IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 September 2024, accepted 13 September 2024, date of publication 16 September 2024, date of current version 30 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3462109

SURVEY

Event Camera-Based Eye Motion Analysis: A Survey

KHADIJA IDDRISU[1,2,3], WASEEM SHARIFF[2,4], (Graduate Student Member, IEEE), PETER CORCORAN[4], (Fellow, IEEE), NOEL E. O'CONNOR[3], (Member, IEEE), JOE LEMLEY[2], (Member, IEEE), AND SUZANNE LITTLE[1,3]

[1]School of Computing, Dublin City University, Dublin 9, D09 DXA0 Ireland
[2]Drowsiness Team, FotoNation-Tobii, Galway, H91 V0TX Ireland
[3]Insight SFI Research Center for Data Analytics, Dublin, D09 DXA0 Ireland
[4]C3I Imaging Laboratory, Department of Electronic and Electrical Engineering, University of Galway, Galway, H91 TK33 Ireland

Corresponding author: Khadija Iddrisu (khadija.iddrisu2@mail.dcu.ie)

This work was supported by the Science Foundation Ireland at the Insight Science Foundation Ireland (SFI) Research Centre for Data Analytics, Dublin City University, in collaboration with FotoNation (Tobii), under Grant 12/RC/2289_P.

ABSTRACT Neuromorphic vision sensors, commonly referred to as Event Cameras (ECs), have gained prominence as a field of research in Computer Vision. This popularity stems from the numerous unique characteristics including High Dynamic Range, High Temporal Resolution, and Low Latency. Of particular interest is their temporal resolution, which proves ideal for human monitoring applications. Capturing rapid facial movements and eye gaze can be effectively achieved with ECs. Recent studies involving the use of ECs for object detection and tracking have demonstrated success in tasks involving Eye Motion Analysis such as Eye tracking, Blink detection, Gaze estimation and Pupil tracking. The objective of this study is to provide a comprehensive review of the current research in the aforementioned tasks, focusing on the potential utilization of ECs for future tasks involving rapid eye motion detection, such as detection and classification of saccades. We highlight studies that may serve as a foundation for undertaking such a task, such as pupil tracking and gaze estimation. We also highlight in our review some common challenges encountered such as the availability of datasets and review some of the methods used in solving this problem. Finally, we discuss some limitations of this field of research and conclude with future directions including real-world applications and potential research directions.

INDEX TERMS Event cameras, eye motion analysis, eye-tracking, pupil segmentation, near-eye, remote-eye.

I. INTRODUCTION

The human eye is a dynamic organ, capable of executing rapid and precise movements that reflect underlying neurological processes and cognitive states. The analysis of eye movements has been a significant area of interest for both clinicians and researchers across various disciplines, including psychology, neurology, and ophthalmology [1]. Fine-grained eye motion analysis can enable or improve real-world applications such as virtual reality [2], attention estimation [3], drowsiness detection [4], diagnostics [5], [6], and many others.

Traditionally measured in a clinical setting, the use of computer vision has provided many advances in eye tracking and gaze estimation [7]. Earlier approaches to eye motion analysis primarily relied on equipment such as eye trackers and coils for scleral search methods. These techniques provided a foundation for further exploration into eye tracking research. A notable shift from these traditional methods has been observed in recent years, with a significant number of studies focusing on video-oculography. This modern approach involves capturing eye movements using cameras and analyzing the data for various tasks.

The associate editor coordinating the review of this manuscript and approving it for publication was Ludovico Minati.

Conventional RGB cameras, though widely used in various applications, are inherently constrained by their limitations in both resolution and event granularity. These constraints significantly hinder their effectiveness in capturing subtle and rapid eye movements, which are essential for understanding complex visual and cognitive processes. The low resolution limits the ability to detect fine details of eye motion, while the restricted event granularity reduces the ability to capture rapid transitions in eye position [8]. These drawbacks become particularly problematic in scenarios where high precision and real-time tracking are critical, such as in neurophysiological studies, gaze-based interaction systems, or diagnostic assessments of eye movement disorders.

Another significant limitation of standard RGB cameras is their high latency, which poses a substantial challenge when attempting to measure real-time changes in dynamic environments. High latency results in delays between the occurrence of an event and its detection, rendering these cameras unsuitable for applications that demand real-time responsiveness [8], [9]. This issue is further aggravated by under-sampling, a problem that occurs with both RGB and near-infrared cameras. Under-sampling refers to the failure to capture sufficient data points within a given timeframe, leading to a loss of critical information, particularly in scenarios involving fast and complex movements. As a result, conventional cameras often fail to provide the temporal and spatial resolution necessary for accurate and detailed eye movement analysis.

