracking-based interaction in VR and AR with head-tracking evaluating the benefits of eye-tracking. They showed that gaze-tracking outperforms head-tracking in many features such as speed, user preference, and task load.

2 Methodology

A wide variety of research articles was reviewed in this work. To this end, Scopus was used as the principal cross-library search tool. Given the purpose of this survey, the Scopus query was performed by seeking the following terms in the title with the AND operator: “eye”, “tracking”, “gaze”, “virtual”, “reality” and “vr”. With these words, the next four searches were performed:

- TITLE (eye AND tracking AND virtual AND reality): 118 results.
- TITLE (eye AND tracking AND vr): 35 results.
- TITLE (gaze AND tracking AND vr): 6 results.
- TITLE (gaze AND tracking AND virtual AND reality): 18 results.

The results from the Scopus search are depicted in Fig. 1, whereas the top ten journals publishing these documents are shown in Fig. 2. The bottom image in Fig. 2 presents the journals from which the works considered in this review come; those marked with a star have been published in the previous top ten journals, whereas blue-colored ones belong to Computer Graphics journals and orange bar agglutinates proceedings of conferences.

With eye-tracking technology rising and continuously evolving, the proposed searches were filtered to obtain studies since 2019 and omit those with discontinued devices and surpassed technology. With this filtering, a total of 134 articles were considered. Next, manual filtering was carried out to ensure the found articles made use of virtual reality and eye-tracking technologies. With this filtering, the number of finally considered studies was 112 by removing those that were unavailable or non-English written documents. From these, the bibliography was enlarged with work found in the state-of-the-art and experimentation of already included documents. This was especially relevant for custom

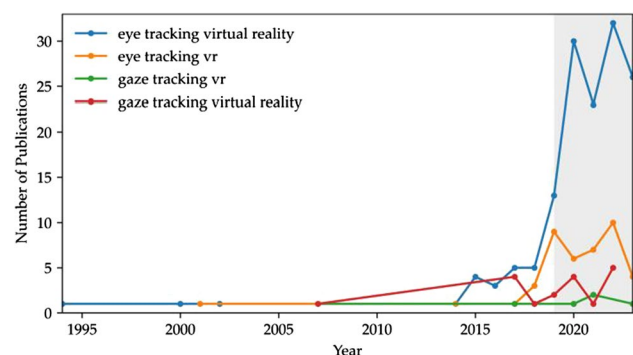


Fig. 1 Number of investigations obtained from the above Scopus queries, from 1994 to 2022

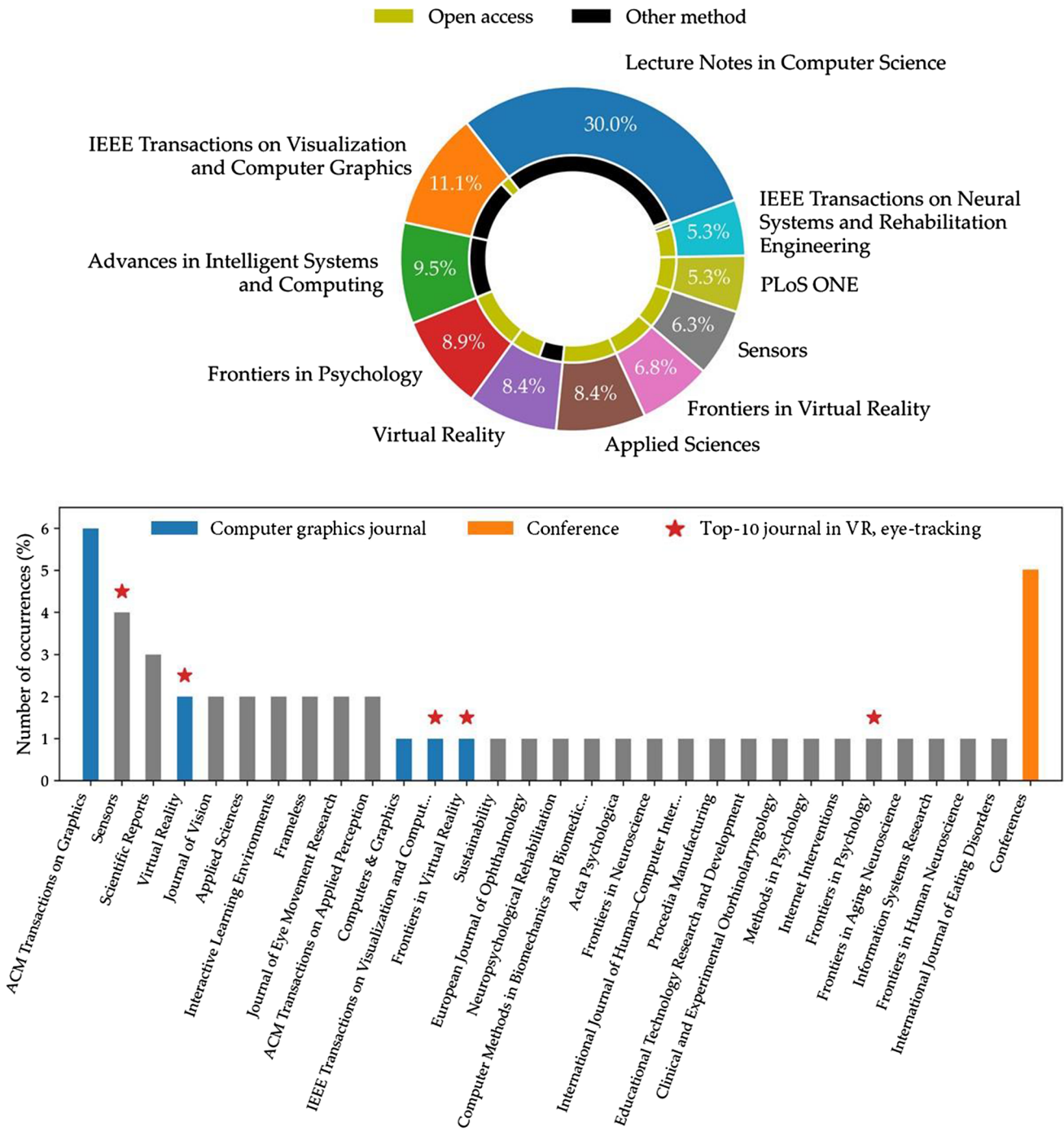


Fig. 2 On the top side, the distribution of research among the top-10 publishing journals in the VR field. The inner color refers to the publication mode. On the bottom image, the most popular journals in our

bibliography (star: top-10 journal from the above image, blue: Computer Graphics journal, orange: conferences)

eye-tracking devices, datasets and eye-tracking methodologies which establish a comparison with previous work.

In the case of rendering techniques, the Scopus search was TITLE-ABS-KEY(foveated AND rendering) since most of the included studies do not focus on VR devices

but check the user’s experience using different foveated rendering approaches. Therefore, 208 documents were found, from which 30 were finally selected according to their alignment with the topic of this survey, their publication date and relevance.

3 The visual perception

There are numerous studies that explain how human vision works by analyzing everything from the eye to the brain processes involved (De Valois and De Valois 1980; Livingstone and Hubel 1988; Wandell 1995; Palmer 1999; Li 2014). In this section, we are going to focus on eye-tracking, how it is performed and why it is useful.

First of all, we must differentiate between the terms “eye-tracking” and “gaze-tracking”, as they are often used interchangeably but denote distinct aspects of visual monitoring. Eye-tracking is a broader term that encompasses the study and measurement of eye movement. This includes the position of the eye and its movement within the socket, which can involve tracking rapid movements known as saccades, periods where the eye remains still (fixations), and the dilation and constriction of the pupil. Eye-tracking provides raw data on where the eyes are positioned and how they move. It is a comprehensive view of ocular activity without necessarily tying it to specific points of focus in the environment.

On the other hand, gaze-tracking is more specific and is concerned with determining where a person is looking in their environment or on a screen. It takes the data from eye-tracking and interprets it to provide a point or region of focus. Essentially, while eye-tracking gives you the mechanism of the eye’s movement, gaze-tracking tells you the outcome or result of that movement in terms of focus points. For instance, in a virtual reality setup, gaze-tracking would indicate what object or scene component the user is looking at.

To determine where an individual is looking, eye-tracking technology often employs a principle called corneal reflection. In this methodology; an infrared light source illuminates the pupil, generating a reflection in the cornea. An infrared camera captures this reflection and delimits the center of the pupil, deducing the rotation of the eye and determining the direction of the gaze. The location of the fovea varies in each individual due to its geometrical peculiarities and thus has to be taken into account for gaze tracking because of the lack of alignment between the optical and visual axis (see Fig. 3). For this reason, a calibration procedure is applied to optimize eye-tracking detection (Tobii 2022a, b). This eye-tracking calibration is covered in more detail in Sect. 5.1.

3.1 Eye movements

From a research perspective, eye movements in VR offer invaluable insights into a user’s attention, cognitive state, and emotional response. This data is instrumental for a

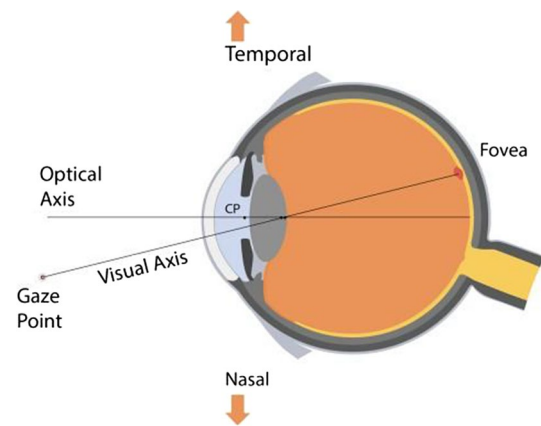


Fig. 3 A diagram illustrating how the fovea, the area of the eye responsible for sharp central vision, can shift from its normal position on the optical axis, requiring the use of calibration to accurately measure eye movements and position in eye-tracking systems (Tobii 2022a, b)

range of fields, from psychology to neurology (Just and Carpenter 1980; Rayner 1998; Jacob and Karn 2003; Leigh and Zee 2015). For VR developers, understanding gaze patterns can optimize content placement, ensuring that key elements capture users’ attention. Furthermore, in training scenarios, eye movements can evaluate a trainee’s observational skills and focus. Eye movements can be classified into three basic types (Duchowski 2017):

- *Fixations*: Occurs when the eye stops over an object or position to collect visual information. The duration of fixation is usually variable, however, the longer the fixation, the more visual information is collected and processed.
- *Saccades*: Rapid, ballistic movements of the eyes that abruptly change the point of fixation. Because saccadic movements occur at high speed, vision is impaired. This is why they are not as important in eye-tracking as fixations. However, they do reveal information about the direction of the user’s gaze and the order of fixations and visual attention.
- *Smooth pursuits*: Much slower tracking movements of the eyes designed to keep a moving stimulus on the fovea.

In addition to these basic movements, other movements can be mentioned that may be of interest depending on the study being performed:

- *Microsaccades*: Extremely small, jerk-like eye movements that occur involuntarily when a person attempts to fixate their gaze on a single point. Even when we try to keep our gaze steady, our eyes are constantly

making these minute adjustments, typically at a rate of around one to two per second.

- *Vestibular Ocular Reflex (VOR)*: An essential reflex that stabilizes images on the retina during head movement by producing an eye movement in the direction opposite to the head movement. This reflex allows us to maintain a clear visual focus on an object even when our head is moving. Discrepancies between the user's real-world VOR and the simulated visual environment can lead to feelings of discomfort or motion sickness.
- *Vergence*: A type of eye movement where the two eyes move in opposite directions (Howard 2002). This mechanism is critical for maintaining binocular vision and depth perception. There are two types of vergence movement:
 1. *Convergence*: The eyes move toward each other when viewing a close object.
 2. *Divergence*: The eyes move away from each other when viewing a distant object.

Discrepancies between real and simulated vergence movements can lead to discomfort or motion sickness.

- *Nystagmus*: A vision condition characterized by involuntary, repetitive eye movements. These movements often result in reduced or limited vision as the eyes uncontrollably oscillate in a quick, jerky, or pendular manner. Nystagmus can occur in a horizontal, vertical, or rotary pattern, and it often involves both eyes. The presence of nystagmus can significantly impact the precision and utility of eye-tracking data, as the erratic eye movements may skew gaze-tracking metrics or result in misinterpretations of gaze direction or intent.
- *Blink*: A natural and involuntary action that serves several essential functions, such as protecting the eye from irritants and keeping the eye moist by spreading tears over its surface. In an average person, blinks occur approximately every 4–6 s, or about 15–20 times per minute.
- *Ambient eye movements*: Generally associated with the early phase of visual perception and are characterized by a series of quick eye movements (or saccades) and brief fixations. They are used to get a rapid overview or 'gist' of a scene and help to orient the viewer in their surroundings.
- *Focal eye movements*: Come into play after the initial ambient phase and involve longer, more deliberate fixations. Focal eye movements are used for detailed examination of objects or features of interest within a scene.

3.2 Conditions of testing

When analyzing eye-tracking data, conditions of testing must be taken into account. This refers to the specific parameters, environment, and controls established for a test or experiment. Properly defined conditions are crucial to ensuring that the results of a test are valid, reliable, and can be replicated. Depending on the context (whether it is a scientific experiment, product testing, clinical trials, or others), these conditions can vary. In this regard, when eye-tracking systems are used, we can speak of two conditions:

- *Head-free*: refers to setups or systems that allow the user's head to move freely without constraining it to a fixed position.
- *Head-still*: typically refers to a condition or requirement in which a participant or subject is asked to keep their head stationary or motionless.

In VR setups, the "head-free" condition is predominant. This approach is favored because it enhances immersion and naturalism, is aligned with the head-tracking capabilities of VR systems, and increases user comfort by reducing motion sickness. Additionally, it supports interactivity, allows full 360° experiences, and distributes movement effort between the eyes and neck, reducing strain. While specific studies might occasionally require limited head movements, such as Sipatchin et al (2020) or Sipatchin et al (2021), general VR applications prioritize a head-free experience for a more comprehensive and immersive user engagement.

This testing condition introduces a new dimension of tracking: head tracking. While eye and gaze tracking concentrate on the eyes and where they focus, head tracking is a broader technology that captures the position, orientation, and movements of a user's head in real-time. This becomes especially relevant in environments where the whole orientation of the viewer's perspective can change based on the movement of their head. Head tracking is crucial in determining how a user is physically orienting themselves within a space. In immersive environments like VR or AR, head tracking ensures that the visual display adapts to the user's head movements, offering a 360-degree perspective. For instance, if a user looks up or turns their head to the side in a VR simulation, the visual scene will adjust accordingly, giving the sensation of "looking around" within the virtual space.

3.3 Metrics

The aim of measuring and analyzing eye movements is to study the user's attention, how he/she distributes it and what determines this distribution, which can be called attention tracking. Attention tracking in VR is a more comprehensive

measure, aiming to gauge the depth of a user's cognitive engagement. It is not just about where users are looking, but how engrossed they are. By merging eye movement data with other metrics, such as the ones listed below, VR systems can deduce how captivating a particular scene or object is for the user (Duchowski 2017).

The following are some of the metrics used in the different studies analyzed in this paper:

- *Fixation count*: Refers to the number of fixations carried out per area of interest (AOI).
- *Time to First Fixation (TTFF)*: A metric used in eye-tracking research to measure the time it takes from the onset of a stimulus to the moment the viewer's gaze fixates on a particular point or area of interest for the first time.
- *Duration of First Fixation (DFF)*: Represents the length of the viewer's initial gaze fixation upon spotting a particular stimulus or area of interest.
- *Transition order between fixations*: Denotes to the sequence or order in which a viewer's gaze moves from one point of fixation to another.
- *Dwell-time*: In gaze analysis, this term represents the duration spent focusing on a single object or position. Its computation relies on:
 1. Identification of a fixation
 2. A temporal window determining a threshold for the fixation's duration.
- *Total fixation duration*: Refers to the time period for their fixation points falling in a certain AOI including the duration of the first fixation.
- *Saccade count*: Quantifies the number of rapid eye movements or shifts made during observation.
- *Saccadic velocity*: Denotes the rate at which the eyes transition during a saccade.
- *Amplitude of saccades*: Measures either the angular or linear distance traversed by the eyes during a saccadic movement.
- *Reaction time*: The interval required for a saccade to commence following the display of a stimulus or cue.
- *Search time*: As gauged using eye-tracking methodologies, this metric determines the duration needed for a participant to visually identify a target or specific area of interest.
- *Microsaccade count*: Represents the tally of minute, involuntary eye movements observed during focused gaze.
- *Fixation/Saccadic ratio*: This metric, commonly employed in eye-tracking research, examines the relationship between phases of ocular stability (fixations) and swift eye motions (saccades). The ratio elucidates how observers assimilate visual data, offering clues about their cognitive condition or the intricacies of their task.
- *Blink rate*: Denotes the frequency of a participant's blinking per minute. It can be used to measure fatigue (Stern et al 1994), engagement (Ranti et al 2020) or cognitive load (Biondi et al 2023), among others.
- *Blink duration*: This metric quantifies the duration during which an individual's eyes remain closed in a blinking episode. It serves analogous purposes to the blink rate.
- *Eye status*: Distinguishes between phases when the eye is open and when it is closed.
- *Gaze direction*: Reflects the eyes' alignment or orientation concerning a specific focal point. It reveals an observer's current visual attention or point of interest.
- *Gaze position*: It is the specific point in space or on a surface where a person's eyes are directed or focused. It represents the spatial location that a viewer is currently looking at.
- *Gaze velocity*: Measures the speed at which one's gaze, or point of focus, changes position.
- *Gaze acceleration*: Denotes the rate at which the speed of one's gaze changes. Just as gaze velocity measures how quickly eyes move from one point to another, gaze acceleration measures how fast this velocity changes, either increasing or decreasing.
- *Eye position*: Illuminates the eyes' orientation or position concerning a reference, be it the observer's head, a display, or an external scene. It denotes the eyes' spatial arrangement and alignment at a given juncture.
- *Pupil diameter*: It can be used to determine pupil dilation/contractions which can determine strong emotional stimuli, acute attention and working memory load (Slovak et al. 2022; Duchowski et al 2020).
- *Pupil position*: Indicates the exact spatial location of the pupil.
- *Ocular deviation*: Pertains to the eyes' misalignment, meaning that the two eyes do not point exactly in the same direction. It is commonly seen in conditions known as strabismus or squint.
- *Spatial accuracy*: Within the realm of eye-tracking, this denotes how precisely the system can determine where a user is looking. It is usually defined as the difference or error between the position recorded by the eye tracker and the actual position of the user's gaze in the real world.
- *Transition entropy*: A metric in eye-tracking analysis to quantify the predictability or randomness of a person's gaze transitions between different areas or points of interest. It can be used as a measure of visual scanning efficiency (Shiferaw et al 2019).
- *Spatial distribution*: Illustrates the spread or arrangement of gaze points or fixations across a particular visual field

or area of interest. In other words, it describes where the user tends to look in a scene or interface.

- Distance between users' gaze and a target: Measures the spatial gap between a user's gaze and a specific target.
- *Visual field*: Represents the full extent of the area that can be seen when the eye is directed forward, encompassing the central and peripheral vision.
- *Near Point of Convergence (NPC)*: The closest point in space to which both eyes can direct their gaze before one or both eyes begin to turn outward, losing binocular alignment. In other words, it is the point at which your eyes can no longer maintain a coordinated focus on a near object, and one eye "breaks" or deviates from the target.
- *Positive Fusional Vergence (PFV)*: Also known as "convergence reserves", represents the ability of the eyes to turn further inward (converge) than is necessary for binocular single vision. It is a measure of the extra convergence capacity the visual system has beyond what's currently being used for a given task.
- *Near/far dissociated phoria*: Gauges the eyes' resting orientation when they are not synchronized to focus. It essentially describes the tendency of one eye to drift either inward (esophoria) or outward (exophoria) when the other eye is covered or when binocular vision is otherwise disrupted.
- *Convergence Insufficiency Symptom Survey*: A validated instrument used to quantify symptoms associated with Convergence Insufficiency (CI), a common binocular vision disorder characterized by the eyes' inability to work together efficiently at near distances.
- *Eye-tracking delay*: Also known as "latency", is the time interval between the occurrence of an eye movement or gaze event and the system's ability to detect, process, and potentially respond to it.
- Conditions of testing.

Table 2 summarizes the most commonly used units for each one of these metrics.

4 VR headsets

In this section, we provide an overview of eye-tracking systems within VR, exploring both custom-built and commercial platforms.

Custom systems emerge as intriguing alternatives in the eye-tracking domain, primarily for their potential cost-effectiveness. The research spotlight in the realm of custom systems shines on three main facets: (1) the construction of the hardware architecture, (2) the development of software aimed at providing the user's gaze vector and (3) the assessment of deviation between ground truth and computed results.

On the commercial side, established devices offer ready-to-use solutions for a wide array of eye-tracking applications. But like any evolving technology, they present their own set of challenges and areas for improvement. Highlighted by studies such as (Llanes-Jurado et al. 2021; Borges et al. 2018), and the investigation conducted by Sipatchin et al. (2021) on the HTC Vive Pro, it is evident that there is a discernible gap between real-world performance and manufacturer claims. For instance, the HTC Vive Pro's spatial accuracy was put to the test in an ophthalmological context using an online virtual perimetry testing application. Two distinct testing conditions were employed: head-still, to assess eye-tracking accuracy across a vast visual field, and head-free, to evaluate the effects of head movements on eye-tracking precision and potential data spillage. The results were illuminating, revealing that the spatial accuracy was not as pristine as manufacturer specifications indicated and that head movements introduced a drop in precision and an increase in data loss.

Accordingly, this section is structured to first explain which are the most frequent commercial headsets, whereas the final subsection is devoted to custom systems that built their own architecture to track the user's line of sight.

4.1 Commercial headsets

Eye-tracking on VR can be approached using notable commercial devices, as depicted in the exhaustive compilation of commercial VR headsets provided in Table 3. These can be integrated into the VR headset, e.g., HTC Vive Pro Eye, or used isolated. Despite the number of solutions combining both technologies having significantly increased, only a few of them fuse both features. However, the vast majority of research concerning eye-tracking nowadays is based on VR-integrated solutions. It represents 57.6% of the reviewed articles published in 2021 (Fig. 4), whereas research from 2019 was mainly dominated by custom and isolated eye-tracking devices such as SensoMotoric Instruments (SMI). On the other hand, integrated solutions barely achieved the 6% of reviewed manuscripts in 2019, whereas it stood for the 28.8% of revised research from 2020.

Besides integrated devices, non-commercial solutions built from scratch are also frequently evaluated (which will be referred to as custom devices from now on). These solutions are harder to use as they require planning and building the hardware architecture, as well as developing the software for eye-tracking and calibration. Most of them use commercial VR headsets as the underlying infrastructure for the eye-tracking system (Table 4), however, recent work has also built custom VR devices that even enable improving headset comfort (Altobelli 2019).

Table 2 Eye-tracking metric units

Metric	Unit
Fixation count	Integer
Time to First Fixation (TTFF)	Milliseconds (ms)
Duration of First Fixation (DFF)	Milliseconds (ms)
Transition order between fixations	Sequence (A-B-C)
Dwell-time	Milliseconds (ms) or Seconds (s)
Total fixation duration	Seconds (s)
Saccade count	Integer
Saccadic velocity	degrees per second ($^{\circ}/s$) or pixels per second (px/s)
Amplitude of saccades	Millimeters (mm) or degrees ($^{\circ}$)
Reaction time	Milliseconds (ms) or seconds (s)
Search time	Milliseconds (ms) or seconds (s)
Microsaccade count	Integer
Fixation/Saccadic ratio	–
Blink rate	Blinks per minute
Blink duration	Milliseconds (ms)
Eye status	Binary (open/closed)
Gaze direction	Coordinates [x,y]
Gaze position	Coordinates [x,y,z]
Gaze velocity	Degrees per second ($^{\circ}/s$) or Pixels per second (px/s)
Gaze acceleration	Degrees per second ² ($^{\circ}/s^2$) or Pixels per second ² (px/s ²)
Eye position	Coordinates [x,y,z]
Pupil diameter	Millimeters (mm)
Pupil position	Coordinates [x,y]
Ocular deviation	Degrees ($^{\circ}$)
Spatial accuracy	Degrees ($^{\circ}$) or percentage (%)
Transition entropy	–
Spatial distribution	–
Distance	Millimeters (mm)
Visual field	Degrees ($^{\circ}$)
Near point of convergence (NPC)	Centimeters (cm)
Positive fusional vergence (PFV)	Prism diopters (Δ or PD)
Near/far dissociated phoria (PFV)	Prism diopters (Δ or PD)
Convergence insufficiency symptom survey	–
Eye-tracking delay (latency)	Milliseconds (ms)
Conditions of testing	Head-free or head-still

4.2 Custom systems

Custom systems are mainly based on infrared (IR) cameras, as they enable acquiring eye data in the absence of lighting, which occurs in cave-like environments such as VR headsets. Nevertheless, they are also approached using cheap visible cameras, such as the ones integrated into mobiles, since they are focused on low-cost setups. Then, acquired data is frequently transformed using traditional image processing algorithms, mostly focused on pupil and iris detection. Image processing pipelines are mainly classified into model-based and feature-based methods whether they seek relevant points on the images (features) or shapes (model). The proposed systems are finally evaluated considering the

difference between the estimated and ground truth gaze vector.

The base hardware for custom devices is mainly given by commercial VR headsets without eye-tracking integration, even reverse-engineered and 3D printed (Altobelli 2019), though a wide range of headsets can be found in the literature (Fig. 5). This includes HTC Vive (Dong et al. 2020; Chugh 2020) and BOE (Sun et al. 2021) head-mounted displays (HMD), inexpensive plastic cases for mobile phones (Drakopoulos et al. 2020, 2021), and other headsets designed specifically for case studies such as Magnetic Resonance Imaging Systems (MRI) (Qian et al. 2021) (see Fig. 6). As previously mentioned, eye-tracking architectures are mainly composed of IR cameras and light

Table 3 Commercial headsets that integrate eye-tracking

Model	Screen	Resolution	Refresh rate	FOV	Sensors
Vive Pro Eye	Dual OLED	1440×1600	90 Hz	110°	SteamVR tracking Eye tracking G-sensor Gyroscope Proximity IPD sensor
FOVE0	WQHD OLED	2560×1440	70 Hz	100°	IMU Eye tracking
Meta Quest Pro	LCD	1920×1080	90 Hz	106°	Eye tracking Face tracking
PlayStation VR2	OLED	2000×2040	90–120 Hz	110°	g IMU Proximity Eye tracking

Resolution is represented in pixels per eye (width×height), and FOV refers to the horizontal field of view. IMU stands for Inertial Measurement Unit

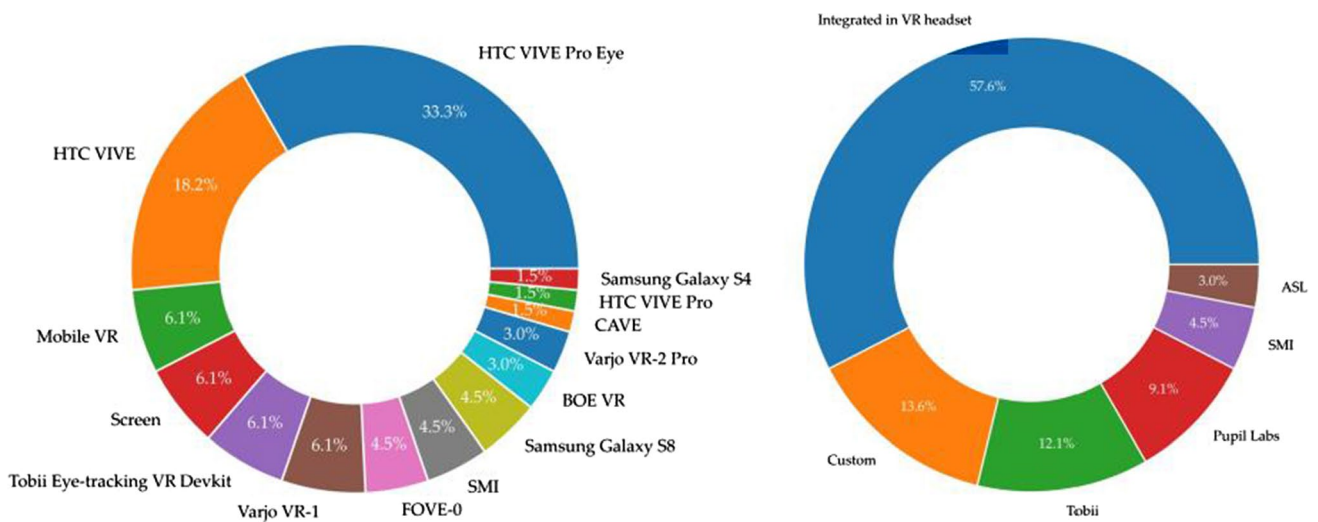


Fig. 4 Most popular virtual reality headsets and eye-tracking devices used in 2021. The left image shows the VR headsets, while the right image shows the eye-tracking devices

sources to perform video-based oculography (VOG). In the absence of natural lighting, IR cameras allow acquiring the participant eye with different grayscale values for pupil and iris (Sun et al. 2021). The purpose of IR light sources is to generate recognizable reflections on the eyeball (Purkinje reflection points), thus helping to estimate the line-of-sight direction (Dong et al. 2020). The number of IR light sources goes from one (Sun et al. 2021; Qian et al. 2021; Dong et al. 2020; Katrychuk et al. 2019) to eight (Lu et al. 2020), forming a ring shape which can be later identified through shape-fitting. However, a large number of IR light sources is reported to negatively affect participants (Qian et al. 2021; Dong et al. 2020) since they increase the risk of causing an undesired red-eye effect, besides causing more specular reflections; therefore, these

flaws influence pupil tracking accuracy and robustness. This is of particular concern when the illumination is placed very close to the face. Rather than creating an eye-tracking architecture from scratch, Chugh (2020) opted for estimating the gaze vector using Pupil Labs hardware. The same hardware also comes with a software application to access the collected images. Yet, a key challenge with prior wireless headset solutions like Google Daydream is their high demand for processing power. To address this issue, Photosensor Oculography (PSOG) has been proposed. This technique measures the IR reflection using a limited number of IR detectors. Nevertheless, an important obstacle lies in sensor shift, which significantly deteriorates spatial accuracy (Katrychuk et al. 2019).

Table 4 List of VR headsets that can work as the underlying infrastructure for custom eye-tracking systems and do not integrate eye-tracking by default. Resolution is represented in pixels per eye (width × height), and FOV refers to the horizontal field of view

Model	Screen	Resolution	Refresh rate	FOV	Sensors
Vive Focus 3	Dual LCD	2448 × 2448	90 Hz	120°	4 tracking Cameras G-sensor Gyroscope Proximity
Vive Focus Plus	Dual AMOLED	144 × 1600	75 Hz	110°	Front facing camera
Vive Pro	Dual AMOLED	144 × 1600	90 Hz	110°	SteamVR Tracking G-sensor
Vive Pro 2	Dual RGB LCD	2448 × 2448	90 120 Hz	120°	G-sensor Gyroscope Proximity IPD sensor SteamVR tracking 2.0
Valve Index	Dual LCD	1440 × 1600	80 Hz 90 Hz 120 Hz 144 Hz	130°	SteamVR tracking 2.0
Vive Cosmos	Dual LCD	1400 × 1700	90 Hz	110°	G-sensor Gyroscope IPD sensor
Oculus Go	Dual LCD	1280 × 1440	60 Hz	90°	Orientalational tracking
Oculus Quest	Dual OLED	1440 × 1600	72 Hz	90°	4 tracking cameras
Oculus Rift S	Dual LCD	2560 × 1440	80 Hz	90°	6 tracking cameras
Oculus Quest 2	Single Fast switching LCD	1832 × 1929	60 Hz 72 Hz 90 Hz	90°	4 tracking cameras

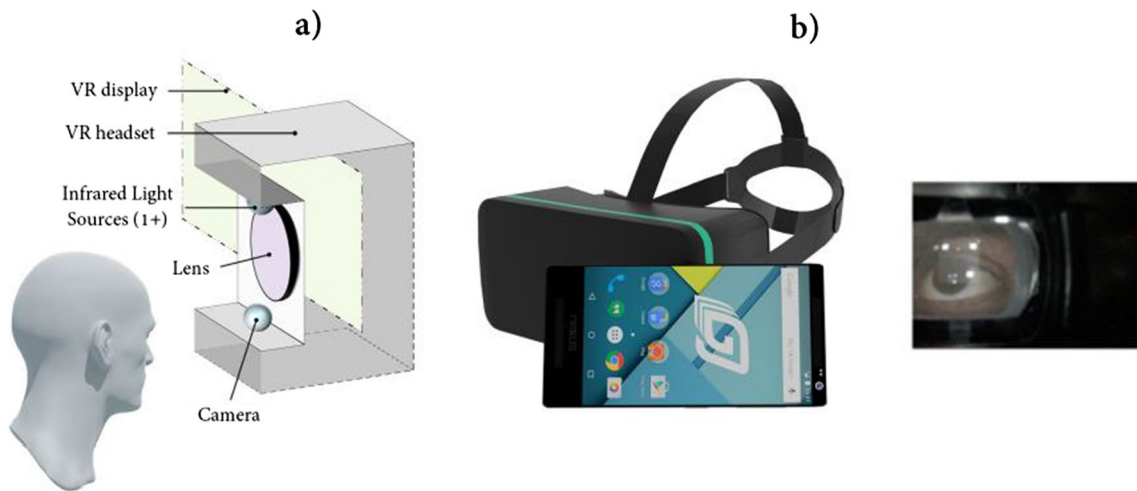


Fig. 5 Comparison of frequent infrastructures for building a custom eye-tracking system. **a** A cross-section of a VR headset coupled with IR lights and a camera for each eye (Sun et al 2021), and **b** a VR

headset for mobile phones, similar to the one employed by Drakopoulos et al (2021)

5 Eye-tracking

On the basis of previously revised devices, this section intends to further explain how is the calibration and eye-tracking stages conducted on them. The following works do not always operate over custom devices, but they implement processing layers over data provided by an eye-tracking system, either custom or commercial.

This section is structured as follows: first, Sect. 5.1 details how the system must be calibrated in a similar manner to commercial devices to ensure accurate eye-tracking. Once configured, Sect. 5.2 explains the algorithms used to track eye movement. These algorithms range from traditional detection methods, based on image processing and the recognition of eye regions and reflections caused by external and controlled light sources, to Artificial Intelligence (AI) algorithms. The first methods present a high computational

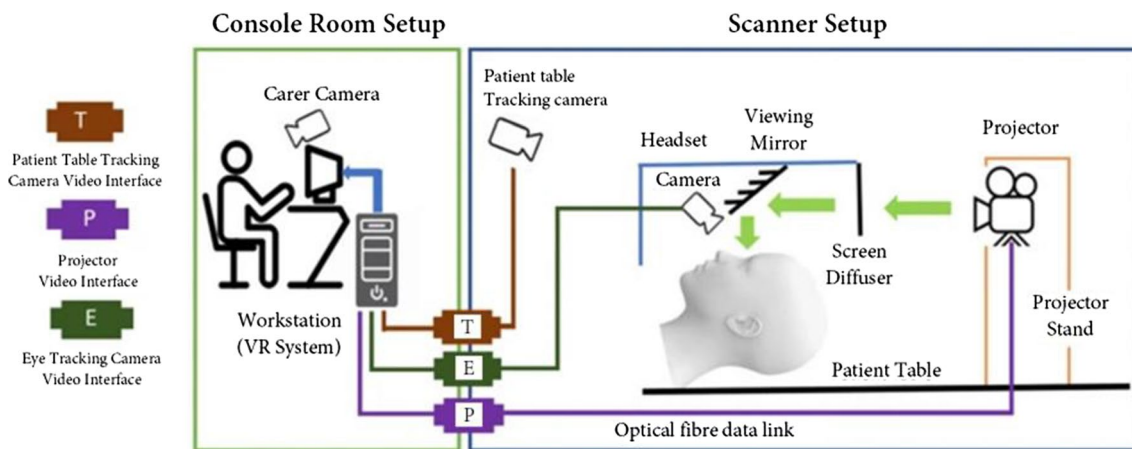


Fig. 6 Eye-tracking architecture within an MRI scanner, as proposed by Qian et al (2021)

cost for real-time tracking and thus have led to AI-based methods that require previous learning. Accordingly, the final subsection is dedicated to datasets for eye-tracking applications.

5.1 Calibration

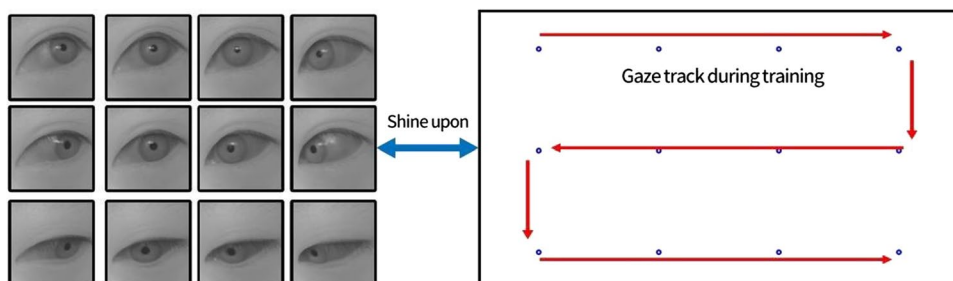
To accurately estimate the gaze vector, these devices are initially calibrated per participant by displaying multiple uniformly distributed points, whose gaze vector is known a priori (Fig. 7). Most studies present calibration methods based on nine points (Qian et al. 2021; Chugh 2020; Lu et al. 2020; Li et al. 2019), although Qian et al. (2021) used fifteen points to select the most appropriate fitting function on a single participant. However, increasing the number of points beyond twelve has been reported to not yield any meaningful improvements (Drakopoulos et al. 2021). Regarding the gaze mapping model that correlates screen coordinates and pupil coordinates with the acquired image, previous research has extensively used linear (Drakopoulos et al. 2021; Dong et al. 2020) and second-order polynomials (Sun et al. 2021; Lu et al. 2020; Li et al. 2019), as high order polynomial models show little improvement. However, quadratic and cubic polynomial models have been successfully applied to case studies where participants have a static pose (Qian et al. 2021). Besides pupil coordinates, other values can also be

integrated into these polynomial models to account for head motion (Qian et al. 2021). During this process, coefficients from mapping functions vary; first, they can be initialized to average expected human values. Then, they can be optimized through algorithms such as Least Squares to minimize the distance between the calibration targets and calculated gaze locations (Chugh 2020; Lu et al. 2020). Some of the calibration measurements may be discarded whether they are considered outliers through statistical analysis, thus avoiding inaccurately estimating the model coefficients (Chugh 2020).

5.2 Eye-tracking

Once calibrated, images acquired either from RGB or IR cameras, as well as numerical data from headset sensors must be processed to estimate the origin of the gaze vector. The most frequent pipelines in computer vision for estimating the gaze vector are feature-based and model-based methods, besides AI ones. Feature-based procedures emphasize the finding of features in images, e.g., the iris. These are mainly guided by simple image transformations, such as changing the image intensity, thresholding and morphological operations. The output is typically a binary mask with the location of the target feature if found. These are known to be more dependent on specific devices, as intensity-based pipelines are more sensitive to changes regarding lighting. On

Fig. 7 Calibration procedure followed by Dong et al (2020), following a predefined path



the other hand, model-based methods are intended to look for specific shapes, such as circular objects. Unlike feature-based methods, the output of this variant is the geometry of a specific part of the eye. However, most of the methods in the latter category frequently perform pre-processing operations similar to those belonging to the feature-based category. Therefore, both kinds of algorithms end up being affected by the recording conditions. In addition, model-based algorithms are more time-consuming due to the shape-fitting phase, at the expense of being more robust and precise.

Otherwise, images and data collected from eye-tracking can be processed with AI algorithms that either extract relevant eye parts from the image or directly estimate 2D screen locations and 3D gaze vectors. Nevertheless, most of these operate as supervised algorithms that require previously annotated datasets that must be representative enough to operate over people from different demographic groups (sex, age, race, etc.) and accessories such as glasses. Furthermore, AI-based techniques require further computational resources for training. In this regard, previous work has investigated the use of simpler networks to operate over lower consumption devices such as Raspberry Pi 3 (Katrychuk et al. 2019).