Event Cameras (ECs), however, offer a promising alternative for addressing these limitations. Unlike traditional cameras that capture frames at fixed intervals, ECs operate by detecting changes in the scene on a per-pixel basis, resulting in a continuous stream of events rather than discrete frames. This unique operational mechanism allows ECs to achieve significantly lower latency, enabling them to capture rapid eye movements in real-time. Additionally, ECs excel in high-speed environments, as they are capable of detecting even minute changes in eye position with a high degree of temporal precision [8]. This makes them particularly well-suited for applications where detailed, real-time analysis of eye motion is crucial.

The potential of ECs extends beyond just high-speed capture. They also offer the ability to conduct more detailed and precise analyses of eye movements, thanks to their superior temporal resolution and low latency [8], [9]. For example, ECs could greatly enhance the accuracy of saccade detection, microsaccades, and other fast ocular movements that are often missed or inadequately represented by traditional cameras. This improved resolution could have wide-reaching implications, from enhancing user interaction in virtual reality environments to providing deeper insights into neurological conditions that manifest through eye movement patterns.

Despite these considerable advantages, the application of Event Cameras in eye movement analysis remains an emerging field with limited exploration. Current research on the use of ECs in this domain is sparse, and much remains to be investigated in terms of their full potential and limitations [8], [10]. While some studies have demonstrated the feasibility of ECs for eye tracking and motion analysis, these investigations are still in their infancy. The lack of extensive research underscores the need for more comprehensive studies that explore the integration of ECs into existing eye-tracking systems, their performance in various conditions, and their potential for advancing the field of eye movement research. As the technology continues to evolve, further research will be essential to fully harness the capabilities of ECs and to develop new methodologies for analyzing eye movements with unprecedented accuracy and speed.

This review seeks to offer a thorough analysis of the current literature, emphasizing the advancements, challenges, and future directions in the field of eye motion analysis. Special attention is given to the potential of Event Cameras (ECs) to improve the precision and effectiveness of research in this domain. Beyond simply cataloging existing studies, this review is intended to serve as a valuable resource for early-stage researchers and professionals interested in exploring the use of ECs for eye-motion analysis. By synthesizing key findings and identifying gaps in the literature, this review aims to guide future work in this emerging field.

The main contributions of this study are:

1) This review will contribute a detailed synthesis of the current state of research on EC-based eye motion analysis, categorizing and critically analyzing existing studies to highlight key advancements, methodologies, and outcomes. This will provide researchers with a clear overview of the progress made and establish a foundation for further research.
2) Another significant contribution is the identification of task-specific challenges, limitations, and gaps in current knowledge. By thoroughly examining the literature, this review will pinpoint areas where additional studies are needed, offering a roadmap for future research efforts that can address these gaps.
3) The review will propose potential strategic future research directions, focusing on how Event Cameras can be further integrated into eye motion analysis. It will outline potential interdisciplinary collaborations, new application areas, and innovative methodologies that could enhance the accuracy and efficacy of ECs in this field.

In this section we define eye motion analysis according to standard classification of eye movements, provide a background description of how event cameras observe eye movements and distinguish between near and remote-eye sensing. The methodology for the literature review is then presented alongside a pictorial overview (Figure 1). We review available datasets and tools related to this topic, discuss how events can be represented and describe the related applications with associated literature. The survey concludes with an overview of the challenges and limitations

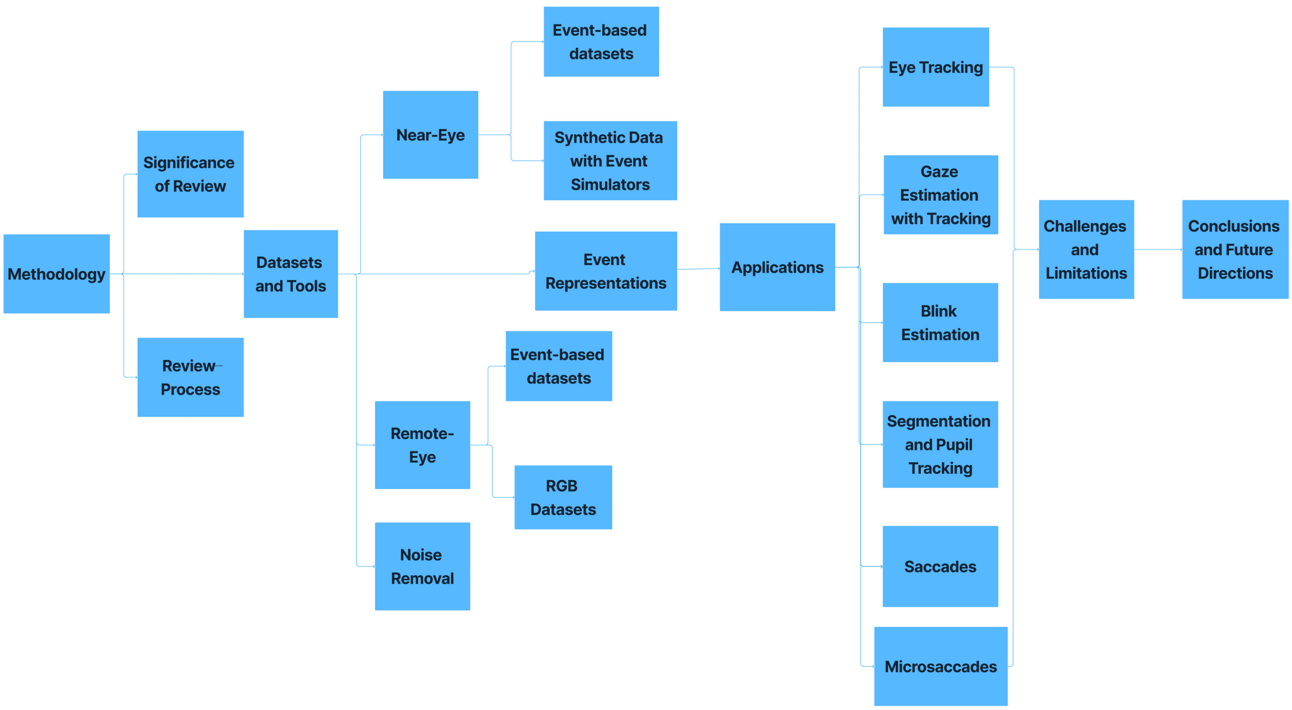

FIGURE 1. Eye motion analysis using event camera. This schematic provides a comprehensive overview of the research topics presented in this manuscript.

in eye motion analysis using event cameras and therefore the potential future directions for research.

A. CLASSIFICATION OF EYE MOVEMENTS

Eye movements includes both voluntary and involuntary movement of the eyes. A specialised type of tissue at the back of the eye, the retina, contains photoreceptors that sense light and eye movements stabilize these images allowing clear vision despite movements of the body and the objects being viewed. Eye motion can range from slight subtle movements to very rapid movements. The minimum angle of eye movements, as measured by Lim et al. [11], ranges from $27.9 \pm 7.6^\circ$ in elevation to 44.9 ± 6.8 in other monocular rotations such as adduction, abduction and depression. The pattern of these eye movements is used in research to understand how people process visual information and make decisions based on what they see.

The primary classifications of eye movements include:

Voluntary Eye Movements

- Fixations: Periods during eye motion when the eyes focus on the visual gaze on a single point.
- Saccades: Contrasted with fixations, saccades are rapid eye movements that change the point of fixation from one location to another.
- Smooth Pursuits: Enables eye gaze to smoothly follow a moving object.
- Listing's Law: Listing's law governs the three-dimensional orientation of the eye and its axes of rotation. It states that, when the head is fixed, there is an eye position called primary position, such that the eye assumes only those orientations that can be reached from primary position by a single rotation about an axis in a plane called Listing's plane [12]

Involuntary Eye Movements

- Micro-saccades: Tiny, involuntary saccades that occur during fixations and play a role in preventing sensory adaptation.
- Vestibulo-ocular Reflex (VOR): Serves to stabilize the gaze during head movement by causing eye movement in response to activation of the vestibular system. This reflex aims to maintain images on the retinas of the eye while the head is in motion.
- Optokinetic Response: A compensatory reflex to stabilize the visual image. This reflex is activated in response to head motion to prevent blurring of the image on the retina. To achieve this, the eyes move reflexively in the same direction as the image motion, thereby minimizing the relative motion of the visual scene on the eye.

The implications of different eye movements in neural responses cannot be underscored. Eye movements such as micro-saccades and blinks have shown associations with sleep and fatigue, a significant indicator for drowsiness estimation, via EEG and EOG signals [13], [14], [15]. Schleicher et al. [16] examine the relationship between eye movement patterns and neural responses in the context of fatigue, highlighting that variations in fixation duration, categorized into very short (less than 150 ms), medium (150-900 ms), and overlong (more than 900 ms), reveal significant changes linked to fatigue. Medium-length fixations decrease, while both very short and overlong fixations increase, indicating a shift from cognitive processing to more reflexive or minimal scanning behaviors as fatigue intensifies. These changes in eye movements, which correlate with EEG behaviors such as changes in blink rate and saccadic motion, reflect a decrease in cognitive engagement and attention, showcasing the potential of eye movement patterns as indicators of underlying neural states and fatigue levels. Additionally, involuntary eye movements like micro-saccades correlate with EEG behaviors, providing insight into underlying cognitive and neural processes such as the rapid evaluation of emotional content in faces, and the disentanglement of initial stimulus processing from subsequent saccade-induced visual responses [17].

B. NEAR-EYE AND REMOTE-EYE SENSING

The utilization of ECs in eye movement analysis falls into two primary categories: Near-eye sensing and Remote-eye sensing. Near-eye sensing involves positioning the camera close to the eyes, a method that enriches the dataset collected, thus improving the accuracy of the analysis. This technique finds its applications primarily in Extended Reality (XR) and Virtual Reality (VR), where precise eye tracking enhances user experiences by making interactions more intuitive and immersive. For example, in VR gaming, near-eye tracking can adjust the game's environment based on where the player is looking, creating a more engaging and realistic experience [18].