There are other kinds of algorithms that have not been included since the most recent works date to one decade ago. For instance, shape-based methods use deformable eye templates that must be fitted with an actual human eye. Others, such as the cross-ratio category described in Kar and Corcoran (2017), have been included in feature-based and model-based categories as they intend to find the projected light sources on the eye.

In summary, feature-based methods are known to be more dependent on recording conditions (observe Fig. 8), due to computer vision processing, although they are also less time-consuming. Model-based works are slightly more complex and typically rely on intensity processing techniques as well. Finally, AI-based studies are not as time-consuming in real-time, but they require training datasets captured in conditions similar to those found in a case study. A challenge in common for all of them is to make them robust to different demographic groups. More insight into these categories is provided in Table 5, where the accuracy of the following revised documents is reported.

Fig. 8 Lens reflection observed by (Drakopoulos et al. 2021), thus hardening the image processing

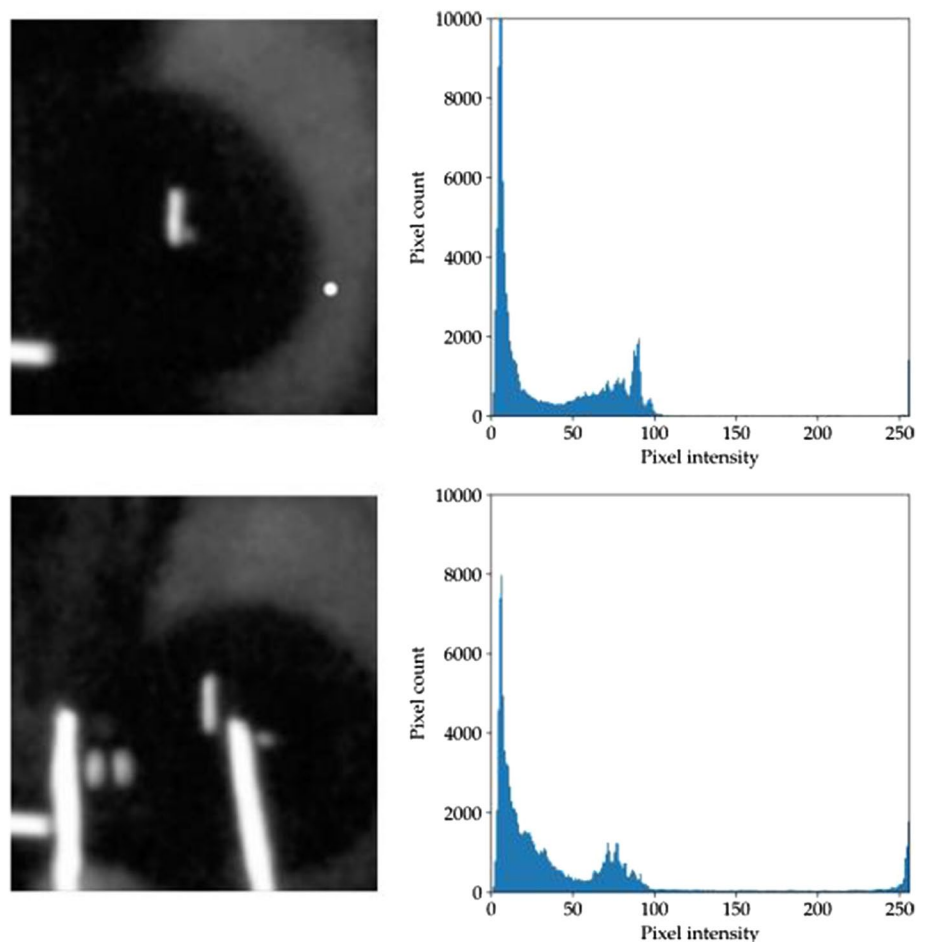


Table 5 Classification of revised eye-tracking methods according to the type, the input data, the recording set-up and the reported results

Category	Input	Conditions	Accuracy	Work
Feature-based	Infrared images of a single eye	Head-free, one IR light source	1°	Dong et al (2020)
	Front-facing mobile camera	Head-free, mobile lighting	1.17°	Drakopoulos et al (2021, 2020)
Model-based	Infrared images of a single eye	Head-free, one IR light source	0.59°	Sun et al (2021)
	Infrared images of both eyes	Head-still due to MRI scenario, one IR light source per eye	0.76°	Qian et al (2021)
Deep learning	Infrared images of a single eye	Head-free, eight IR light sources	1.08°	Lu et al (2020)
	Infrared images of both eyes	Head-free, five IR light sources per eye	0.55°	Katrychuk et al. (2019)
	HMD rotation and gaze data	OpenNEEDs dataset	≥ 0.58°	Illahi et al (2022)
	Infrared image of a single eye	Head-free, two IR light sources	1°	Chugh (2020)
	RGB nearfield eye image	Head-free, natural lighting and desk lamp	2.43°	Ou et al (2021)
	Infrared images of both eyes	Head-still, two Pupil Labs cameras	4.38°	Lu et al (2022)
	Semantic labels of eye images	OpenEDS dataset	f1: 99.3%	Chaudhary et al (2019)
	RGB image containing both eyes from laptop camera	Head-free, natural lighting	Accuracy: 91.30%	de Lope and Grana (2022)
	Infrared image of face and both eyes	Head-free, natural lighting, six infrared light sources	Accuracy: 92.8%	Naqvi et al (2018)
	RGB image of the face and both eyes, head pose (angles)	GazeCapture dataset	3.05 cm	Wong et al (2019)
	RGB face image, images containing both eyes and individual eye images	MPIIFace Gaze and EYEDIAP datasets	4.21 cm	Gudi et al (2020)

The latter are either annotated in degrees, metric units or metrics typically used in DL and ML (accuracy, f1, etc.)

5.2.1 Feature-based

A naive approach is based on the finding of the iris, from where the gaze vector can be cast. To this end, feature-based algorithms detect key features with the help of intensity values. Accordingly, the pupil is known to be the darkest element within the eye region (Drakopoulos et al. 2021). Captured images can be slightly enhanced for later detection through the suppression of reflections, defined as image regions with abrupt intensity peaks. To achieve this, bright regions are averaged with their neighbors. Due to their low contrast, an histogram equalization such as Contrast Limited Adaptive Histogram Equalization (CLAHE) can also be applied (Drakopoulos et al. 2020, 2021). Concerning the procedure core, Dong et al. (2020) proposed to enhance the image contrast, apply a Haar-cascade eye detector (implemented in OpenCV) and threshold the image to output a binary image. Then, connected graphs are calculated and wrapped in a convex hull. The center of the largest connected component is considered to be the target feature within the eye pupil. Finally, Drakopoulos et al. (2020, 2021) described a variant of the traditional Hough transform, accelerated in 2D and focused on circular features to detect the iris. Then, extracted circular shapes are assigned a confidence metric using a linear transformation combining two visual features and weights

extracted from experimentation. Therefore, challenging conditions lead to lower confidence values.

Despite feature-based methods being more efficient, most of them cropped the images in real-time to reduce their size. The area to be cropped is determined either by the Haar-cascaded filter or in the calibration process. Hence, a safe eye area can be represented through a rectangle-shaped Region of Interest (ROI) whose corners are given by the minimum and maximum coordinates of detected iris center points (Drakopoulos et al. 2021).

5.2.2 Model-based

Another relevant procedure is based on the finding of ellipses representing the iris or pupil. A frequent step prior to such a fitting process is the binarization of the image to extract features with elliptical shapes. Sun et al. (2021) seeks the image area with the lowest average grey value, which is binarized to perform ellipse fitting on the pupil. For a single IR light source per eye, Qian et al. (2021) propose to segment the pupil by applying adaptive intensity thresholding, dilation and erosion, edge detection and ellipse fitting. The main objective with multiple IR light sources is to binarize their reflection and therefore apply the ellipse fitting on such points (Lu et al. 2020) (Fig. 9). However, they enhance the fitted ellipse with image processing, with the pupil center

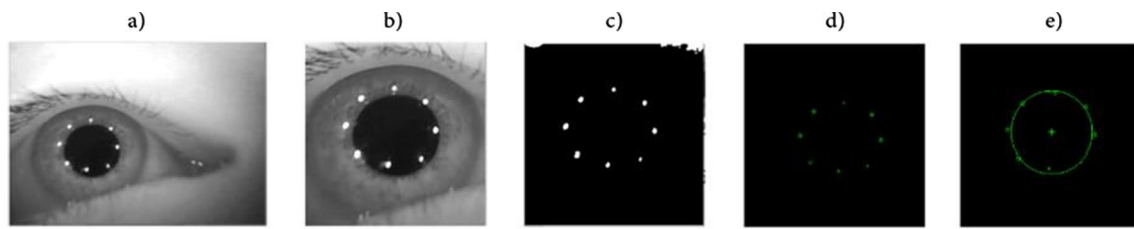


Fig. 9 Ellipse fitting process of Lu et al (2020). From left to right: initial image cropping, cropping after pupil detection, binarization, detection of light spots and ellipse fitting

being detected using a combination of morphological operators (dilation, erosion), smoothing, watershed starting from the darkest area (pupil) and edge detection. Hence, the first ellipse is adjusted using the Least Squares method and such polygon.

Regarding the efficiency of model-based methods, Sun et al. (2021) did not crop a predefined area, instead, the image was partitioned into smaller windows and only the one containing the pupil is further processed. The eye corners, and thus its bounding box, can also be manually marked during calibration and tracked in the following frames using state-of-the-art methods such as Discriminative Correlation Filter with Channel Spatial Reliability (DCF-CSR) (Qian et al. 2021).

5.2.3 Machine learning and deep learning

The increasing use of AI has also favored the proliferation of gaze-tracking works using Deep Learning (DL) and Machine Learning (ML) over infrared and visible images. More specifically, Convolutional Neural Networks (CNN) are the most widespread networks applied to gaze tracking and segmentation of eye parts (Chugh 2020; Katrychuk et al. 2019; Fuhl et al. 2020; Kothari et al. 2022), achieving the top-most accuracy in Table 5. Yet, calibration is required unless working with publicly available datasets.

The most frequent procedure is to use head or eye images to estimate the gaze vector, especially for laptop applications without 3D interactions. Wong et al. (2019) proposed a ResNet network for inferring the gaze vector, using as input a face image with color normalization and the head pose, given by yaw, roll and pitch angles. In the case of driving, the location is not always as accurate and instead, previous works have defined relevant gaze areas that must be estimated. Naqvi et al. (2018) created their own driving dataset with 17 different gaze zones and used a CNN with a significant number of convolutions to estimate such zones. On the other hand, Illahi et al. (2022) proposed to predict gaze targets in real-time by solely using the normalized X, Y screen coordinates as well as the gaze velocity, head rotational velocity, gaze acceleration and head rotational acceleration. These variables were the input of a Recurrent

Neural Network (RNN). Katrychuk et al. (2019) compared a Multilayer Perceptron (MLP) against a shallow CNN, using several configurations from low-power to high-power. These set-ups are intended to optimize real-time tracking in low-consumption systems such as Raspberry Pi 3. The shallow CNN obtained better results with the high-power configuration (deviation of 0.55°), as expected; however, other configurations may be preferred for speeding up the training and real-time tracking.

Other works are intended to estimate the pupil location from imagery, supported by previous calibrations that help to transform the detected location into a gaze vector (Ou et al. 2021). Ou et al. (2021) used the YOLOv3 network to predict the pupil's center using their own dataset of nearfield visible images. Otherwise, it is possible to conduct semantic segmentation over images, rather than solely detecting the pupil location. Chaudhary et al. (2019) proposed a U-Net-like network, in contrast to SegNet, to perform semantic segmentation over the OpenEDS dataset. Similarly, Kothari et al. (2022) trained an encoder-decoder network, DenseEINet, using multiple publicly available datasets, concluding the use of multiple of them effectively helped to obtain better results. Chugh (2020) described a simple CNN with feature upsampling and downsampling to estimate pupil location from infrared images of individual eyes, obtaining a mean distance of 1.2 pixels from the expected output. Lu et al. (2022) trained a simple CNN with 3 layers to de-refract eye images and shape the pupil's ellipse with five parameters (center, axes and tilt angle). They achieved an error of at least 2 mm in the estimation of the 3D pupil.

Another not-that-frequent set-up is to use webcams. In this case, the expected output is typically the pixel at which the user is looking. Gudi et al. (2020) split their methodology into two steps: (1) estimate the gaze vector from input images and (2) transform gaze vectors into gaze locations in the screen. The first step was solved with a VGG16 pre-trained network, whereas the second step was solved in three manners: (1) estimating the coefficients for translating one space into another, (2) using ML and (3) a hybrid method, where ML is used to estimate the coefficients. The latter outperformed the other two, while still obtaining errors above 4 cm. de Lope and Granã (2022) checked several

pre-trained networks over their own dataset, extracted from a laptop camera, and found that DenseNet had the highest accuracy (91.30% with multi-user dataset).

Rather than making DL intricate for the general public, there exist frameworks able to optimize the network and its hyperparameters (Bublea and Căleanu 2020). With this approach, a shallow CNN was used to obtain an accuracy of 85% over the Columbia Gaze dataset.

5.3 Eye-tracking datasets

There is a plethora of research concerning the publication of eye-tracking datasets for the intensive training of ML-based solutions. The main concern of gaze prediction datasets is to cover the population with a wide range of physical features, based on different genders, ethnicities, eye color, age and accessories (e.g., glasses, makeup, etc.). Eye-tracking datasets can be classified according to the task that participants carry out during acquisition, mainly split into real-world

and elicited tasks. Despite this, they do not present patterns regarding the dataset properties. Following the classification of Palmero et al. (2021), datasets are mainly divided according to their illumination, sampling frequency, image resolution, number of participants, number of image sequences, annotation and whether head movements were allowed or not, as regarded in Table 6. Note that a considerable number of previously revised works construct their own datasets; however, these are typically small in contrast to the ones that will be following presented.

In this way, Palmero et al. (2021) presented an outstanding dataset to assess both gaze prediction and sparse segmentation in AR and VR fields. As part of the prediction challenge, two different datasets were published. The first dataset consists of sequences of images from 87 subjects that were asked to gaze at specific dot patterns, thus allowing to collect saccade and fixation eye movements and the corresponding ground-truth vector. The obtained images were curated by avoiding blinks, incorrect detections, subject

Table 6 Summary of publicly available eye-tracking datasets, regarding their illumination, sampling frequency, image resolution, number of participants, allowed head movement, number of images and provided annotations

Dataset	Illumination	Freq	Resolution	Part	Head mov	Images	Annotation
OpenEDS Palmero et al (2021)	Continuous (VR)	100	640×400	87	No	550 k (Gaze) 30 k (Seg.)	Eye region; iris and pupil; segmentation masks; 3D gaze vector; gaze point; cornea center
Fuhl and Kasneci (2021)	Continuous (natural)	25	192×144	19	Yes	800 k	Pupil and eyelid center; distance between two frames; opening of eyelid
Chugh (Chugh 2020)	Continuous (VR)	NA	640×480	10	Yes	40 k (incl. part of (Kim et al. 2019))	Pupil center; corneal reflections
GW (Kothari et al. 2020)	Continuous (indoor)	120	640×480	19	Yes	Video sequences	Eye-in-head coordinate; type of eye movement; 3D gaze vector
NVGaze (Kim et al. 2019)	Varying (indoor)	120	640×480	30 (real) 10 (synthetic)	No	2 M (synthetic) 2.5 (real)	Real Dataset: 2D gaze vector; pupil position; blink labels
LPW (Tonsen et al. 2016)	Continuous (natural/artificial)	95	640×480	22	N.A	130 k	Pupil center; segmented masks
PoG (McMurrough et al. 2012)	Continuous (indoor)	29.97	768×480	20	Yes	Video sequences	Gaze point; head location
(Garbin et al. 2020)	Continuous (VR)	200	640×400	152	Yes	356 k	Gaze point Head location
UnityEyes (Wood et al. 2016)	Continuous (Outdoor, synthetic)	–	6480×480	–	Yes	1 M	3D gaze vector; Eye region; shape parameters; lighting information; 2D and 3D; facial landmarks
Driving monitoring dataset (Ortega et al. 2022)	Continuous (outdoor)	29.98	1280×720	–	Yes	Video (41 h)	Face/head/eyes/objects bounding box; gaze location (0–9); hands/body actions; Objects

Datasets are ordered following their publication date, from newest to oldest

distractions and randomly selecting frames. Furthermore, the dataset was augmented by flipping horizontally each image, thereby providing a training dataset both for left and right eyes. Hence, the notation of each image is its corresponding 3D gaze vector within the headset coordinate system. Second, a dataset for eye region, iris and pupil segmentation was provided by means of image masks manually annotated. Similarly, it can be flipped to augment the dataset to train models appropriately for both eyes.

As most of the publicly available datasets are based on pupil and iris detection, Chugh (2020) collected a dataset for the extraction of corneal reflections by combining both their own dataset and an unlabelled dataset from NVIDIA (Kim et al. 2019). Then, corneal reflections were marked manually, thereby generating one binary mask per light source and image. The resulting dataset is finally generated and augmented by cropping images and applying image-based operators, such as gaussian and motion blurring, contrast adjustment or adding synthetic reflections. However, this dataset may be constrained to devices with similar IR light configurations.

Besides real datasets obtained from participants, Kim et al. (2019) augments their dataset with synthetic images (Fig. 10b). To this end, a set of models were used along with virtual light sources and realistic rendering to acquire synthetic eye imagery, thus allowing to obtain high-resolution data. Similarly to their real dataset, 4 IR light sources were simulated.

Instead of applying frequent data gathering operations, such as point tracking for the collection of saccades, fixations and smooth pursuits (Palmero et al. 2021; McMurrough et al. 2012) (Fig. 10a), other studies propose

real-world tasks, ranging from indoor navigation or visual search (Kothari et al. 2020) to car riding (Fuhl and Kasneci 2021). However, they require automatic labelling with high confidence (Kothari et al. 2020; Fuhl et al. 2019) that may be preceded by manual labelling (Kothari et al. 2020; Tonsen et al. 2016). Then, data must be curated to avoid erroneous training samples, either by discarding events with abnormal duration (Kothari et al. 2020) or error entries (Fuhl and Kasneci 2021) derived from subject distraction, blinks or incorrect data. Note that false positives are far more harmful than false negatives for training purposes, as the second only reduces the available dataset. Furthermore, most of the experiments are not performed using VR headsets. Instead, most of them are solely based on eye-tracking glasses (Tonsen et al. 2016; Katrychuk et al. 2019; McMurrough et al. 2012; Kothari et al. 2020; Fuhl and Kasneci 2021), allowing better handling of the hardware and varying lighting conditions. Thus, avoiding continuous lighting may be desirable for applying the proposed datasets to other environment configurations. Nevertheless, Kim et al. (2019) managed to alter lighting conditions within VR headsets, whereas other works present a constant light source (Chugh 2020; Palermo et al. 2021).

Regarding user driving, Ortega et al. (2022) published a large dataset that provides, among other features, the location at which the user is looking (pre-defined, from zero to nine) as well as the bounding boxes of face, head, eyes and other objects that are used as distractors while driving. The dataset also includes frames with users yawning, having microsleeps, texting or drinking.

Other recent publications, such as Lu et al. (2022), have published 3D datasets aimed at providing the 3D pupil

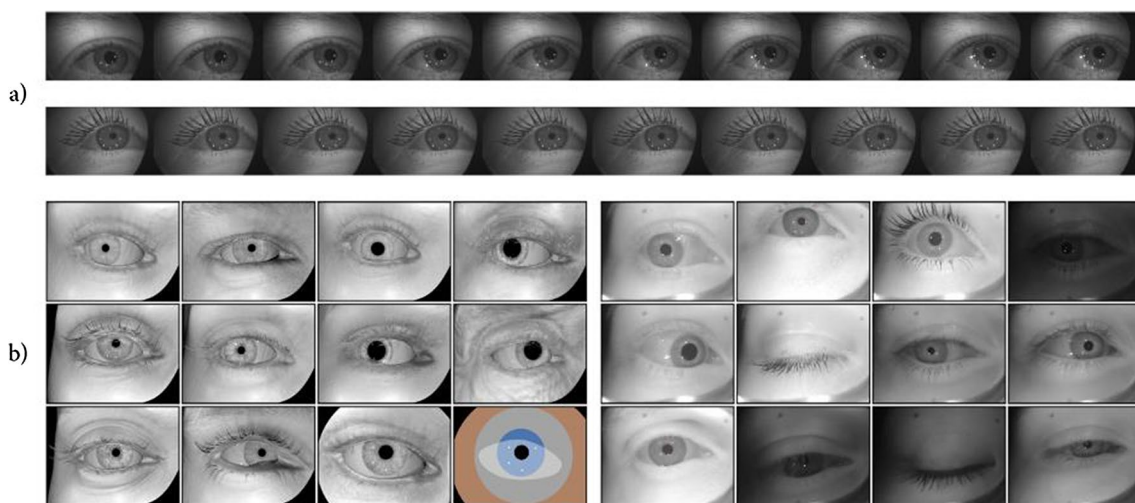


Fig. 10 Eye-tracking imagery from some cited datasets. **a** Comparison of saccades and smooth pursuit from OpenEDS dataset (Palmero et al. 2021), whereas **b** shows both synthetic and real-world datasets,

although only the second one presents varying lighting conditions (Kim et al. 2019)

parameters together with the 2D location and the gaze vector. To this end, the users' head was stabilized during the data collection, and the pupil detection process was significantly improved to surpass previous work. The 3D detection was observed to have a mean error above 2 mm, whereas the gaze vector estimation had an error of 4.38° . Garbin et al. (2020) collected a dataset that split the eye region into eyelid, pupil and iris, and even is able to provide 3D point clouds of the corneal topography. The dataset is composed of 12,759 annotated images collected from 286 subjects.

A summary of the described studies for eye-tracking dataset generation is shown in Table 6. These are the latest and most relevant regarding their size, though other notable work precedes those included in the summary (Fuhl et al. 2015, 2016, 2019). NVGaze (Kim et al. 2019) is solely described according to the real dataset, though a synthetic dataset is also generated. This approach allowed them to generate more advanced region maps. Accordingly, they were able to produce image samples along with the following labels: 2D gaze vector, head position, eye-lid states, pupil size, 2D iris center and pupil center, as well as accurate region maps that separate skin, pupil, iris, sclera, and corneal reflections.

As a future research line, Emery et al. (2021) offer new possibilities since they provide the head and hand orientations as well as the rendered frames, with the 3D gaze vector being the ground truth. Bozkir et al. (2020) introduced a protocol for collecting eye-tracking data in VR remotely. Miller et al. (2021) crafted a post-processing framework combining mobile eye-tracking with motion capture, enabling the computation of a 3D gaze vector tied to object and body positioning in VR, considering metrics such as gaze direction. Finally, Demir and Ciftci (2021) investigated the importance of gaze-tracking to discern fake videos from real ones.

Due to the high number of eye-tracking datasets, Kothari et al. (2022) studied the use of several of them applied to the segmentation of pupil and iris with an encoder-decoder network. As a result, they concluded that the use of several datasets helped to better generalize. Instead, for specific configurations regarding lighting, allowed head movement or scenario configuration, datasets ought to be filtered.

6 Rendering

Rendering optimization is a recurrent topic in Computer Science research, especially for realistic image synthesis. The nature of VR also leads to worsening the performance of the rendering pipeline, as the trivial approach requires rendering the scene twice, one for each eye. Furthermore, the trend in VR is to increase both refresh rate and image resolution to provide a better user experience. Several optimizations have been proposed to deal with VR shortcomings, although most of the reviewed works focus on using devices

that already implement these techniques as their underlying rendering core. Besides improving the rendering, another objective of these optimizations is to avoid discomfort from cyber-sickness. Otherwise, factors such as high latency may lead users to feel they are not there due to discrepancies between visual, vestibular, and pro-prioceptive awareness. As regarded by previous work (Bayramova et al. 2021; Valori et al. 2020), proprioception is the sense of position and movement in space.

Rendering techniques on VR must be interpreted following the development of eye-tracking, as it presents a timeline guided by incremental steps that finally lead to the recent integration of VR and eye-tracking, with the latter being supported via hardware and software. Figure 11 shows the milestones that occurred since the 1800s, finishing with the commercialization of the VR devices that were previously mentioned in Sect. 4. Matthews et al. (2020) jointly reviewed the history of VR and eye-tracking, and thus, this chapter is briefly summarizing some of the identified main events. Accordingly, the development of VR started with the concept of stereoscopy (Brewster 1856) and evolved in waves. In parallel, eye-tracking and gaze-tracking were rudimentarily started in 1910 with a device affixed to the user's eye (Huey 1968). The first non-intrusive manner of recording the eye's movement emerged in 1937, where the beam light reflections were recorded in a piece of film (Hartridge and Thomson 1948). Then, a disruptive work was published in 1968 by addressing optimizations in the eye-tracking procedure (Yarbus 1967). From here, eye-tracking was mainly investigated as an alternative human-computer interaction (HCI) (Bolt 1982), and evolved into non-intrusive solutions such as webcams and IR cameras. The two most recent milestones come with (1) the main VR industry leaders emerging in 2016, including Oculus, HTC, Sony or Valve, and (2) eye-tracking being integrated into VR headsets (Vive Pro and Vive Cosmos along with Tobii) and the first 8 K headsets being announced.

During this development, optimizations on the rendering using eye-tracking were studied and laid the foundations of nowadays current foveated rendering, i.e., concentrating the computational resources at the point at which the user is looking. Following this approach, areas falling out of the focus area can be rendered with a lower level of detail (LOD). These LOD variations were first handled with geometric simplifications (Zheng et al. 2018), followed by adaptive rendering with variable resolution across the image (Meng et al. 2020). Similar to geometry, shading simplifications are also trivial; more accurate and time-consuming techniques are applied over the focused area, whereas more efficient, and less accurate methods are applied over peripheral areas (Xiao et al. 2018). On the other hand, a lower effort has been put into spatiotemporal deterioration. In the latter, the refresh rate is configured across the image, and

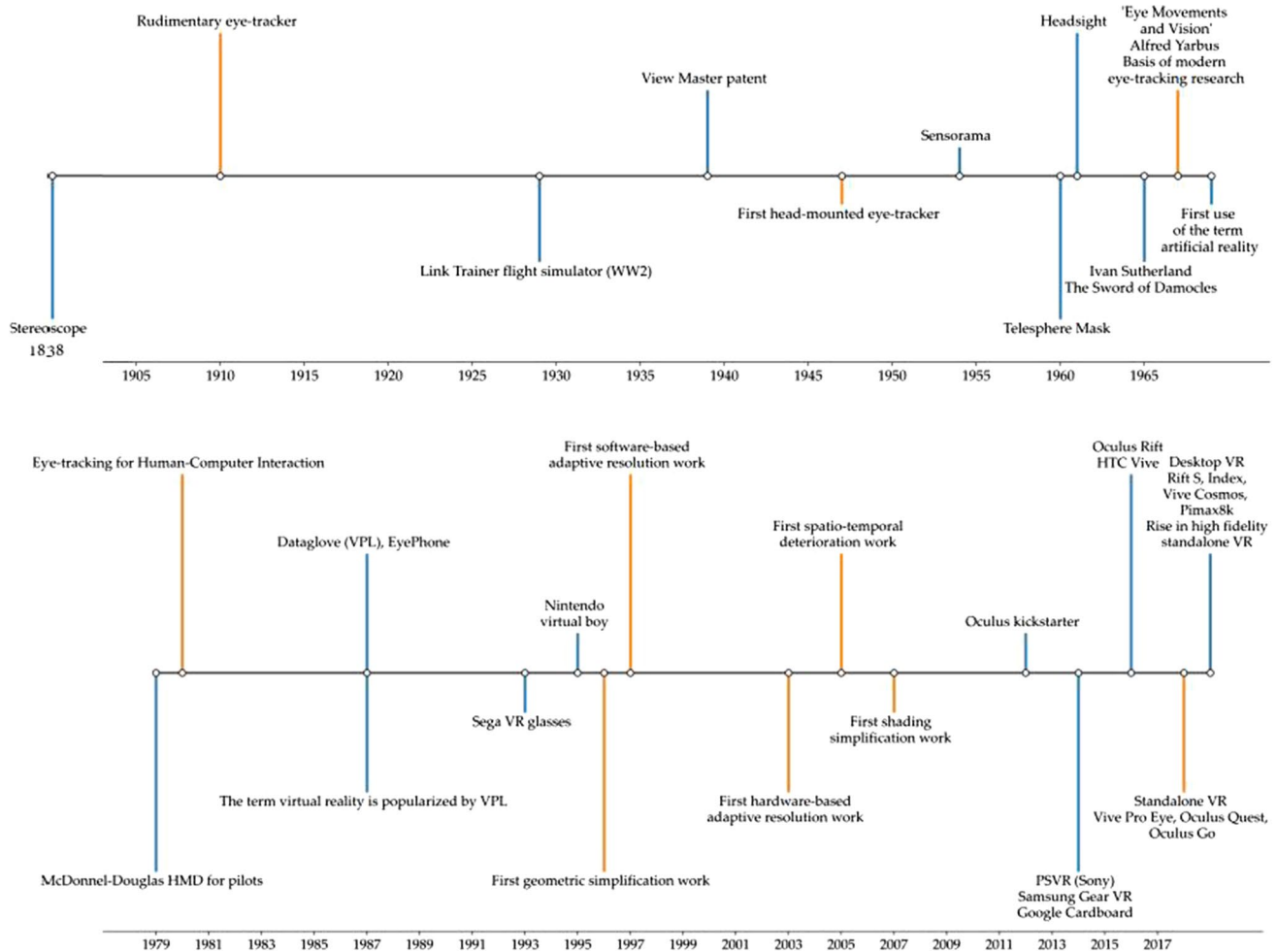


Fig. 11 Timeline from 1838 to nowadays showing the main milestones related to rendering advancements

even data stored in the cache can be used multiple times in non-relevant areas (Franke et al. 2021). With this in mind, VR rendering optimizations are following classified into traditional and perception-based optimizations.

6.1 Traditional optimizations

The naive approach is based on duplicating the draw calls to render the view of both eyes, thereby introducing a large latency overhead for dense scenarios and intricate rendering pipelines, including ray-tracing. This increased latency is even higher for lighting techniques that require additional draw calls per light, such as cascade shadow mapping (White et al. 2021). Instead of applying multi-pass rendering for each view, Multi-View rendering (MVR) has been the preferred solution for addressing this problem. The underlying concept in MVR is to avoid several draw calls, which are known to be time-consuming, and instead, calculate the outcome of several views during a single one, in contrast to Single View Rendering (SVR). This problem has been

addressed for decades, starting from the rendering of splats which were later used as impostors for different viewpoints (Schaufler and Stürzlinger 1996). Storing previous shading calculations in a cache (Sitthi-amorn et al. 2008) and enforcing restrictions in the discrepancies between different viewpoints have also been investigated (Halle 1998). Other optimizations rely on determining the potentially visible set (PVS), mainly in a camera $\leftarrow \rightarrow$ pixel relation, though more recent work has studied the correlation of camera movements and pixels (Hladky et al. 2019) to speed up the rendering.

MVR was first supported by Nvidia hardware with Pascal architecture for generating up to two simultaneous views, as regarded in Fig. 12. The same draw call is able to render two different views of the scene, each one with a different camera orientation, which are stored in two different textures. Lately, it has been improved for GPUs with Turing architecture, enabling the rendering of four different views for ultra-wide FOV. These new features can be accessed through the extension named OVR_multiview (NVIDIA Corporation

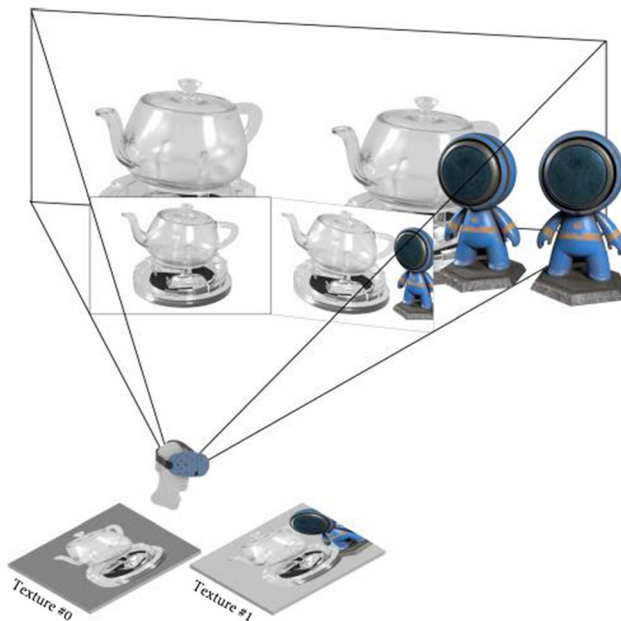


Fig. 12 Multi-View Rendering of two different versions of the scene, with two target textures

2018; Unterguggenberger et al. 2020) in the OpenGL standard (Open Graphics Library), Vulkan, DirectX11 and DirectX12. However, MVR still requires multiple fragment shader calls despite using one pass. With this in mind, Unterguggenberger et al. (2020) explores the latency derived from MVR with a flexible framework that proposes several possibilities regarding geometry instancing, framebuffers as well as culling and clipping. Depending on the number of subdivisions, scene size and GPU architecture, the best pipeline was shown to vary among configurations.

6.2 Perception-based rendering

Rather than optimizing the rendering pipeline, the resulting image can be adapted according to the psychophysical aspects of the human's eye, which can be split into three regions, from lower to higher angular covering: (1) foveal region, (2) inter-foveal and (3) peripheral. The two latter regions present lower visual acuity, though they are also sensitive to motion, and thus could be used to guide the user's attention. This knowledge leads to the widespread foveated rendering (Fig. 13).

In the initial stages of GPU development, only an overall fixed sampling frequency could be utilized, with fragment shaders being executed once per pixel. Still, this remains the standard approach unless indicated otherwise. On the other hand, multi-sampling enable fragment shaders to be run multiple times per pixel; however, the number of shader calls per pixel remains constant over the entire

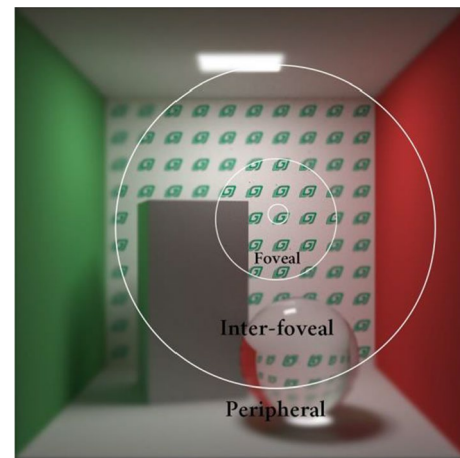


Fig. 13 Adaptive rendering of a scene according to eye-tracking, thereby allowing to optimize rendering in the foveal region

image. The first solutions to overcome this were based on three draw calls, each one with lesser quality (Guenter et al. 2012), as well as geometric techniques based on targeted mesh simplifications (Weier et al. 2014) and LOD (Mohanto et al. 2022). More recently, this limitation has been resolved with Multi-Rate shading, which allows varying the number of calls per pixel and even requiring calls for a group of pixels (NVIDIA Corporation 2020). This technique is currently supported by hardware with the NV_shading_rate_image extension for Turing GPUs. With the Nvidia extension, we can control the number of calls per pixel, SHADING_RATE_N_INVOCATIONS_PER_PIXEL_NV, and narrow it to a few calls for surrounding pixels, SHADING_RATE_1_INVOCATION_PER_IxJ_PIXELS_NV, according to a predefined enumeration.

However, this variation based on the eye's target also leads to aliasing and image disruptions that can be recognized by users. To solve this, the adaptive image has been processed, for instance, by addressing the contrast variation (Patney et al. 2016). Similarly, image contrast has been used to increase or reduce the number of samples depending on the contrast of an image area (Tursun et al. 2019). Even the dominance of one eye over the other can be exploited to reduce the rendering latency of an eye with lower visual acuity (Meng et al. 2020). Also, Generative Adversarial Networks (GANs) have been applied to the generation of plausible peripheral images from the foveal area (Kaplanyan et al. 2019). Tariq et al. (2022) included procedural noise for specific spatial frequencies that otherwise would be detected by users in oversimplified areas. Besides isolated frames, the encoding and compression of foveated-rendering videos have also been studied (Illahi et al. 2020).

7 Applications

We have reviewed different fields of applications for VR and eye-tracking, thus leading to the following classification: medicine and biology, neuroscience and marketing, engineering and architecture and education and training. These categories highlight how some disciplines leverage the benefits of VR and eye-tracking showing different ways of using these technologies. Figure 14 shows the number of papers that have been found for the different applications.

7.1 Medicine and biology

The use of virtual reality in medicine has been increasing in recent years as the technology has improved. One of the improvements that added great value to VR was eye-tracking. This allowed the introduction of improvements in different medical studies and meant a new form of innovation in fields such as rehabilitation, treatment of eye problems, vertigo, anxiety, Alzheimer's disease, among others.

For rehabilitation, Fromm et al. (2019) introduced the potential of home rehabilitation through VR, enhanced by eye-tracking. Park et al. (2019) explored the feasibility of eye-tracking-assisted vestibular rehabilitation. Their experiments, which employed saccadic eye exercises, primarily measured spatial accuracy. They concluded that eye-tracking algorithms enhance vestibular rehabilitation using HMDs. Similarly, Lee et al. (2020) adopted vestibular rehabilitation exercises in VR, including Cawthorne–Cooksey and Herdman training, emphasizing the spatial accuracy measurements. Their findings suggest VR-based methods, coupled

with eye-tracking, could offer safer and more engaging rehabilitation.

In the domain of ocular diseases, Tan et al. (2020) introduced an eye-tracking-aided VR system for amblyopia pediatric care, utilizing metrics like fixations count, dwell-time and transitions order. Their system dynamically adjusted the difficulty based on eye-tracking data, enhancing amblyopia treatment outcomes. Also, knowing the gaze position, eye-tracker can provide assistance to help the patient achieve the goal by providing a tip. Yaramothu et al. (2019) assessed the VERSE video game for vision therapy, capturing clinical measures such as Near Point of Convergence (NPC), Positive Fusional Vergence (PFV), near/far dissociated phoria and Convergence Insufficiency Symptom Survey (CISS). Results highlighted the game's efficacy. Yeh et al. (2021) employed an eye-tracking VR system to measure ocular deviation in strabismus patients, finding a strong correlation with the traditional alternate prism cover test. Martinez-Almeida Nistal et al. (2021) analyzed glaucoma patients' gaze patterns, emphasizing metrics like saccadic velocity, fixations count and fixation/saccadic ratio. Lastly, Mehringer et al. (2021) contrasted Hess Screen Test results in VR using eye-tracking with monitor-based methods, noting variations in measured visual deviation angles.

Beyond these domains, VR integrated with eye-tracking has eased Magnetic Resonance Imaging (MRI) procedures by minimizing patient anxiety, as noted by Qian et al. (2021). Al-Ghamdi et al. (2020) showcased eye-tracked VR as a potential analgesic during painful medical procedures. Davis (2021) evaluated VR and eye-tracking's application for Alzheimer's patients, with fixations being a key metric to monitor the disease's progression. In biology, Gunther et al.