Remote-eye sensing on the other hand, places the camera at a distance from the subject. This approach is less intrusive and considered to be more efficient in scenarios where direct interaction with the device is impractical. A notable application is in driver monitoring systems, where cameras installed within the vehicle continuously assess the driver's gaze direction to evaluate attentiveness. By identifying patterns of eye movement that suggest drowsiness or distraction, these systems can trigger alerts, improving driver and road safety. Additionally, remote-eye tracking is instrumental in enhancing user interfaces for accessibility, enabling users with mobility impairments to interact with computers through eye movements alone [19], [20].

C. HOW EVENT CAMERAS VIEW EYE MOVEMENT

Event Cameras (ECs), inspired by their bio-vision sensing technology, operate based on the presence of motion, reporting changes in intensity at the pixel level. These pixel changes are reported asynchronously and immediately, producing a data stream known as events. An event is typically represented as a tuple (t, x, y, p) that contains information about the the time at which the change occurred, t, the exact location of the brightness change (x, y) and the direction of the change referred to as the polarity, p. The polarity is represented as a binary value, $(0, 1)$ or $(-1, 1)$ in some sensors, indicating a decrease or increase in brightness change.

Figure 2 illustrates how ECs produce events when facing a black dot on a rotating disk. Unlike traditional cameras which capture frames in intervals, ECs generate events asynchronously producing data as long as a brightness change occurs in the scene.

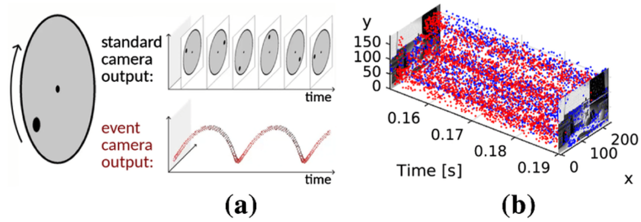

FIGURE 2. **(a) Comparison of the output from a standard frame-based camera and an event-based camera. The standard camera captures discrete frames over time, resulting in multiple snapshots that may miss critical motion details. In contrast, the EC continuously records changes in the scene, capturing the trajectory of movement more accurately with time-resolved data points. (b) A 3D representation of event data, showing the x, y spatial coordinates versus time (t) where each point represents a detected change in the scene. Red and blue dots represent positive and negative polarities, respectively, highlighting the ECs ability to capture fine temporal dynamics [21].**

This results in the elimination of redundant information as it solely generates data when there is motion. This is particularly advantageous for tasks that are dependent on motion, such as eye movements. Furthermore, since events are continuously reported, there is no loss of information in situations involving rapid motion, which often results in under-samping and motion blur in traditional counterparts.

Figure 3 illustrates how different eye parts are captured by ECs. As observed, most eye features are present in some frames while some eye parts are missing in other frames. This shows how ECs produce data by only reporting information in the parts of the eyes where motion is generated during eye movements. For example in the top frames, *(a,b,c,d)*, The iris, pupil, and eyelids are all visible. Whereas, only certain parts of the eyes are visible in the bottom frames, *(e,f,g,h)*, due to less motion generated thus no events are reported for these eye parts. The polarity information is also included to illustrate changes in pixel intensity with pixels with positive polarities indicated in green and red for negative polarities showing the polarity behaviour observed for certain types of eye movement, such as a full and half blink in frame g and h respectively.

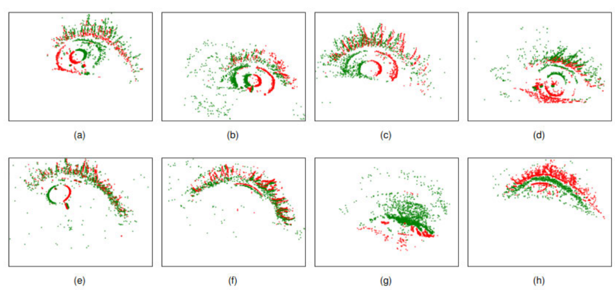

FIGURE 3. **Illustration of eyes captured by an event camera showing different variations of appearance of eye features when triggered by motion. Image source: [22].**

II. METHODOLOGY

This section details the methodology employed in this review, focusing on the distinct advantages of Event Cameras (ECs) for eye motion analysis. It also outlines the review's significance, describes the process followed, and highlights how this review contributes novel insights to the field.

A. SIGNIFICANCE OF REVIEW

The study of eye motion has evolved significantly over the years. Historically, eye motion analysis began with manual observation and has progressed through the use of various devices and computational techniques [23], [24]. Recent advancements include the application of Machine Learning and Deep Learning approaches to enhance eye motion analysis [24], [25], [26]. Eye motion analysis encompasses both invasive methods, which require direct contact with the eyes, and non-invasive methods, which use external devices such as cameras.