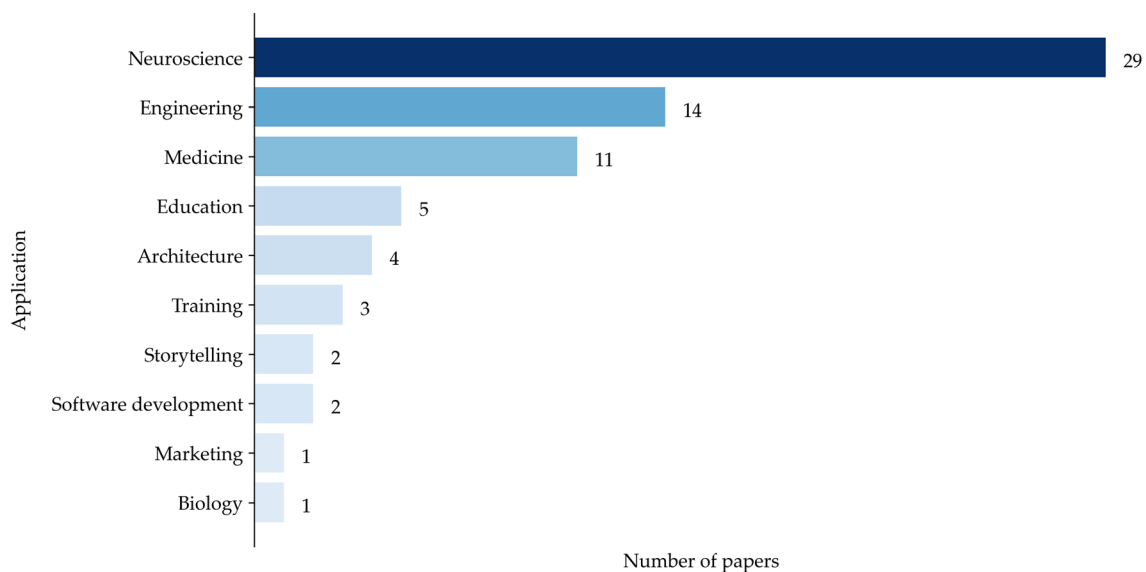
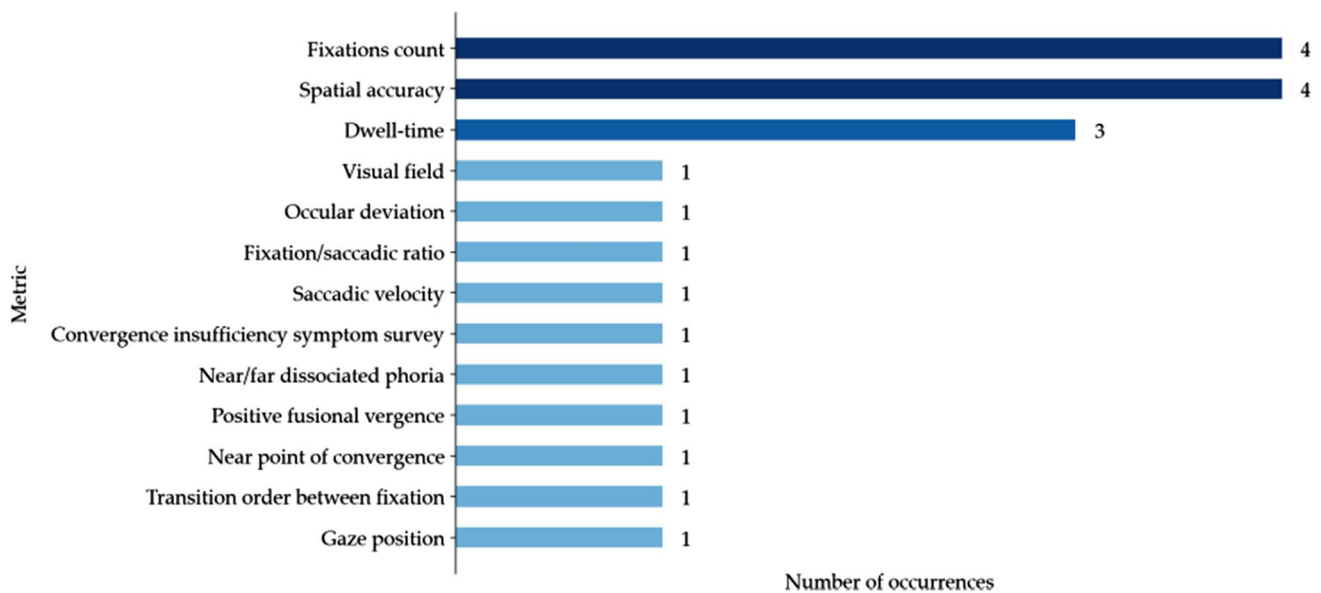


Fig. 14 Number of studies devoted to each one of the considered application fields

Table 7 Brief summary of some of the most important studies carried out in the field of medicine and biology using eye tracking and VR

Applications	Example studies	Brief description
Rehabilitation	Park et al (2019)	Test the feasibility of interactive components using eye tracking assisted strategy through neurophysiologic evidence
Ocular diseases	Tan et al (2020)	Evaluate and apply dynamic difficulty adjustment to amblyopia pediatric care
RMI anxiety avoiding	Qian et al (2021)	VR framework with immersive physical sensations (noise, vibration, etc.) to avoid anxiety and stress during MRI scanning
Analgesic	Al-Ghamdi et al (2020)	Study of VR effectiveness for reducing pain during thermal pain stimuli
Alzheimer	Davis (2021)	Feasibility of using VR and eye tracking with older adults with and without Alzheimer
Biological cell tracking	Gunther et al.(2020)	Bionic tracking highlighting the benefits and limitations

**Fig. 15** Number of occurrences of different metrics in medicine and biology applications

(2020) introduced Bionic Tracking in VR using eye-tracking to trace biological cells, demonstrating its accuracy compared to traditional methods. Table 7 shows an overview of some of the applications mentioned, together with the most cited article in each of them. The most frequently observed metrics in these fields are shown in Figure 15.

7.2 Neurosciences and marketing

Virtual reality, enhanced by eye-tracking, is reshaping our understanding of the human mind in areas such as neuroscience and marketing. This combined approach offers deeper insights into human behavior, brain disorders, and even emotion recognition.

Within the realm of neglect disorders, Hougaard et al. (2021) demonstrated the value of VR and eye-tracking in assessing spatial neglect subtypes in stroke patients. Their experiment utilized metrics like dwell-time, fixations count and eye orientation, revealing significant differences

in eye-tracking data between stroke and healthy patients. Similarly, Ogura et al. (2019) designed a VR application that quantitatively assessed the visual field in patients with unilateral spatial neglect using color changes in observed blocks. Additionally, Porrás-García et al. (2019) explored attentional bias toward body parts in VR and eye-tracking, emphasizing the differences in fixations between genders on weight-related and non-weight-related areas.

On human behavior, Reichenberger et al. (2020) analyzed how social anxiety influences attention in VR, especially concerning emotionally threatening stimuli, using dwell-time and fixations count. Wang et al. (2019)'s study on VR advertising highlighted that commercial objects are not the primary focal points, based on measures such as fixations count, duration of first fixation and total fixation duration. Pettersson (2021) conducted a thesis harnessing VR-based eye-tracking metrics, namely gaze direction, pupil position, and pupil diameter. These data were subsequently analyzed using a neural network to decipher user behavior.

Melendrez-Ruiz et al. (2021) concluded that pulses were not effective eye-catchers in a VR supermarket setting after considering metrics like dwell-time and fixations count. In contrast, Tian et al. (2019) show-cased the potential of VR and eye-tracking in fire escape behavior analysis, emphasizing the efficiency of the approach.

For prediction purposes, Stein (2021) is pioneering a method to predict locomotion paths in VR, linking various behavioral data including eye-tracking. He examined metrics like eye-tracking latency and distance between users' gaze and a target for predicting future paths. Wechsler et al. (2019) indicated that gaze behavior, when analyzed using fixations count and dwell-time, could predict physiological stress responses. Furthermore, Huizeling et al. (2021) asserted that hesitation words during speech impact prediction capability based on eye-tracker fixations count.

Concerning emotion and recognition, Liu et al. (2020) discovered that stereoscopic images enhance facial recognition, as evidenced by measures such as fixations, dwell-time and pupil diameter. Geraets et al. (2021) compared facial emotion recognition across different media using VR and eye-tracking metrics like fixations count, dwell-time and total fixation duration and highlighted the potential of VR for emotion recognition training. Similarly, Tabbaa et al. (2021) compiled a dataset integrating eye-tracking data (gaze position, eyes status) and physiological measurements for VR emotion recognition, and Bozkir et al. (2019) presented an approach for recognizing driver cognitive load in VR and eye-tracking by collecting pupillary information, gaze position and performance measures (inputs on the accelerator, brake, and steering wheel) from a VR driving experiment to train multiple classifiers. Lim et al. (2021) undertook an initial study utilizing pupil position in VR eye-tracking to discern emotions. Their findings suggest the promising potential of pupil position as an emotion recognition metric. In another significant study,

Hickson et al. (2019) introduced an algorithm that utilizes eye-tracking to infer facial expressions, even with partial face occlusion. Their trials with various convolutional neural networks yielded an impressive mean accuracy of 73%, surpassing the proficiency of advanced human raters.

Additional studies like Kobylinski and Pochwatko (2020) focused on movement detection in VR narration. Sterna et al. (2021) proposed an ideal design for psychophysiological and eye-tracking measurement in VR, emphasizing the fixations count. Mirault et al. (2020) analyzed transposed-word effects on reading using eye-tracking metrics in VR such as fixations count, dwell-time and gaze position. Maraj et al. (2021) investigated immersion and comfort using eye-tracked VR devices, concluding no significant difference in user responses. Jurik et al. (2019) explored the use of eye-tracking in VR, advocating for its potential in cross-cultural studies on human perception and cognition. Ryabinin et al. (2021) examined hierarchically segmented images, such as historical paintings, employing metrics like gaze position, fixations count, and saccades count. In the broader context, Meißner et al. (2019) discussed the role of VR and eye-tracking in marketing research, stressing the importance of metrics like spatial accuracy, gaze position, fixations count and saccades count, while Soret et al. (2020) explored how auditory and visual stimuli impact attention in VR by evaluating saccadic reaction time. Marwecki et al. (2019) introduced "Mise-Unseen" software in VR, leveraging eye-tracking to determine optimal scene change moments based on user attention, spatial memory, and metrics such as gaze position, spatial accuracy and pupil diameter. In the realm of VR storytelling, Yang et al. (2021) conducted a study to discern if eye-tracking could serve as implicit interactions, exploring the boundaries between implicit and explicit interactions in VR eye-tracking. Lastly, Wang et al. (2020) evaluated which stimuli sources are the most effective to

Table 8 Summary of some important studies in neuroscience and marketing using eye tracking and VR

Applications	Example studies	Brief description
Neglects and disorders	Porras-Garcia et al (2019)	Assess gender differences in attentional bias toward specific weight or non-weight-related body parts
Human behavior	Reichenberger et al (2020)	Socially anxious fear ratings, physical behavior, hypervigilance, and attentional avoidance during social fear conditioning and extinction
Predictions	Pettersson and Falkman (2020, 2021)	Classify and predict human movement direction
Recognitions	Hickson et al (2019)	Infer facial expressions using the user's eyes captured from an IR camera within a VR headset
Marketing	Meißner et al (2019)	Review key advantages of different eye-tracking technologies and show its potential combined with VR on shopper assistance
Cinematic virtual Reality (CVR)	Wang et al (2020)	Experiment with different attention-guiding methods in watching VR movies and offer new basis and references for narrative mode in VR movies
Cognitive psychology	Jurik et al.(2019)	Propose the use of eye-tracking and VR for cross-cultural research of human cognitive processes and perception

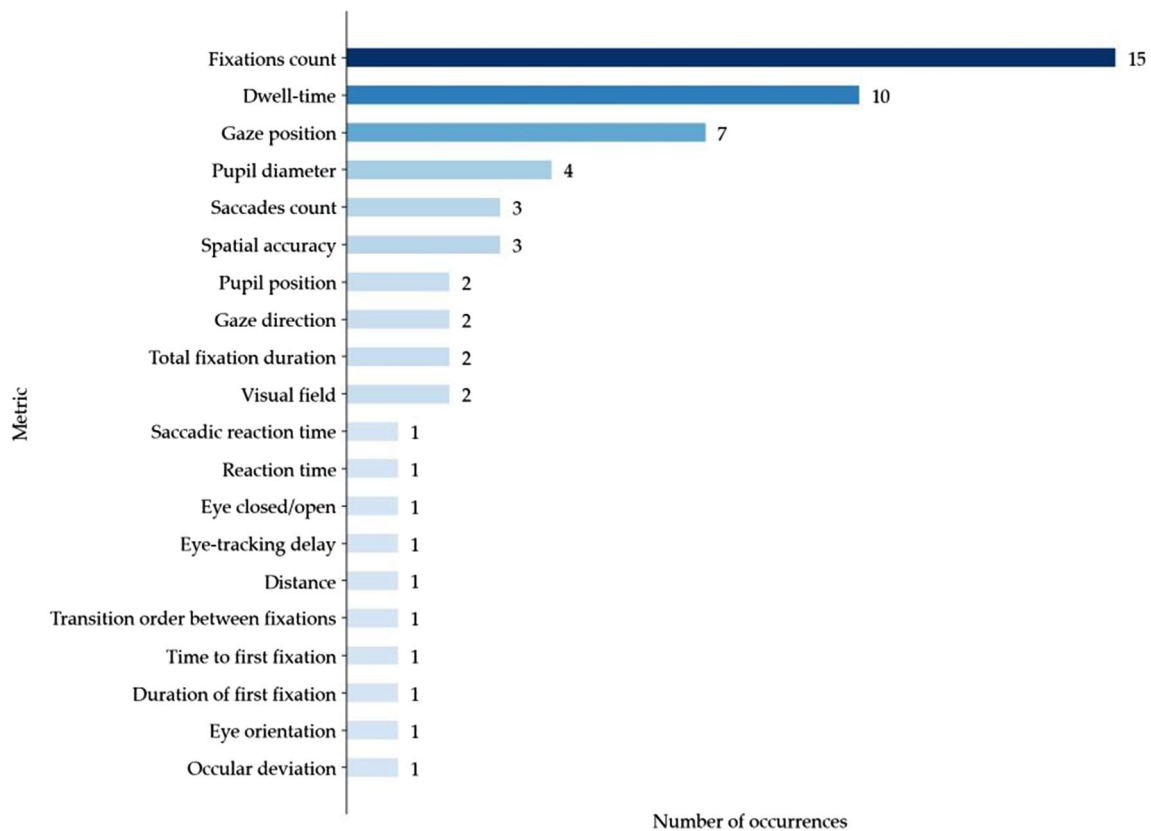


Fig. 16 Number of occurrences of different metrics in neuroscience and marketing applications

guide users in Cinematic Virtual Reality (CVR) by using eye-tracking data such as gaze position and reaction time.

Table 8 shows an overview of some applications mentioned selected on the criteria of relevance (citations), whereas the popularity of different metrics in this field is depicted in Fig. 16.

7.3 Engineering and architecture

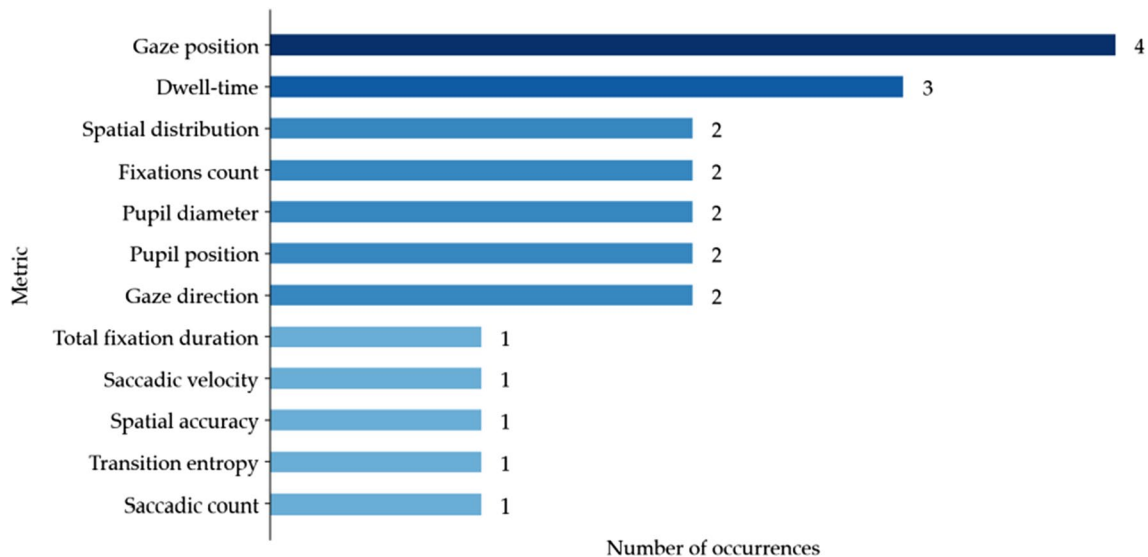
In the fields of engineering and architecture, eye-tracking within VR has been applied to diverse areas from risk assessment and situational awareness to architectural design and algorithm enhancement. Kang et al. (2020) delved into the correlation between visual paths and situational awareness, utilizing eye-tracking metrics like fixations and saccades in an oil rig anomaly detection scenario. Their findings revealed distinct visual paths between participants with varying levels of situational awareness. Khatri et al. (2020) focused on refining age classification during a virtual shopping task by optimizing the Dispersion Threshold Identification algorithm using eye-tracking spatial distribution. Cubero (2020) predicted user choices using time-series data and LSTM, a type of RNN, analyzing gaze position. Dong et al. (2020) introduced the “Central-Eye” eye-tracking

method to discern human gaze focus, utilizing metrics like spatial accuracy and saccadic velocity. Additionally, Pettersson and Falkman (2020) classified human movement direction in a virtual environment using metrics like gaze direction, pupil position and pupil diameter to enhance collaborative robot intelligence, with a follow-up study by Pettersson and Falkman (2021) on predicting human movements using neural networks. In architectural design, Bar-san-Pipu (2020) blended brain-computer interface (BCI), eye-tracking, VR, and AI-driven neurofeedback to discern designers’ conceptual design intentions, providing dynamic responses. Zhang et al. (2019) proposed a comprehensive method for cityscape design and protection, merging cognitive psychology, spatial behavior, and sociology, while emphasizing metrics like gaze position. Lastly, Özel (2019) developed a hazard recognition system for construction sites using VR and eye-tracking, investigating the impact of work experience and education on hazard recognition using metrics such as dwell-time and total fixation duration and total fixation duration.

Table 9 shows an overview of some of the applications mentioned. The selection has been made based on the citations of the articles, choosing the most cited article of each application. On the other hand, Fig. 17 shows the number

Table 9 Some relevant studies in the area of engineering and architecture using eye tracking and VR

Applications	Example studies	Brief description
Situation awareness	Kang et al (2020)	Investigate how situation awareness can be associated with visual scan paths in an anomaly detection task
Hazard recognition	Özel (2019)	Provide a methodology for a practical, ethical and feasible way of conducting safety research
Design and protection of cityscape	Zhang et al (2019)	Identify cityscape features and propose a multi-cultural strategy for protecting cityscape
Optimization	Khatri et al (2020)	Proposed a method to optimize dispersion thresholds in order to improve the accuracy of age classification based on different visual patterns achieved during a virtual shopping task

**Fig. 17** Number of occurrences of different metrics in engineering and architecture applications

of occurrences of previously revised metric in engineering and architecture fields.

7.4 Education and training

Wang et al. (2021) employed VR and eye-tracking to explore how students, varying in prior knowledge, process visual behavior related to Japanese mimicry and onomatopoeia in learning Japanese as a second language. They *ye-tracking on virtual reality*: utilized eye-tracking metrics like fixation count and dwell-time, expanding on the application of visual behavior in real-time VR environments. Meanwhile, Khokhar et al. (2019) introduced an architectural framework for educational VR, aiming to make VR pedagogical agents responsive to user attention changes monitored via eye-tracking, leveraging metrics such as pupil diameter, gaze direction or distance between user's gaze and a target. Additionally, Bacca-Acosta and Tejada (2021) delved into efficient eye-tracking data collection in 3D virtual environments and further investigated students' visual behavior during English preposition learning

in Bacca-Acosta et al. (2021) by analyzing fixations count and dwell-time from eye-tracking. Their findings indicate the efficacy of dynamic IVR environments with integrated scaffolds for enhanced learning performance, though they also highlighted challenges with the preposition "on". In addition, Gadin (2021) examined legibility in VR text displays through eye-tracking metrics, including fixation count, dwell-time and amplitude of saccades. Lastly, Komoriya et al. (2021) devised a system enabling hand-capped individuals to compose text on computers through eye blinking and shifting. Post-feasibility analysis, notable enhancements in desktop maneuvers, success rates, and reduced execution times were observed.

In the field of training, Laivuori (2021)'s thesis employed VR eye-tracking in a simulation training sea captains, using metrics like gaze position and pupil diameter. Burova et al. (2020) developed a training application emphasizing safety awareness, considering metrics like fixation count and gaze position. For sports, Mutasim et al. (2020) harnessed eye-tracking in VR to boost user performance, analyzing aspects like reaction time or search time.

Table 10 A selection of studies with eye tracking and VR applied in the field of education

Applications	Example studies	Brief description
VR pedagogical agent	Khokhar et al (2019)	changes in user attention
System from handicapped users	Komoriya et al (2021)	Implementation of a flick function in the user interface to enable handicapped people to operate computers only with eyes
Training	Burova et al (2020)	Development of a VR/AR safety kit framework and evaluating AR features (in-field guidance and safety warning) in a realistic industrial process with experts

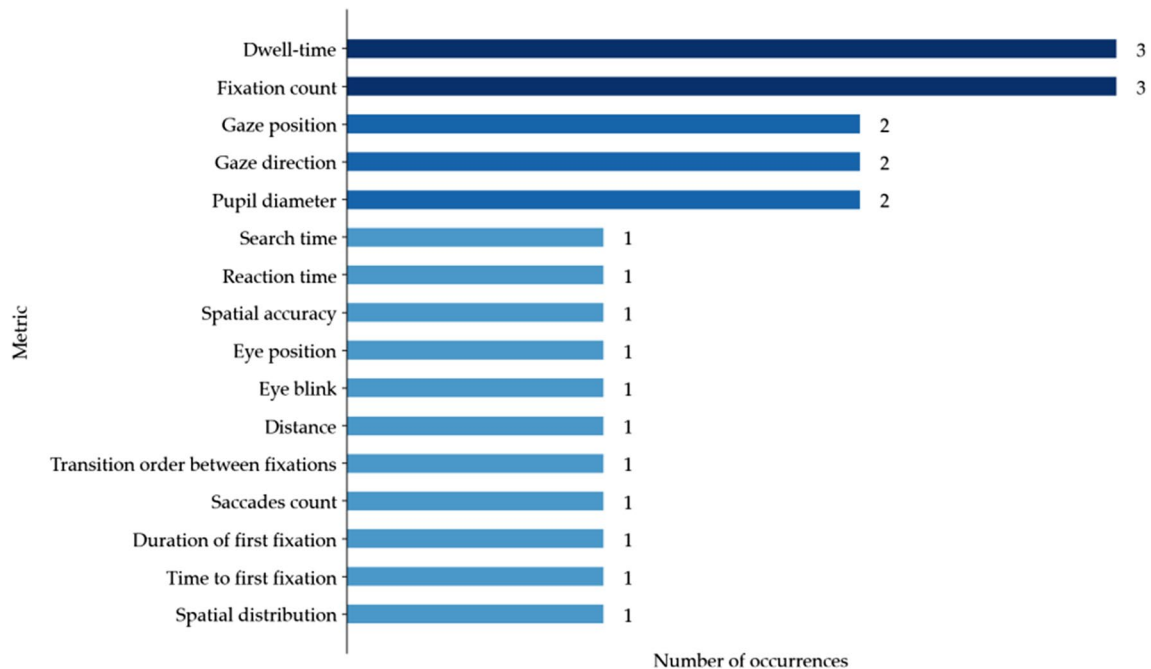
**Fig. 18** Number of occurrences of different metrics in education applications

Table 10 shows an overview of some of the applications mentioned taking into account the most cited studies. As in previous applications, Fig. 18 depicts the number of occurrences of different metrics.

8 Discussion

The latest major addition to virtual reality has been eye-tracking technology. Traditionally, studies using the user's gaze direction were conducted on a computer screen where external eye trackers could be attached. However, with the addition of this technology to VR headsets, a new range of possibilities opens up to design more immersive applications where the user's gaze can play an important role, not only for studies where the gaze provides relevant information but also for the inclusion of new interaction and optimization mechanisms. This integration not only allows us to explore more fields of applications for VR but also to improve the

use of VR in existing fields: for example, it is possible to improve the rendering performance of VR applications with eye-tracking by concentrating computational resources at the gaze location. Among the opportunities offered by this technology, we can highlight better monitoring of user attention with the help of precise estimation of gaze vectors. Also, the registration of pupil dilations related to stimuli with a relevant meaning for the user helps to determine emotions without the need for external hardware such as heart rate or respiratory monitors (Finke et al. 2021).

However, there are still open problems as well as current and future works that must be pointed out. As reviewed in this survey, eye-tracking integration cannot yet be found in many virtual reality headsets (see Table 3) and these are often expensive. In addition, the usefulness of these integrated eye-tracking systems depends on their configuration and accuracy, and over this, the software layer is another key factor to get the most out of this technology. A nuanced understanding requires an examination of the technical

specifications of various devices. HMDs with integrated eye-tracking typically report an accuracy ranging between 0.5° and 1° . However, in certain real-world evaluations with devices like the FOVE0 and HTC VIVE Pro Eye, this accuracy has been observed to deviate, reaching up to 2° (Chernyak et al. 2021; Lamb et al. 2022), contingent on the experimental conditions. Comparatively, desktop eye-tracking devices, exemplified by the Tobii Pro Spectrum, boast a superior accuracy threshold of 0.3° (Tobii 2022a, b) under optimal conditions. Yet, it is worth noting that even this touted accuracy has been challenged in practical applications, with some studies suggesting a deviation nearing 1° (De Kloe et al. 2022). Additionally, these may face accuracy issues due to the prism effect. This phenomenon arises especially when the center of the eyeglass lens does not align seamlessly with the center of the pupil, introducing potential error margins (Yeh et al. 2021). This juxtaposition between HMD-based and desktop eye-tracking systems elucidates a tangible challenge: the pressing need for the further refinement of HMD systems to approach, if not surpass, the accuracy benchmarks set by their desktop counterparts.

Not all scenarios mandate the highest degrees of accuracy. For instance, in entertainment VR applications or basic interactive platforms, where the primary objective might be gauging user interest or attention, a moderate level of accuracy could be sufficient. Conversely, more specialized domains have stringent accuracy requirements. In medical training simulations or precision skill training modules, where the distinction of a few millimetres in gaze direction can make a significant difference, the demand for impeccable accuracy becomes paramount. Similarly, in research contexts where detailed gaze patterns are analyzed to derive cognitive or behavioral insights, the granularity and accuracy of eye-tracking are crucial. The aforementioned variations underscore the importance of tailoring eye-tracking systems and their accuracy thresholds to the specific objectives and requirements of individual VR applications.

A significant hindrance is the susceptibility of VR users to motion sickness, which can elevate drop-out rates in experiments, potentially skewing results (Clay et al. 2019). There is also the concern of position shifts in the VR headset after calibration, perhaps resulting from abrupt head movements or improper adjustments. Such shifts can considerably undermine accuracy. Addressing the aforementioned challenges necessitates a multi-pronged approach to enhance the robustness and reliability of eye-tracking within VR environments. One primary avenue for future exploration is the mitigation of motion sickness in participants. Advanced algorithms that predict and adjust for potential motion sickness triggers, based on individual user profiles, could be developed. Additionally, it might be beneficial to explore real-time re-calibration mechanisms (Plopski et al. 2016). These calibrations would serve to consistently maintain the

accuracy of eye-tracking without disrupting the user experience. Furthermore, enhancing the ergonomic design of VR headsets to minimize position shifts will be pivotal. Collaborations with biomechanical engineers could lead to the development of headsets that balance comfort with secure placements, reducing the likelihood of inadvertent shifts during experiments.

Similarly, the requisite minimum frequency for eye-tracking, contingent on the type of eye movements being observed, inherently places hardware constraints. This stipulation might preclude the use of certain devices or necessitate specific configurations (Geraets et al. 2021). This limitation is not merely a question of selecting an appropriate device but embodies broader challenges in ensuring universal applicability and reproducibility of findings. One pivotal avenue for future investigation will be the development of adaptive tracking systems. These systems would intelligently modulate their tracking frequency based on the specific requirements of the ongoing task or experiment. Such adaptability could potentially obviate the need for high-frequency tracking during phases where it is unnecessary, conserving computational resources and power. Simultaneously, advancements in hardware miniaturization and processing capabilities are imperative. Collaborative efforts between eye-tracking research and hardware engineers might yield devices that, despite being compact, do not compromise on tracking frequency or accuracy. Innovations in semiconductor technologies and algorithmic optimizations can play a significant role in this direction. Additionally, exploring cloud-based processing solutions, where the eye-tracking data is processed remotely rather than on the device itself, might alleviate some hardware constraints (Zou et al. 2021). From a design perspective, the incorporation of an eye-tracking mechanism can alter the ergonomic properties of VR headsets. The additional weight or altered balance might impede prolonged usage, posing a challenge, particularly for extended experimental sessions or immersive experiences. Research into materials and design paradigms that allow for lightweight yet robust eye-tracking integration will be paramount. This would ensure that users can engage in prolonged VR sessions without discomfort, ensuring the integrity of extended experiments or applications.

Alternate to nowadays HMDs with integrated eye-tracking, a relevant number of studies have investigated custom eye-tracking devices using existing HMDs as the underlying structure. However, the optimal number of IR light sources that must be used per eye is not clear, varying from one to eight in the literature, although using a larger number is known to have negative effects on users (Qian et al. 2021). In addition, calibration must be performed from scratch by showing patterns and locations on the screen and calculating the parameters of a model, whose complexity is frequently reduced to second-order polynomial expressions. Custom

devices are simply the baseline over which alternative eye-tracking methods are checked. These methods range from feature-based, intended to find features such as the iris, to model-based, focused on the detection of specific shapes, such as a circle on the iris and pupil locations, as well as ML and DL models. The main shortcomings of all these methods are that they are highly dependent on the tested recording conditions, and therefore, reflections and accessories such as glasses harden the recognition of eye parts. Note that, feature-based and model-based frequently use image processing techniques, including thresholding and morphological operators, to process the recorded images. Hence, minor deviations from the expected intensity distribution have notable effects on the outcomes. Despite this, revised articles reported accuracies even below 1° ; however, these are typically observed against custom datasets and should be further checked to assure that these solutions are more effective than current commercial solutions.

On the other hand, machine learning techniques are starting to gain interest as they quickly determine the gaze point with low latency. CNNs are the most studied networks by far, either to perform semantic segmentation and output labelled eye parts or to calculate the normalized gaze location/vector. The majority of works use shallow CNN, which is expected to have a low memory footprint, and pre-trained networks, which are already known to perform well in classification tasks, despite having a larger memory footprint. On the downside, these kinds of methods require large datasets for learning relevant features from collected eye-tracking data, either from images or numerical data (e.g., head orientation). Rather than showing a similar data scheme, widespread datasets have different features or similar ones captured in different lighting conditions, thus hardening the simultaneous use of various datasets. For instance, two of the previously revised driving datasets do not detect the gaze vector, but a gaze zone; Ortega et al. (2022) labelled ten zones, whereas Naqvi et al. (2018) distinguished seventeen. Still, Kothari et al. (2022) proved that feeding networks with multiple datasets contributed to obtaining better results. Another yet barely explored area is the synthetic generation of datasets, which together with realistic shading, may help to construct huge datasets including users from every possible demographic group, without the time-consuming tasks of gathering participants, labelling and cleaning data. Furthermore, few works have exploited numerical data together with images in machine learning (Wong et al. 2019). In this regard, the most recent datasets are providing further data, such as head and hand pose (Emery et al. 2021). Thus, the main limitations concerning machine learning on eye-tracking are the shortfall of datasets and the disparate range of features published along with them.

Further complicating the scenario is the potential data overload. The granular capture of eye movements within

an intricate VR environment can generate vast datasets. The resultant data not only poses storage challenges but also demands robust algorithms for efficient processing and meaningful analysis. Advanced data compression algorithms, real-time analysis, and cloud-based solutions can help manage the vast amount of data generated. Machine learning models can further aid in filtering out noise and focusing on significant patterns within the data.

Although machine learning techniques are expected to be more robust by learning from huge datasets, these have not yet extensively surpassed feature and model-based methods, or at least, should be checked against similar data. Besides this, the use of one over others may be suffixed to the system requirements. While most of the current research is devoted to desktop applications, there exist other use cases which may not fit in this field. For instance, (Qian et al. 2021) integrated a custom eye-tracking system in an MRI scanner, whereas Katrychuk et al. (2019) checked DL networks with different capacities to be integrated into a Raspberry Pi 3. In conclusion, there is no better technique regarding efficiency for every possible scenario; computer vision pipelines are typically fast enough due to the low dimensionality of IR images, especially in feature-based methods. Machine learning offers, on the other hand, low latency while having higher memory requirements, although this depends on the number of trained weights. Otherwise, it is possible to develop shallow CNN with few parameters that are still able to recognize high-level features.

Another key factor in eye-tracking is rendering. Optimizations concerning VR have achieved a notable level of maturity with MVR, and the horizon points toward supporting a higher number of simultaneous renderings as GPUs increase their capacity. However, optimizations focused on eye-tracking must be still polished. These techniques are referred to as foveated rendering, and despite having their origin decades ago, various bottlenecks remain. Among these, spatial artefacts that make VR experiences not seamless, including aliasing, flickering (motion aliasing) and temporal artefacts, are the most frequent. These artefacts arise due to the rendering differences between peripheral and non-peripheral areas, regardless of the degradation technique employed. Furthermore, the peripheral areas are especially sensitive to contrast changes, even more than stereoscopic depth. Studies combating this problem, for instance by blurring the limits between distinct areas, have not yet completely solved it since they also modify the image contrast (Mohanto et al. 2022). As occurred in eye-tracking, rendering can also be helped by machine learning, although it is still in an early stage which makes these solutions not robust enough for their commercialization. Nevertheless, previous work has already achieved the completion of images using GAN networks (Kaplanyan et al. 2019). Matthews et al. (2020) have

also suggested that multi-rate shading may be implemented with machine learning.

While the aforementioned challenges provide a candid understanding of the current landscape, they simultaneously shed light on the immense potential awaiting realization. As technology and research methodologies evolve, the integration of eye-tracking in VR is poised to open up a plethora of transformative applications that transcend existing paradigms. Delving into the imminent horizon, we discern several promising trends and applications that harness the combined prowess of eye-tracking and VR, sculpting the future trajectory of immersive experiences.

The ever-evolving landscape of VR, augmented by eye-tracking capabilities, is poised to reshape several domains of human–computer interaction. A relatively untapped avenue is the operability of these systems with multiple users on non-single-view displays. This future line is especially relevant as displays tend to grow in size, together with light field displays that enable watching a scenario from different perspectives (Spjut et al. 2020). Hence, narrowing down the number of perspectives to be rendered and discarding those not directed toward any viewer may help in reducing computations. Building upon this, the adaptation of user interfaces utilizing gaze patterns offers tantalizing prospects. Predictive algorithms can leverage eye-tracking data to present a real-time, user-centric interface, thereby enhancing usability and reducing cognitive load (Plopski et al. 2022). This principle is extended further in interactive gaming. As gaming continues to be at the forefront of VR innovations, integrating eye-tracking could redefine gameplay mechanics, making them more immersive and challenging. (Heilemann et al. 2022; Gemicioglu et al. 2023). Simultaneously, as the virtual domain becomes more intricate, the nuances of human behavior become even more crucial. In this context, the integration of realistic ocular movements within virtual entities enhances the verisimilitude of social interaction in VR. The subtleties of eye movements, intrinsic to authentic human communication, are paramount. Through the accurate replication of these nuances, virtual entities can attain a higher degree of anthropomorphic realism, thus facilitating genuine human-avatar interactions (Visconti et al. 2023).

Moreover, in the academic world and professional world, this confluence of VR and eye-tracking is proving invaluable. For the research community, particularly those in cognitive sciences, the amalgamation of VR and eye-tracking presents a robust tool for experimental paradigms. By monitoring ocular movements within controlled VR scenarios, it becomes feasible to derive insights into intricate cognitive processes and behavioral processes (McNamara and Mehta 2020). Our further studies will be focused on this line of research. Furthermore, eye-tracking technology holds significant promise for professional skill augmentation, particularly in fields where precision and focus are paramount.

Real-time feedback mechanisms, derived from eye-tracking in VR modules, can be employed in specialized training scenarios, such as surgical simulations or athletic drills, thereby facilitating accelerated and refined skill acquisition (Cowan et al. 2021; Stoeve et al. 2022; Galuret et al. 2023; Pastel et al. 2023). The data-rich domain that this confluence promises can also be a game-changer for machine learning. The voluminous data from eye-tracking can refine predictive models, aiding in understanding nuanced user behaviors or even identifying potential health concerns. This has therapeutic implications as well. By analyzing ocular responses to specific virtual stimuli, therapeutic strategies can be optimized for conditions such as Post-traumatic stress disorder (PTSD) or specific phobias (Diemer et al. 2023; Fehlmann et al. 2023). In addition, the vast data generated from eye-tracking in VR stands to redefine content recommendation algorithms. Through gaze-based data, platforms could offer hyper-personalized content suggestions, further enhancing user experience (Pfeiffer et al. 2020). This gaze-data also has significant implications in marketing. Analogous to the analysis of click-through rates on contemporary digital platforms, the scrutiny of gaze durations on specific virtual entities or promotional content can be leveraged to fine-tune marketing paradigms, tailoring them to individual user inclinations (Burke and Leykin 2014).

Lastly, the convergence of VR and eye-tracking presents novel opportunities in the realm of digital security. Unique biometric signatures derived from eye movements and retinal patterns could serve as robust authentication mechanisms within virtual environments (Lohr and Komogortsev 2022).

In essence, while challenges persist, the future teems with opportunities, promising a synergy between virtual reality and eye-tracking that could reshape numerous facets of our digital interactions.

9 Conclusions

Although other surveys have discussed some features of eye-tracking systems in different areas such as performance, usability, or trends, this study performs a comprehensive analysis of eye-tracking technology embedded in HMDs in terms of use, integration, and implementation. Besides commercial devices, the infrastructure and implementation of custom and inexpensive devices was also revised to tackle the shortage of the former, as well as to provide a more customized and tailored solution for specific needs and requirements. In addition, this technology has been widely adopted by a large number of applications in recent years, including research reviewed in this survey in fields such as psychology, marketing, and human-computer interaction, as well as practical applications in areas such as assistive technology and user experience design.

Acknowledgements This result has been partially supported by the Spanish Ministry of Science, Innovation and Universities via grants to the first (PGC2018097769-B-C22) and second author (FPU/00100).