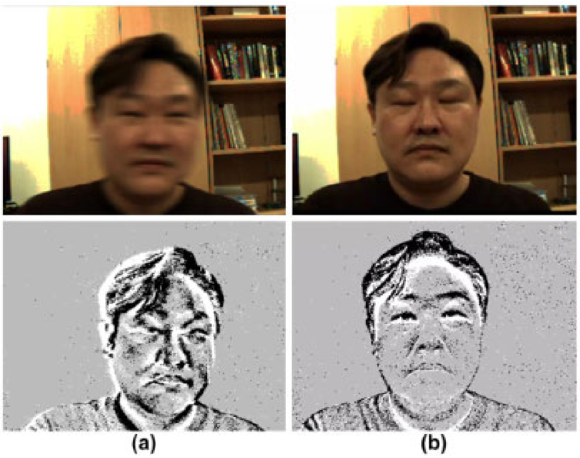

FIGURE 4. An illustration of eyes captured by an RGB camera and their corresponding event frames demonstrating the advantages of event cameras over traditional cameras in terms of motion blur. Image source: [27].

ECs have demonstrated advantages in eye motion analysis tasks such as eye-tracking [28], pupil tracking [27], and blink detection [29] and with more advanced tasks such as eye-gaze estimation [30]. The High Temporal Resolution of ECs has offered the opportunity to capture subtle and rapid eye movements with high temporal precision which may not otherwise be observed in traditional counterparts, making them a useful alternative for the mentioned tasks. For instance, RGB cameras may not fully capture certain eye movements such as blinks due to how quickly they occur. Additionally, common issues such as motion blur (as shown in Figure 4) in conventional cameras have been mitigated. Furthermore, due to their asynchronous nature, ECs report changes in intensity in real-time, resulting in significantly lower latency compared to conventional cameras. This is particularly beneficial for applications requiring real-time feedback, such as adaptive vision aids or human-computer interaction systems. Moreover, ECs have a very High Dynamic Range (HDR) of about 140 dB compared to about 60 dB for standard cameras, enabling them to perform well in a wide range of lighting conditions.

This is particularly useful for eye motion analysis, especially in environments where lighting conditions are highly variable, such as outdoors compared to indoors. They extend the potential use of ECs in more dynamic settings. For instance, ECs are highly beneficial for driver monitoring in vehicular cabins, where lighting levels fluctuate significantly between day and night [20], [31].

ECs are significantly more energy-efficient compared to traditional frame-based cameras. i.e They typically consume only around 30 mW of power, compared to 10W or more for high-speed frame-based cameras, making them well-suited for battery-powered applications [9]. The asynchronous event-based operation of ECs allows them to avoid the energy waste of capturing redundant static background information. Additionally, this sparseness of the data from ECs is more efficient to process than full video frames, reducing the computational load and power requirements compared to traditional cameras.

Finally, ECs are generally considered more privacy-preserving compared to traditional cameras. The key privacy-preserving properties are that they capture only a fraction of the visual information compared to normal cameras, naturally hiding sensitive details, thereby revealing less detailed visual data about the scene and people within it [32]. However, sensitive information might still be recoverable from event data, despite the lack of intensity information through high resolution reconstruction [33]. To address this, techniques like "event encryption" have been developed to enhance privacy protection by encrypting the event data to prevent reconstruction of the original visual scene [34]. Event-based person identification methods have also been developed that can preserve privacy by anonymizing the visual information captured by ECs to ensure privacy [35]. Another study proposed an event scrambling technique that makes the stream uninterpretable to the human eye while still allowing effective application of computer vision models [36].

These advantages make ECs particularly suited for eye motion analysis, offering opportunities for advances in medical diagnostics [37], augmented and virtual reality [38], and other various forms of interactive technology. Nonetheless, this field of research is still evolving, and thus, integrating with existing deep learning systems poses its own set of challenges, such as the need for new algorithms and processing techniques.

B. REVIEW PROCESS

To conduct a comprehensive review of the literature, the initial focus was placed on identifying the state-of-the-art applications of ECs in the analysis of saccades. However, due to a lack of relevant results, the scope of the review was broadened to encompass the wider field of eye motion analysis using ECs. To achieve this, we conducted a systematic and rigorous literature search using a variety of sources, including Google Scholar, Springer, arXiv, IEEE Explore, ScienceDirect (Elsevier), ACM Digital Library and other sources.

Search Keywords: From these sources, we conducted an initial search using the keywords: ''event cameras'', ''neuromorphic sensors'', ''eye tracking'', ''gaze tracking'', ''pupil tracking'', ''blink detection'', ''saccade detection'', ''remote eye tracking'', ''near-eye tracking'', and ''Machine Learning''. Our search was inclusive of articles published from the inception of the database up to the current date of writing. To ensure a comprehensive and up-to-date review, this search was performed daily until the manuscript submission. This approach allowed us to capture the latest studies and developments in the field, ensuring the relevance and timeliness of our review.