Funding Funding for open access publishing: Universidad de Jaén/CBUA.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The file is attached to the submission.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Ghamdi NA, Meyer WJ, Atzori B et al (2020) Virtual reality analgesia with interactive eye tracking during brief thermal pain stimuli: a randomized controlled trial (crossover design). *Front Hum Neurosci* 13:467. <https://doi.org/10.3389/fnhum.2019.00467>
- Altobelli F (2019) ElectroOculoGraphy (EOG) Eye-tracking for virtual reality. PhD thesis, Delft University of Technology, URL <https://repository.tudelft.nl/islandora/object/uuid%3A59c0e444-72c8-42e3-b29d-b0db150b1450>
- Bacca-Acosta J, Tejada J (2021) Eye tracking in virtual reality for educational technology research. *Frameless* 3(1):16
- Bacca-Acosta J, Tejada J, Fabregat R et al (2021) Scaffolding in immersive virtual reality environments for learning English: an eye tracking study. *Educ Tech Res Dev*. <https://doi.org/10.1007/s11423-021-10068-7>
- Barsan-Pipu C (2020) Artificial intelligence applied to brain-computer interfacing with eye-tracking for computer-aided conceptual architectural design in virtual reality using neurofeedback. In: Yuan PF, Xie YMM, Yao J, et al (eds) *Proceedings of the 2019 DigitalFUTURES*. Springer, Singapore, pp 124–135. https://doi.org/10.1007/978-981-13-8153-9_11
- Bayramova R, Valori I, McKenna-Plumley PE et al (2021) The role of vision and proprioception in self-motion encoding: An immersive virtual reality study. *Atten Percept Psychophys* 83(7):2865–2878. <https://doi.org/10.3758/s13414-021-02344-8>
- Biondi FN, Saberi B, Graf F et al (2023) Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Appl Ergon* 106(103):867. <https://doi.org/10.1016/j.apergo.2022.103867>
- Blaschek T, Kurzhals K, Raschke M et al (2017) Visualization of eye tracking data: a taxonomy and survey. *Comput Graph Forum* 36(8):260–284. <https://doi.org/10.1111/cgf.13079>
- Blattgerste J, Renner P, Pfeiffer T (2018) Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In: *Proceedings of the workshop on communication by gaze interaction*. Association for computing machinery, New York, NY, USA, COGAIN'18, pp 1–9. <https://doi.org/10.1145/3206343.3206349>
- Bolt RA (1982) Eyes at the interface. In: *Proceedings of the 1982 conference on human factors in computing systems*. Association for computing machinery, New York, NY, USA, CHI '82, pp 360–362. <https://doi.org/10.1145/800049.801811>
- Borges M, Symington A, Coltin B, et al (2018) HTC vive: analysis and accuracy improvement. In: *2018 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp 2610–2615. <https://doi.org/10.1109/IROS2018.8593707>
- Bozkir E, Geisler D, Kasneci E (2019) 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)*, pp 1834–1837. <https://doi.org/10.1109/VR.2019.8797758>
- Bozkir E, Eivazi S, Akgu'n M, et al (2020) 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)*, pp 397–401. <https://doi.org/10.1109/AIVR50618.2020.00083>
- Brewster D (1856) *The stereoscope; its history, theory, and construction: with its application to the fine and useful arts and to education*. John Murray, google-Books-ID: V185AAAAcAAJ
- Bublea A, Căleanu CD (2020) Deep learning based eye gaze tracking for automotive applications: an auto-keras approach. In: *2020 international symposium on electronics and telecommunications (ISETC)*, pp 1–4. <https://doi.org/10.1109/ISETC50328.2020.9301091>
- Burke RR, Leykin A (2014) Identifying the drivers of shopper attention, engagement, and purchase. In: *Shopper marketing and the role of in-store marketing, review of marketing research*, vol 11. Emerald Group Publishing Limited, pp 147–187. <https://doi.org/10.1108/S1548-643520140000011006>
- Burova A, Mäkelä J, Hakulinen J, et al (2020) 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*. Association for computing machinery, New York, NY, USA, p 1–13, URL <https://doi.org/10.1145/3313831.3376405>
- Chaudhary AK, Kothari R, Acharya M, et al (2019) RITnet: real-time semantic segmentation of the eye for gaze tracking. In: *2019 IEEE/CVF international conference on computer vision workshop (ICCVW)*, pp 3698–3702. <https://doi.org/10.1109/ICCVW.2019.00568>
- Chernyak I, Chernyak G, Bland JKS, et al (2021) 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*. Association for Computing Machinery, New York, NY, USA, ETRA '21 Full Papers, pp 1–9. <https://doi.org/10.1145/3448017.3457383>
- Chugh S (2020) *An eye tracking system for a virtual reality headset*. M.A.S., University of Toronto, Ann Arbor, United States, URL <https://www.proquest.com/docview/2467610518/abstract/A2F4AD0C9231462CPQ/1>, ISBN: 9798698549024
- Clay V, König P, König S (2019) Eye Tracking in Virtual Reality. *J Eye Mov Res*. <https://doi.org/10.16910/jemr.12.1.3>
- Cowan A, Chen J, Mingo S et al (2021) Virtual reality versus dry laboratory models: comparing automated performance metrics and cognitive workload during robotic simulation training. *J Endourol* 35(10):1571–1576. <https://doi.org/10.1089/end.2020.1037>
- Cubero CG (2020) *Prediction of choice using eye tracking and VR*. PhD thesis, Aalborg University, Aalborg, URL <https://projekter.aau.dk/projekter/en/studentthesis/predi>

- ction-of-choice-using-eye-tracking-and-vr(99c2b0d8-5a45-477b-9308-ebf2d04a1b1a).html
- Davis R (2021) The feasibility of using virtual reality and eye tracking in research with older adults with and without Alzheimer's disease. *Front Aging Neurosci* 13:350. <https://doi.org/10.3389/fnagi.2021.607219>
- De Valois RL, De Valois KK (1980) Spatial Vision. *Annu Rev Psychol* 31(1):309–341. <https://doi.org/10.1146/annurev.ps.31.020180.001521>
- De Kloe YJR, Hooge ITC, Kemner C et al (2022) Replacing eye trackers in ongoing studies: a comparison of eye-tracking data quality between the Tobii Pro TX300 and the Tobii Pro Spectrum. *Infancy off J Int Soc Infant Stud* 27(1):25–45. <https://doi.org/10.1111/infa.12441>
- de Lope J, Grana M (2022) Deep transfer learning-based gaze tracking for behavioral activity recognition. *Neurocomputing* 500:518–527. <https://doi.org/10.1016/j.neucom.2021.06.100>
- Demir I, Ciftci UA (2021) Where do deep fakes look? Synthetic face detection via gaze tracking. In: ACM symposium on eye tracking research and applications. Association for computing machinery, New York, NY, USA, ETRA '21 Full Papers, pp 1–11, <https://doi.org/10.1145/3448017.3457387>
- Diemer J, Muhlberger A, Yassouridis A et al (2023) Distraction versus focusing during VR exposure therapy for acrophobia: A randomized controlled trial. *J Behav Ther Exp Psychiatry* 81(101):860. <https://doi.org/10.1016/j.jbtep.2023.101860>
- Dong M, Zhao J, Wang D, et al (2020) Central-eye: gaze tracking research on visual rendering method in industrial virtual reality Scene. In: Proceedings of the ACM turing celebration conference -China. Association for computing machinery, New York, NY, USA, ACM TURC'20, pp 51–57, <https://doi.org/10.1145/3393527.3393537>
- Drakopoulos P, Ga K, Mania K (2021) Eye tracking interaction on unmodified mobile VR headsets using the selfie camera. *ACM Trans Appl Percept* 18(3):111–1120. <https://doi.org/10.1145/3456875>
- Drakopoulos P, Koulieris GA, Mania K (2020) Front camera eye tracking for mobile VR. In: 2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), pp 642–643, <https://doi.org/10.1109/VRW50115.2020.00172>
- Duchowski AT (2017) Eye tracking methodology: theory and practice, 3rd edn. Springer, Cham
- Duchowski AT, Krejtz K, Gehrler NA, et al (2020) The low/high index of pupillary activity. In: Proceedings of the 2020 CHI conference on human factors in computing systems. association for computing machinery, New York, NY, USA, CHI '20, pp 1–12, <https://doi.org/10.1145/3313831.3376394>
- Emery KJ, Zannoli M, Warren J, et al (2021) 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. Association for Computing Machinery, New York, NY, USA, ETRA '21 Short Papers, pp 1–7, <https://doi.org/10.1145/3448018.3457996>
- Fehlmann B, Mueller FD, Wang N et al (2023) Virtual reality gaze exposure treatment reduces state anxiety during public speaking in individuals with public speaking anxiety: a randomized controlled trial. *J Affect Disord Rep* 14:100627. <https://doi.org/10.1016/j.jadr.2023.100627>
- Finke J, Roesmann K, Stalder T et al (2021) Pupil dilation as an index of Pavlovian conditioning. A systematic review and meta-analysis. *Neurosci Biobehav Rev* 130:351–368. <https://doi.org/10.1016/j.neubiorev.2021.09.005>
- Franke L, Fink L, Martschinke J et al (2021) Time-warped foveated rendering for virtual reality headsets. *Comput Graph Forum* 40(1):110–123. <https://doi.org/10.1111/cgf.14176>
- Fromm CA, Huxlin K, Diaz GJ (2019) Using virtual reality with integrated eye tracking for visual rehabilitation. *Frameless* 1(1):1–2. <https://doi.org/10.14448/Frameless.01.003>
- Fuhl W, Kubler T, Sippel K et al (2015) ExCuSe: robust pupil detection in real-world scenarios, vol 9256. Springer, Cham. <https://doi.org/10.1007/978-3-319-23192-14>
- Fuhl W, Kasneci E (2021) A multimodal eye movement dataset and a multimodal eye movement segmentation analysis. In: ACM symposium on eye tracking research and applications. Association for computing machinery, New York, NY, USA, ETRA '21 Short Papers, pp 1–7, <https://doi.org/10.1145/3448018.3458004>
- Fuhl W, Santini T, Kasneci G, et al (2016) PupilNet: convolutional neural networks for robust pupil detection. arXiv <https://doi.org/10.48550/arXiv.1601.04902>
- Fuhl W, Rosenstiel W, Kasneci E (2019) 500,000 images closer to eyelid and pupil segmentation. In: Vento M, Percannella G (eds) Computer analysis of images and patterns. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 336–347, <https://doi.org/10.1007/978-3-030-29888-327>
- Fuhl W, Gao H, Kasneci E (2020) Tiny convolution, decision tree, and binary neuronal networks for robust and real time pupil outline estimation. In: ACM symposium on eye tracking research and applications. Association for Computing Machinery, New York, NY, USA, ETRA '20 Short Papers, pp 1–5, <https://doi.org/10.1145/3379156.3391347>
- Gadin V (2021) Factors for good text legibility: eye-tracking in virtual reality. PhD thesis, Uppsala University, Uppsala, URL <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-447167>
- Galuret S, Vallee N, Tronchot A et al (2023) Gaze behavior is related to objective technical skills assessment during virtual reality simulator-based surgical training: a proof of concept. *Int J Comput Assist Radiol Surg*. <https://doi.org/10.1007/s11548-023-029618>
- Garbin SJ, Komogortsev O, Cavin R, et al (2020) Dataset for eye tracking on a virtual reality platform. In: ACM symposium on eye tracking research and applications. Association for Computing Machinery, New York, NY, USA, ETRA '20 Full Papers, pp 1–10, <https://doi.org/10.1145/3379155.3391317>
- Gemiciglu T, Winters RM, Wang YT, et al (2023) Gaze and tongue: a subtle, hands-free interaction for head-worn devices. In: Extended abstracts of the 2023 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, USA, CHI EA '23, pp 1–4, <https://doi.org/10.1145/3544549.3583930>
- Geraets CNW, Klein Tunte S, Lestestuiver BP et al (2021) Virtual reality facial emotion recognition in social environments: an eye-tracking study. *Internet Interv* 25(100):432. <https://doi.org/10.1016/j.invent.2021.100432>
- Gunther U, Harrington KIS, Dachselt R, et al (2020) Bionic tracking: using eye tracking to track biological cells in virtual reality. In: Bartoli A, Fusiello A (eds) Computer vision – ECCV 2020 Workshops. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 280–297, <https://doi.org/10.1007/978-3-030-66415-218>
- Gudi A, Li X, van Gemert J (2020) Efficiency in real-time webcam gaze tracking. In: Bartoli A, Fusiello A (eds) Computer vision – ECCV 2020 workshops. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 529–543, <https://doi.org/10.1007/978-3-030-66415-2>
- Guenter B, Finch M, Drucker S et al (2012) Foveated 3D graphics. *ACM Trans Graph* 31(6):164:1-164:10. <https://doi.org/10.1145/2366145.2366183>
- Halle M (1998) Multiple viewpoint rendering. In: Proceedings of the 25th annual conference on computer graphics and interactive techniques. Association for Computing Machinery, New York,

- NY, USA, SIGGRAPH '98, pp 243–254, <https://doi.org/10.1145/280814.280884>
- Hartridge H, Thomson LC (1948) Methods of investigating eye movements. *Brit J Ophthalmol* 32(9):581–591
- Heilemann F, Zimmermann G, Münster P (2022) Accessible hands-free input methods for VR games. In: Miesenberger K, Kouroupetoglou G, Mavrou K, et al (eds) *Computers helping people with special needs*. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 371–379, <https://doi.org/10.1007/978-3-031-08648-943>
- Hickson S, Dufour N, Sud A, et al (2019) Eyemotion: classifying facial expressions in VR using eye-tracking cameras. In: 2019 IEEE winter conference on applications of computer vision (WACV), pp 1626–1635, <https://doi.org/10.1109/WACV.2019.00178>
- Hladky J, Seidel HP, Steinberger M (2019) The camera offset space: real-time potentially visible set computations for streaming rendering. *ACM Trans Graph* 38(6):231:1–231:14. <https://doi.org/10.1145/3355089.3356530>
- Hougaard BI, Knoche H, Jensen J et al (2021) Spatial neglect midline diagnostics from virtual reality and eye tracking in a free-viewing environment. *Front Neurosci* 12(742445):1–13. <https://doi.org/10.3389/fpsyg.2021.742445>
- Howard IP (2002) *Seeing in depth, Vol. 1: Basic mechanisms*. Seeing in depth, vol. 1: Basic mechanisms. University of Toronto Press, Toronto, ON, Canada, p v, 659
- Huey EB (1968) *The Psychology and Pedagogy of Reading*. URL <https://mitpress.mit.edu/9780262580106/the-psychology-and-pedagogy-of-reading/>
- Huizeling E, Peeters D, Hagoort P (2021) Prediction of upcoming speech under fluent and disfluent conditions: eye tracking evidence from immersive virtual reality. *Lang Cognit Neurosci*. <https://doi.org/10.1080/23273798.2021.1994621>
- Illahi GK, Siekkinen M, Kamarainen T, et al (2020) On the interplay of foveated rendering and video encoding. In: 26th ACM symposium on virtual reality software and technology. Association for computing machinery, New York, NY, USA, VRST'20, pp 1–3, <https://doi.org/10.1145/3385956.3422126>
- Illahi GK, Siekkinen M, Kamarainen T, et al (2022) Real-time gaze prediction in virtual reality. In: Proceedings of the 14th international workshop on immersive mixed and virtual environment systems. Association for Computing Machinery, New York, NY, USA, MMVE'22, pp 12–18, <https://doi.org/10.1145/3534086.3534331>
- Jacob RJK, Karn KS (2003) Commentary on section 4 -eye tracking in human-computer interaction and usability research: ready to deliver the promises. In: Hyona J, Radach R, Deubel H (eds) *The Mind's Eye*. North-Holland, Amsterdam, pp 573–605, <https://doi.org/10.1016/B978-044451020-4/50031-1>
- Joo HJ, Jeong HY (2020) A study on eye-tracking-based Interface for VR/AR education platform. *Multimed Tools Appl* 79(23):16719–16730. <https://doi.org/10.1007/s11042-019-08327-0>
- Jurik V, Sidekova Z, Ugwitz P, et al (2019) Eye-tracking in VR setting: implementation for cross-cultural research. In: 20th European conference on eye movements, alicante (Spain), URL <https://is.muni.cz/publication/1550558/cs/Eye-tracking-in-VR-setting-implementation-for-cross-cultural-research/Jurik-Jurikova-Pavel-Ugwitz-Chmelik>
- Just MA, Carpenter PA (1980) A theory of reading: From eye fixations to comprehension. *Psychol Rev* 87(4):329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Kang Z, Jeon J, Salehi S (2020) Eye tracking data analytics in virtual reality training: application in deepwater horizon oil drilling operation. *Proc Hum Factors Ergon Soc Annu Meet* 64(1):821–825. <https://doi.org/10.1177/1071181320641191>
- Kaplanyan AS, Sochenov A, Leimkuhler T et al (2019) DeepFovea: neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Trans Graph* 38(6):212:1–212:13. <https://doi.org/10.1145/3355089.3356557>
- Kar A, Corcoran P (2017) A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access* 5:16,495–16,519. <https://doi.org/10.1109/ACCESS.2017.2735633>, conference Name: IEEE Access
- Katrychuk D, Griffith HK, Komogortsev OV (2019) Power-efficient and shift-robust eye-tracking sensor for portable VR headsets. In: Proceedings of the 11th ACM symposium on eye tracking research and applications. Association for computing machinery, New York, NY, USA, ETRA '19, pp 1–8, <https://doi.org/10.1145/3314111.3319821>
- Khatri J, Moghaddasi M, Llanes-Jurado J, et al (2020) Optimizing virtual reality eye tracking fixation algorithm thresholds based on shopper behavior and age. In: Stephanidis C, Antona M (eds) *HCI International 2020 -Posters*. Springer International Publishing, Cham, Communications in computer and information Science, pp 64–69, <https://doi.org/10.1007/978-3-030-50729-99>
- Khokhar A, Yoshimura A, Borst CW (2019) Pedagogical agent responsive to eye tracking in educational VR. In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR), pp 1018–1019, <https://doi.org/10.1109/VR.2019.8797896>
- Kim J, Stengel M, Majercik A, et al (2019) 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. Association for Computing Machinery, New York, NY, USA, CHI '19, pp 1–12, <https://doi.org/10.1145/3290605.3300780>
- Kobylinksi P, Pochwatko G (2020) Detection of strong and weak moments in cinematic virtual reality narration with the use of 3D eye tracking. In: The thirteenth international conference on advances in computer-human interactions, Valencia (Spain), pp 185–189
- Komoriya K, Sakurai T, Seki Y, et al (2021) User interface in virtual space using VR device with eye tracking. In: Ahram T, Taiar R, Langlois K, et al (eds) *Human interaction, emerging technologies and future applications III*. Springer International Publishing, Cham, Advances in intelligent systems and computing, pp 316–321, <https://doi.org/10.1007/978-3-030-55307-4>
- Kothari R, Yang Z, Kanan C et al (2020) Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Sci Rep* 10(1):2539. <https://doi.org/10.1038/s41598-020-59251-5>
- Kothari RS, Bailey RJ, Kanan C et al (2022) EllSeg-Gen, towards domain generalization for head-mounted eyetracking. *Proc ACM Human-Comput Interact* 6:139:1–139:17. <https://doi.org/10.1145/3530880>
- Koulieris GA, Aksit K, Stengel M et al (2019) Near-Eye display and tracking technologies for virtual and augmented reality. *Comput Graph Forum* 38(2):493–519. <https://doi.org/10.1111/cgf.13654>
- Miller HL, Raphael Zurutuza I, Fears N, et al (2021) Post-processing integration and semi-automated analysis of eye-tracking and motion-capture data obtained in immersive virtual reality environments to measure visuomotor integration. In: ACM symposium on eye tracking research and applications. 17, Association for computing machinery, New York, NY, USA, pp 1–4 <https://doi.org/10.1145/3450341.3458881>
- Laivuori N (2021) Eye and hand tracking in VR training application. URL <http://www.theseus.fi/handle/10024/503405>, accepted: 2021-06-14T10:18:30Z
- Lamb M, Brundin M, Perez Luque E et al (2022) Eye-tracking beyond peripersonal space in virtual reality: validation and best practices. *Front Virtual Real*. <https://doi.org/10.3389/frvir.2022.864653>

- Lee S, Hong M, Kim S et al (2020) Effect analysis of virtual-reality vestibular rehabilitation based on eye-tracking. *KSII Trans Internet Inf Syst (TIIS)* 14(2):826–840. <https://doi.org/10.3837/tiis.2020.02.020>
- Leigh RJ, Zee DS (2015) *The neurology of eye movements*. Oxford University Press, google-Books-ID: v2s0BwAAQBAJ
- Li J, Barmaki R (2019) Trends in virtual and augmented reality research: a review of latest eye tracking research papers and beyond. *Math Comput Sci*. <https://doi.org/10.20944/preprints201909.0019.v1>
- Li B, Zhang Y, Zheng X, et al (2019) A smart eye tracking system for virtual reality. In: 2019 IEEE MTT-S international microwave biomedical conference (IMBioC), pp 1–3. <https://doi.org/10.1109/IMBIOC.2019.8777841>
- Li F, Lee CH, Feng S, et al (2021) Prospective on eye-tracking-based studies in immersive virtual reality. In: 2021 IEEE 24th international conference on computer supported cooperative work in design (CSCWD), pp 861–866. <https://doi.org/10.1109/CSCWD49262.2021.9437692>
- Li Z (2014) *Understanding vision: theory, models, and data*. Oxford University Press, google-Books-ID: 9DlsAwAAQBAJ
- Lim JZ, Mountstephens J, Teo J (2021) Exploring pupil position as an eye-tracking feature for four-class emotion classification in VR. *J Phys Conf Ser* 2129(1):012069. <https://doi.org/10.1088/1742-6596/2129/1/012069>
- Liu H, Laeng B, Czajkowski NO (2020) Does stereopsis improve face identification? A study using a virtual reality display with integrated eye-tracking and pupillometry. *Acta Physiol (oxf)* 210(103):142. <https://doi.org/10.1016/j.actpsy.2020.103142>
- Livingstone M, Hubel D (1988) Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240(4853):740–749. <https://doi.org/10.1126/science.3283936>
- Llanes-Jurado J, Marin-Morales J, Guixeres J et al (2020) Development and calibration of an eye-tracking fixation identification algorithm for immersive virtual reality. *Sensors* 20(17):4956. <https://doi.org/10.3390/s20174956>
- Llanes-Jurado J, Marin-Morales J, Moghaddasi M, et al (2021) Comparing eye tracking and head tracking during a visual attention task in immersive virtual reality. In: Kurosu M (ed) *Human-computer interaction. interaction techniques and novel applications*. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 32–43. <https://doi.org/10.1007/978-3-030-78465-23>
- Lohr D, Komogortsev OV (2022) Eye know you too: toward viable end-to-end eye movement biometrics for user authentication. *IEEE Trans Inf Forensics Secur* 17:3151–3164. <https://doi.org/10.1109/TIFS.2022.3201369>
- Lu S, Li R, Jiao J et al (2020) An eye gaze tracking method of virtual reality headset using a single camera and multi-light source. *J Phys Conf Ser* 1518(1):012020. <https://doi.org/10.1088/1742-6596/1518/1/012020>
- Lu C, Chakravarthula P, Liu K, et al (2022) 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), pp 375–383. <https://doi.org/10.1109/ISMAR558272022.00053>
- Luro FL, Sundstedt V (2019) A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In: *Proceedings of the 11th ACM symposium on eye tracking research and applications*. Association for Computing Machinery, New York, NY, USA, ETRA '19, pp 1–9. <https://doi.org/10.1145/3317956.3318153>
- Maraj C, Hurter J, Pruitt J (2021) Using head-mounted displays for virtual reality: investigating subjective reactions to eye-tracking scenarios. In: Chen JYC, Fragomeni G (eds) *Virtual, augmented and mixed reality*. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 381–394. <https://doi.org/10.1007/978-3-030-77599-527>
- Martin D, Malpica S, Gutierrez D et al (2022) Multimodality in VR: a survey. *ACM Comput Surv* 54(10s):216:1–216:36. <https://doi.org/10.1145/3508361>
- Martinez-Almeida Nistal I, Lampreave Acebes P, Martinez-de-la Casa JM et al (2021) Validation of virtual reality system based on eye-tracking technologies to support clinical assessment of glaucoma. *Eur J Ophthalmol* 31(6):3080–3086. <https://doi.org/10.1177/1120672120976047>
- Marwecki S, Wilson AD, Ofek E, et al (2019) *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*. Association for Computing Machinery, New York, NY, USA, UIST '19, pp 777–789. <https://doi.org/10.1145/3332165.3347919>
- Matthews S, Uribe-Quevedo A, Theodorou A (2020) Rendering optimizations for virtual reality using eye-tracking. In: 2020 22nd symposium on virtual and augmented reality (SVR), pp 398–405. <https://doi.org/10.1109/SVR51698.2020.00066>
- McMurrough CD, Metsis V, Rich J, et al (2012) An eye tracking dataset for point of gaze detection. In: *Proceedings of the symposium on eye tracking research and applications. association for computing machinery*, New York, NY, USA, ETRA '12, pp 305–308. <https://doi.org/10.1145/2168556.2168622>
- McNamara A, Mehta R (2020) Additional insights: using eye tracking and brain sensing in virtual reality. In: *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, CHIEA'20, pp 1–4. <https://doi.org/10.1145/3334480.3375060>
- Mehring W, Wirth M, Risch F, et al (2021) Hess screen revised: how eye tracking and virtual reality change strabismus assessment. In: 2021 43rd annual international conference of the IEEE engineering in medicine biology society (EMBC), pp 2058–2062. <https://doi.org/10.1109/EMBC46164.2021.9631002>
- Meißner M, Pfeiffer J, Pfeiffer T et al (2019) Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *J Bus Res* 100:445–458. <https://doi.org/10.1016/j.jbusres.2017.09.028>
- Melendrez-Ruiz J, Goisbault I, Charrier JC et al (2021) An exploratory study combining eye-tracking and virtual reality: are pulses good “eye-catchers” in virtual supermarket shelves? *Front Virtual Real* 2:68. <https://doi.org/10.3389/frvir.2021.655273>
- Meng X, Du R, Varshney A (2020) Eye-dominance-guided foveated rendering. *IEEE Trans Visual Comput Graph* 26(5):1972–1980. <https://doi.org/10.1109/TVCG.2020.2973442>
- Mirault J, Guerre-Genton A, Dufau S et al (2020) Using virtual reality to study reading: an eye-tracking investigation of transposed-word effects. *Methods Psychol* 3(100):029. <https://doi.org/10.1016/j.metip.2020.100029>
- Mohanto B, Islam AT, Gobbetti E et al (2022) An integrative view of foveated rendering. *Comput Graph* 102:474–501. <https://doi.org/10.1016/j.cag.2021.10.010>
- Mutasim AK, Stuerzlinger W, Batmaz AU (2020) Gaze tracking for eye-hand coordination training systems in virtual reality. In: *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, CHI EA '20, pp 1–9. <https://doi.org/10.1145/3334480.3382924>
- Naqvi RA, Arsalan M, Batchuluun G et al (2018) Deep learning-based gaze detection system for automobile drivers using a NIR camera sensor. *Sensors* 18(2):456. <https://doi.org/10.3390/s18020456>
- NVIDIA Corporation (2018) *VRWorks-Multi-View Rendering (MVR)*. URL <https://developer.nvidia.com/vrworks/graphics/multiview>

- NVIDIA Corporation (2020) GL nv shading rate image. URL <https://registry.khronos.org/OpenGL/extensions/NV/NVshading-rate-image.txt>
- Ogura K, Sugano M, Takabatake S, et al (2019) VR application for visual field measurement of unilateral spatial neglect patients using eye tracking. In: 2019 IEEE international conference on healthcare informatics (ICHI), pp 1–2, <https://doi.org/10.1109/ICHI.2019.8904558>
- Ortega JD, Canas PN, Nieto M et al (2022) Challenges of large-scale multi-camera datasets for driver monitoring systems. *Sensors* 22(7):2554. <https://doi.org/10.3390/s22072554>
- Ou WL, Kuo TL, Chang CC et al (2021) Deep-learning-based pupil center detection and tracking technology for visible-light wearable gaze tracking devices. *Appl Sci* 11(2):851. <https://doi.org/10.3390/app11020851>
- Özel E (2019) Construction site hazard recognition skills measurement via eye-tracking and immersive virtual reality technologies. Master's thesis, Middle East Technical University, URL <https://open.metu.edu.tr/handle/11511/45135>
- Palmer SE (1999) Vision science: photons to phenomenology. MIT Press, google-Books-ID: mNrxCwAAQBAJ
- Palmero C, Sharma A, Behrendt K et al (2021) OpenEDS2020 challenge on gaze tracking for VR: dataset and results. *Sensors* 21(14):4769. <https://doi.org/10.3390/s21144769>
- Park JH, Jeon HJ, Lim EC et al (2019) Feasibility of eye tracking assisted vestibular rehabilitation strategy using immersive virtual reality. *Clin Exp Otorhinolaryngol* 12(4):376–384. <https://doi.org/10.21053/ceo.2018.01592>
- Pastel S, Marlok J, Bindow N et al (2023) Application of eye-tracking systems integrated into immersive virtual reality and possible transfer to the sports sector -a systematic review. *Multimed Tools Appl* 82(3):4181–4208. <https://doi.org/10.1007/s11042-022-13474-y>
- Patney A, Salvi M, Kim J et al (2016) Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans Graph* 35(6):179:1-179:12. <https://doi.org/10.1145/2980179.2980246>
- Pettersson J, Falkman P (2021) Human movement direction prediction using virtual reality and eye tracking. In: 2021 22nd IEEE international conference on industrial technology (ICIT), pp 889–894, <https://doi.org/10.1109/ICIT46573.2021.9453581>
- Pettersson J, Falkman P (2020) Human movement direction classification using virtual reality and eye tracking. *Proc Manuf* 51:95–102. <https://doi.org/10.1016/j.promfg.2020.10.015>
- Pettersson J (2021) Data-driven human intention analysis: supported by virtual reality and eye tracking. licentiate, Chalmers University of Technology, Ann Arbor, United States, URL <https://www.proquest.com/docview/2611626919/abstract/5C36052A9DA244BAPQ/1>, ISBN: 9798496572606
- Pfeiffer J, Pfeiffer T, Meißner M et al (2020) Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. *Inf Syst Res* 31(3):675–691. <https://doi.org/10.1287/isre.2019.0907>
- Plopski A, Hirzle T, Norouzi N et al (2022) The eye in extended reality: a survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Comput Surv.* <https://doi.org/10.1145/3491207>
- Plopski A, Orlosky J, Itoh Y, et al (2016) Automated spatial calibration of HMD systems with unconstrained eye-cameras. In: 2016 IEEE international symposium on mixed and augmented reality (ISMAR), pp 94–99, <https://doi.org/10.1109/ISMAR.2016.16>
- Porras-Garcia B, Ferrer-Garcia M, Ghita A et al (2019) The influence of gender and body dissatisfaction on body-related attentional bias: an eye-tracking and virtual reality study. *Int J Eat Disord* 52(10):1181–1190. <https://doi.org/10.1002/eat.23136>
- Qian K, Arichi T, Price A et al (2021) An eye tracking based virtual reality system for use inside magnetic resonance imaging systems. *Sci Rep* 11(1):16301. <https://doi.org/10.1038/s41598-021-95634-y>
- Ranti C, Jones W, Klin A et al (2020) Blink rate patterns provide a reliable measure of individual engagement with scene content. *Sci Rep* 10(1):8267. <https://doi.org/10.1038/s41598-020-64999-x>
- Rappa NA, Ledger S, Teo T et al (2019) The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review. *Interact Learn Environ.* <https://doi.org/10.1080/10494820.2019.1702560>
- Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 124(3):372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Reichenberger J, Pfaller M, Muhlberger A (2020) Gaze behavior in social fear conditioning: an eye-tracking study in virtual reality. *Front Psychol* 11:35. <https://doi.org/10.3389/fpsyg.2020.00035>
- Ryabinin K, Belousov K, Chumakov R (2021) Visual analytics tools for polycode stimuli eye gaze tracking in virtual reality. In: Proceedings of the 31th international conference on computer graphics and vision. Volume 2. Keldysh Institute of Applied Mathematics, pp 211–222, <https://doi.org/10.20948/graphicon-2021-3027-211-222>
- Schauffer G, Sturzlinger W (1996) A three dimensional image cache for virtual reality. *Comput Graph Forum* 15(3):227. <https://doi.org/10.1111/1467-8659.1530227>
- Shiferaw B, Downey L, Crewther D (2019) A review of gaze entropy as a measure of visual scanning efficiency. *Neurosci Biobehav Rev* 96:353–366. <https://doi.org/10.1016/j.neubiorev.2018.12.007>
- Sipatchin A, Wahl S, Rifai K (2020) Accuracy and precision of the HTC VIVE PRO eye tracking in head-restrained and head-free conditions. *Invest Ophthalmol vis Sci* 61(7):5071
- Sipatchin A, Wahl S, Rifai K (2021) Eye-tracking for clinical ophthalmology with virtual reality (VR): a case study of the HTC vive pro eye's usability. *Healthcare* 9(2):180. <https://doi.org/10.3390/healthcare9020180>
- Sithi-amorn P, Lawrence J, Yang L et al (2008) An Improved shading cache for modern GPUs. *Eurogr Assoc.* <https://doi.org/10.2312/EGGH/EGGH08/095-101>
- Slovak M, Anyz J, Erlebach J et al (2022) Emotional arousal in patients with functional movement disorders: a pupillometry study. *J Psychosom Res.* <https://doi.org/10.1016/j.jpsychores.2022.111043>
- Soret R, Charras P, Khazar I, et al (2020) Eye-tracking and Virtual Reality in 360-degrees: exploring two ways to assess attentional orienting in rear space. In: ACM symposium on eye tracking research and applications. Association for computing machinery, New York, NY, USA, ETRA '20 Adjunct, pp 1–7, <https://doi.org/10.1145/3379157.3391418>
- Spjut J, Boudaoud B, Kim J et al (2020) Toward standardized classification of foveated displays. *IEEE Trans Visual Comput Graph* 26(5):2126–2134. <https://doi.org/10.1109/TVCG.2020.2973053>
- Stein N (2021) Analyzing visual perception and predicting locomotion using virtual reality and eye tracking. In: 2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), pp 727–728, <https://doi.org/10.1109/VRW52623.2021.00246>
- Stern JA, Boyer D, Schroeder D (1994) Blink rate: a possible measure of fatigue. *Hum Factors* 36(2):285–297. <https://doi.org/10.1177/001872089403600209>
- Sterna R, Cybulski A, Igras-Cybulska M, et al (2021) Psychophysiology, eye-tracking and VR: exemplary study design. In: 2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), pp 639–640, <https://doi.org/10.1109/VRW52623.2021.00202>
- Stoeve M, Wirth M, Farlock R et al (2022) Eye tracking-based stress classification of athletes in virtual reality. *Proc ACM Comput*

- Graph Interact Tech 5(2):191–1917. <https://doi.org/10.1145/3530796>
- Sun J, Zhang H, Chen L et al (2021) 29–3: an easy-to-implement and low-cost VR gaze-tracking system. *SID Symp Digest Tech Pap* 52(1):373–375. <https://doi.org/10.1002/sdtp.14693>
- Tabbaa L, Searle R, Bafti SM et al (2021) VREED: virtual reality emotion recognition dataset using eye tracking and physiological measures. *Proc ACM Interact Mobile Wearable Ubiquitous Technol* 5(4):178:1-178:20. <https://doi.org/10.1145/3495002>
- Tan S, Lo Y, Li C, et al (2020) Eye-tracking aided VR system for amblyopic pediatric treatment difficulty adjustment. In: 2020 international conference on virtual reality and intelligent systems (icvr18), pp 47–50, <https://doi.org/10.1109/icvr18.2020.00019>
- Tariq T, Tursun C, Didyk P (2022) Noise-based enhancement for foveated rendering. *ACM Trans Graph* 41(4):143:1-143:14. <https://doi.org/10.1145/3528223.3530101>
- Tian P, Wang Y, Lu Y, et al (2019) Behavior analysis of indoor escape route-finding based on head-mounted vr and eye tracking. In: 2019 international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (Smart-Data), pp 422–427, <https://doi.org/10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00090>
- Tobii (2022) Most advanced eye tracking system—Tobii Pro Spectrum. URL <https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-spectrum>
- Tobii (2022) Tobii Customer Portal. URL <https://connect.tobii.com>
- Tonsen M, Zhang X, Sugano Y, et al (2016) Labelled pupils in the wild: a dataset for studying pupil detection in unconstrained environments. In: Proceedings of the Ninth Biennial ACM symposium on eye tracking research and applications. Association for computing machinery, New York, NY, USA, ETRA '16, pp 139–142, <https://doi.org/10.1145/2857491.2857520>,
- Tursun OT, Arabadzhyska-Koleva E, Wernikowski M et al (2019) Luminance-contrast-aware foveated rendering. *ACM Trans Graph* 38(4):98:1-98:14. <https://doi.org/10.1145/3306346.3322985>
- Unterguggenberger J, Kerbl B, Steinberger M et al (2020) Fast multi-view rendering for real-time applications. *The Eurogr Assoc.* <https://doi.org/10.2312/pgv.20201071>
- Valori I, McKenna-Plumley PE, Bayramova R et al (2020) Pro-priocceptive accuracy in immersive virtual reality: a developmental perspective. *PLoS ONE* 15(1):e0222. <https://doi.org/10.1371/journal.pone.0222253>
- Visconti A, Calandra D, Lamberti F (2023) Comparing technologies for conveying emotions through realistic avatars in virtual reality-based meta-verse experiences. *Comput Animat Virtual Worlds.* <https://doi.org/10.1002/cav.2188>
- Wandell BA (1995) Foundations of vision. Foundations of vision, Sinauer Associates, Sunderland, MA, US, p xvi 476
- Wang CC, Hung JC, Chen HC (2021) How prior knowledge affects visual attention of Japanese mimicry and onomatopoeia and learning outcomes: evidence from virtual reality eye tracking. *Sustainability* 13(19):11058. <https://doi.org/10.3390/su131911058>
- Wang CC, Wang SC, Chu CP (2019) Combining Virtual Reality Advertising and Eye Tracking to Understand Visual Attention: A Pilot Study. In: 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), pp 160–165, <https://doi.org/10.1109/IIAI-AAI.2019.00041>,
- Wang G, Gan Q, Li Y (2020) Research on attention-guiding methods in cinematic virtual reality based on eye tracking analysis. In: 2020 International conference on innovation design and digital technology (ICIDDT), pp 68–72, <https://doi.org/10.1109/ICIDDT52279.2020.00020>,
- Wechsler T, Bahr LM, Muhlberger A (2019) Can gaze behaviour predict stress response and coping during acute psychosocial stress? A virtual-reality-based eyetracking study. *Eur Neuropsychopharmacol* 29:S524. <https://doi.org/10.1016/j.euroneuro.2018.11.777>
- Weier M, Maiero J, Roth T, et al (2014) Lazy details for large high-resolution displays. In: SIGGRAPH Asia 2014 Posters. Association for Computing Machinery, New York, NY, USA, SA '14, p 1, <https://doi.org/10.1145/2668975.2669016>
- White S, Natalie J, Coulter D, et al (2021) Cascaded Shadow Maps. URL <https://learn.microsoft.com/en-us/windows/win32/dxtechharts/cascaded-shadow-maps>
- Wong ET, Yean S, Hu Q, et al (2019) Gaze Estimation Using Residual Neural Network. In: 2019 IEEE international conference on pervasive computing and communications workshops (PerCom Workshops), pp 411–414, <https://doi.org/10.1109/PERCOMW.2019.8730846>
- Wood E, Baltrušaitis T, Morency LP, et al (2016) Learning an appearance-based gaze estimator from one million synthesised images. In: Proceedings of the ninth Biennial ACM symposium on eye tracking research and applications. Association for computing machinery, New York, NY, USA, ETRA '16, pp 131–138, <https://doi.org/10.1145/2857491.2857492>
- Xiao K, Liktov G, Vaidyanathan K (2018) Coarse pixel shading with temporal supersampling. In: Proceedings of the ACM SIGGRAPH symposium on interactive 3D graphics and games. Association for computing machinery, New York, NY, USA, I3D '18, pp 1–7, <https://doi.org/10.1145/3190834.3190850>
- Yang TH, Huang JY, Han PH, et al (2021) Saw It or triggered it: exploring the threshold of implicit and explicit interaction for eye-tracking technique in virtual reality. In: 2021 IEEE Conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), pp 482–483, <https://doi.org/10.1109/VRW52623.2021.00123>
- Yaramothu C, Bertagnolli JVd, Santos EM et al (2019) Proceedings #37: virtual eye rotation vision exercises (VERVE): a virtual reality vision therapy platform with eye tracking. *Brain Stimul Basic Transl Clin Res Neuromodul* 12(2):e107–e108. <https://doi.org/10.1016/j.brs.2018.12.206>
- Yarbus AL (1967) Eye movements and vision. Springer, New York
- Yeh PH, Liu CH, Sun MH et al (2021) To measure the amount of ocular deviation in strabismus patients with an eye-tracking virtual reality headset. *BMC Ophthalmol* 21(1):246. <https://doi.org/10.1186/s12886-021-02016-z>
- Zhang LM, Zhang RX, Jeng TS et al (2019) Cityscape protection using VR and eye tracking technology. *J vis Commun Image Represent* 64(102):639. <https://doi.org/10.1016/j.jvcir.2019.102639>
- Zheng Z, Yang Z, Zhan Y, et al (2018) Perceptual model optimized efficient foveated rendering. In: Proceedings of the 24th ACM symposium on virtual reality software and technology. association for computing machinery, New York, NY, USA, VRST '18, pp 1–2, <https://doi.org/10.1145/3281505.3281588>
- Zou W, Feng S, Mao X, et al (2021) Enhancing Quality Of Experience For Cloud Virtual Reality Gaming: An Object-Aware Video Encoding. In: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp 1–6, <https://doi.org/10.1109/ICMEW53276.2021.9455970>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.