To ensure the quality and relevance of the papers included in the review, we applied a set of strict inclusion and exclusion criteria. The inclusion criteria mandated that papers must focus on the use of event cameras for tasks related to eye motion analysis, be written in English, be publicly accessible, and report original research or review existing literature on the topic. Conversely, papers were excluded if they were not written in English, were not available in full-text format, or focused on tasks unrelated to eye motion analysis using ECs.

After identifying a total of 150 potentially relevant papers through the initial search, we screened the titles and abstracts to determine their relevance to the review. This resulted in the selection of 48 papers for full-text review. We then hand-searched the reference lists of these papers to identify any additional relevant studies, resulting in a final total of 38 papers included in the task-specific review. Through this rigorous methodology, we provided a comprehensive and up-to-date review of the current state-of-the-art in the use of ECs for eye motion analysis, highlighting the key challenges and opportunities in this exciting and rapidly evolving field.

III. DATASETS AND TOOLS

As with any machine learning task, a well-curated dataset is essential for training neural networks and obtaining reliable results. In the domain of eye motion analysis using event cameras, data collection plays a particularly crucial role. Two main types of datasets are typically used: Near-Eye Datasets and Remote-Eye Datasets. These datasets are categorized based on the sensing method used for data collection. In this section, we will discuss both types of datasets and provide a comprehensive summary of the tasks they have been employed for. It is important to highlight that, as at the time of this writing, only a few large datasets qualify as benchmarks, as many of the existing datasets either have limited samples or are not publicly accessible. Since event cameras represent a relatively new technology, publicly available datasets remain scarce. Furthermore, another promising approach for training machine learning models in this field is through the use of event simulators. Event simulators can convert conventional RGB datasets into event-based data, providing an alternative means for generating training data. We will also explore simulators, the studies that have utilized them, and discuss potential large RGB datasets that could be processed for event camera analysis using these tools.

A. NEAR-EYE

Near-eye datasets are collected by positioning the EC close to the subject's eye, typically within a range of 5 to 50 centimetres [28], [39]. This proximity allows ECs to focus solely on generating events related to the eye, thereby minimizing noise from occlusions and other objects in the environment. Figure 4 illustrates the data collection setup for near-eye event-based data collection.

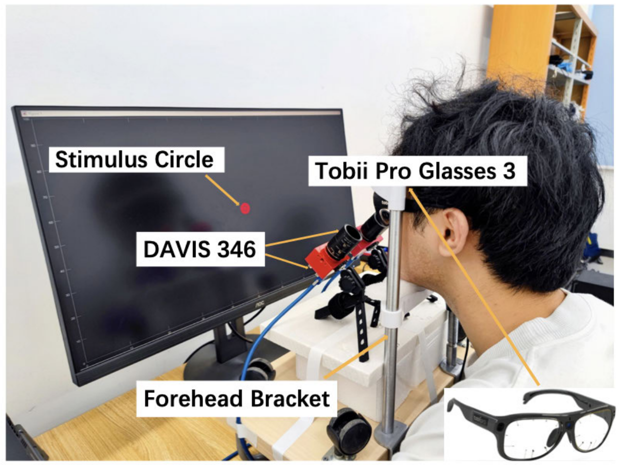

FIGURE 5. **An example of a near-eye data collection setup with 2 DAVIS cameras placed in direct contact with the eyes and the screen at 33cm from the subject [28].**

Near-eye data collection can be done either with one camera to collect data from only one eye or with two cameras collecting data from each eye simultaneously. The reasoning behind the former approach is the fact that during some eye movements such as blinks, both eyes move at the same time removing the need to record both eyes [39], [40]. Other researchers have approached this task by placing two separate cameras for each eye to record more information for a particular task [28]. It is worth noting that the choice of number of cameras to use depends on several factors including the objective of the study and the specific requirements of the tasks. For instance, If the study aims to understand basic eye movement or the gaze direction relative to a simple task or a static environment, recording one eye might suffice whereas a task involving estimating the difference in image location of an object seen by the left and right eyes, require data from both eyes [26]. An example of an image frame with a corresponding event accumulated visualisation is given in Figure 6.

1) EVENTS-BASED DATASETS

Over the decades, several near-eye datasets collected for studies in eye motion analysis with ECs have emerged. Some datasets which have been utilized in research are depicted in

TABLE 1. Event-based datasets available for eye motion analysis tasks. Size is reported, where available, as the number of subjects / number of samples.

Dataset	Gaze Estimation	Pupil Tracking	Blink Detection	Eye Tracking	Availability	Size	Annotation
EV-Eye [28]	✓	✓	–	✓	✓	48 / 92	✓
Neuromorphic Helen [20]	–	–	✓	✓	–	– / 330	–
Lenz *et al*. [29]	–	–	✓	–	✓	7 / 50	✓
Angelopous *et al*. [41]	✓	–	–	✓	✓	24 / 48	✓
Kang *et al*. [42]	–	✓	–	–	–	–	–
SEET-3ET [43]	–	✓	–	✓	✓	– / 66	✓
Ini-30 [44]	–	–	–	✓	✓	22 / –	✓
Gaze-FELL [40]	✓	–	-	–	✓	– / 5	–
3ET+ (AIS 2024) [45]	–	–	–	✓	✓	13 / 78	✓

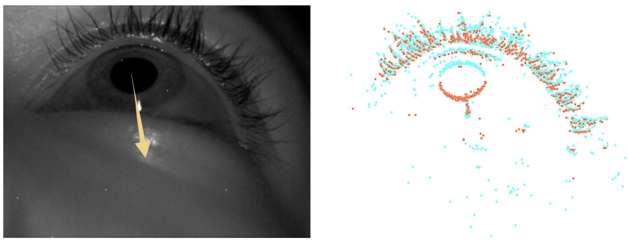

FIGURE 6. An example of a near-eye dataset illustrating a frame and its corresponding event visualization. As Events are triggered by the movement of the pupil, The events on the right hand side are generated by the eye a few milliseconds after the frame shown on the left, during a downward saccadic eye movement. The bottom part of the frame generates negative (red) events, while the top part generates positive (blue) ones. This is due to a negative contrast change produced by the pupil in its direction of motion. Image source: [41].

Table 1, accompanied by indications of the task they have been applied to. These datasets comprise of both pure EC data collected by real ECs as well as synthetically generated data. As noted, certain datasets are specialized for specific tasks, while others have not been made publicly accessible.

Key insights reveal that the EV-Eye dataset [28] stands out as the most versatile, supporting multiple tasks such as gaze estimation, pupil tracking, and eye tracking. It is publicly available and well-annotated, making it a valuable resource for a range of research endeavors. In contrast, datasets like Neuromorphic Helen [20] and Lenz et al. [29] are more specialized, primarily focusing on blink detection and eye tracking. These datasets vary in terms of availability and annotation, with Lenz et al. being both accessible and annotated, while Neuromorphic Helen lacks detailed annotation and availability. Another notable dataset is SEET-3ET [43] , which supports both pupil and eye tracking, offering good availability and annotations. Datasets like Gaze-FELL [40] are more limited in scope, focusing primarily on gaze estimation, but still contribute to specialized research applications.

Overall, the table illustrates the diversity of available datasets, with EV-Eye [28] emerging as the most comprehensive in terms of task coverage. Meanwhile, other datasets offer more targeted data for specific tasks such as blink detection or eye tracking. Researchers should select datasets based on their specific needs, considering factors like availability and annotation quality to ensure the suitability of the data for their studies.

B. REMOTE-EYE

Remote eye datasets are collected by placing the event camera at a distance from the subject's eye, typically ranging from 50 centimetres to several meters. This setup allows for eye tracking without the need for head-mounted devices, making it less intrusive and enabling more natural behaviour during data collection. Remote eye systems are well-suited for studies involving interactions with screens or in environments where maintaining a natural user experience is crucial. However, the increased distance can introduce more noise and reduce the precision of the captured eye movement data, especially in dynamic or complex scenes. The choice of camera placement and distance depends on the study's objectives. For instance, tasks that involve tracking gaze behaviour across a screen or monitoring eye movements in a natural setting might only require a single camera placed at a moderate distance, while more detailed tasks might need multiple cameras positioned strategically to capture comprehensive data. Figure 4 illustrates the data collection setup for remote-eye event-based data collection

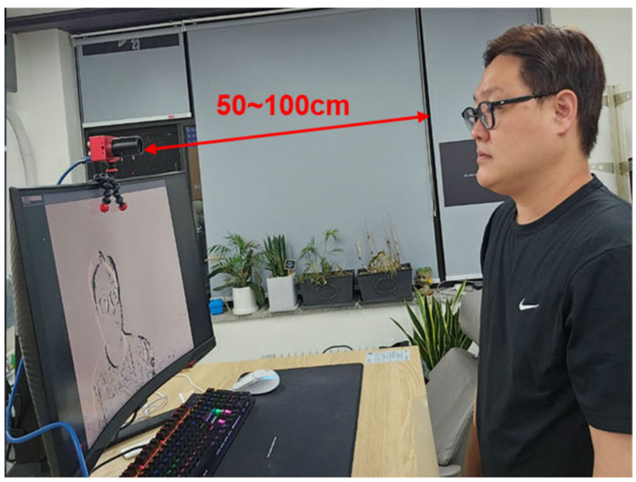

FIGURE 7. An example of a remote-eye data collection setup with a single DAVIS camera placed within a range of 50-100cm from the subject [27].

Remote-eye datasets gather information from the entire facial region. Consequently, in these datasets, both eyes are visible within the frame, unlike near-eye datasets, which typically only contain data on one eye per frame or event stream. To perform eye motion analysis on such data, a method for detecting eye regions, such as using an ROI filter, is employed to extract events from the eye region for further analysis. As of the time of writing, only a few studies have utilized remote-eye datasets for tasks involving eye motion analysis, and none of these datasets have been made publicly available [19], [46]. An example of an RGB remote-eye dataset with a corresponding synthetic event accumulated frame is illustrated in Figure 8.

FIGURE 8. An example of a remote-eye dataset illustrating an original RGB frame from the dataset and a corresponding event visualization reconstructed from simulated events. Image source: [29].

1) EVENT-BASED DATASETS

Over the past decades, a few remote event-based datasets have emerged. Some of these datasets include Faces in Event Streams (FES) dataset [51], Color event dataset (CED) [52] etc. However, most of these datasets are not directly suited for tasks involving eye motion analysis. To the best of the authors' knowledge, there are currently no event-based datasets for remote eye-gaze estimation.

However, there are few studies which utilized event based remote eye-datasets with local acquisitions. Ryan et al. [20] utilised synthetic events simulated from the publicly available Helen Dataset for eye tracking. Ryan et al. [46] further evaluated worlds first ever event-based driver monitoring system with locally acquired event dataset. Alessi et al. [19] collected a purely event-based in-house dataset for a driver scenario and subsequently applied filter fitting to the pupil for eye-tracking. However, these datasets have not been made publicly available. To address this limitation, the event camera community has developed event simulators that generate synthetic event data streams. This topic is discussed in the following sub-section.

C. EVENT SIMULATORS

The introduction of event simulators has enabled the translation of RGB datasets into event-based counterparts. Event simulators are designed to replicate the way ECs operate in capturing visual information. Hence, they simulate the generation of events by converting sequences of frames into a stream of events, usually by detecting changes in pixel intensity between subsequent frames. In the case of static RGB images, some studies have applied techniques to simulate motion (using 6-degree-of-freedom) from the RGB images before event simulation [20], [53]. Early event simulators exploited the sparse nature of events to generate simulations. However, this simulation was inaccurate as it relied on the difference between two successive frames to produce images similar to those produced by ECs and thus leading to every event having the same timestamp [54].

Mueggler et al. [55] developed a simulator by rendering thousand images along a specified trajectory of the sensor's movement within a 3D scene. Each pixel's activity is tracked, including the time of the last event triggered at that location, enabling the generation of event streams, intensity frames, and depth maps. Time interpolation of the rendered image is employed to determine brightness changes between consecutive images, effectively providing continuous timestamps for event generation.

A similar approach is presented in [56] where images are rendered at a high frame rate and low resolution to produce events whenever an intensity threshold is met. However, this fixed frame rate methods can still lead to inaccuracies in situations where the brightness change fluctuates at a pace exceeding the rendering frame rate arbitrarily selected for the simulation. A more suitable approach [57] is the introduction of a significant departure: rather than fixing an arbitrary rendering frame rate and uniformly sampling frames at this rate. This adaptive approach dynamically adjusts the sampling rate based on the anticipated dynamics of the visual signal.

Other studies have proposed different methods to event simulation including a region based model which computes scores for each image region and conducts rendering solely for these regions with a Region of Interest (ROI) model [58]. V2e (video to events) simulator [59], addresses common issues like motion blur in dynamic scenes, achieved through pixel-level Gaussian event threshold variability and intensity-dependent bandwidth, reflecting the camera's changing responsiveness under different lighting conditions; and light intensity-dependent noise, which adds realistic sensor noise based on environmental lighting.

1) POTENTIAL RGB DATASETS

To address the challenge of unavailable remote-eye datasets, researchers have explored converting RGB datasets into event-based formats using simulators. This approach leverages existing video datasets to create event-based training data, which is crucial for developing EC-based eye-tracking systems capable of real-time analysis in varying lighting conditions. The energy-efficient nature of ECs further enhances the benefits of such systems, including improved power consumption and battery life.

Non-event-based RGB datasets are valuable for generating synthetic event data. These datasets, captured under diverse conditions, provide essential benchmarks for training and evaluating vision algorithms. By integrating insights from both event-based simulations and RGB datasets, researchers can develop robust and versatile vision systems suitable for a wide range of real-world scenarios. Table 2 lists several RGB datasets used for remote-eye motion analysis tasks, such as gaze estimation, pupil tracking, and blink

TABLE 2. Non event-based datasets available for remote-eye motion analysis tasks.

Dataset	Modality	Gaze Estimation	Pupil Tracking	Blink Detection	Eye Tracking
mEBAL [47]	Multimodal	–	–	✓	–
MpiiGaze [48]	RGB	✓	–	–	
GazeCapture [49]	RGB	✓	–	–	✓
ETH-XGaze [50]	RGB	✓	–	–	–

detection. Each dataset is categorized by modality and the specific tasks it supports, illustrating the diverse range of available datasets and their applications. By utilizing both event-based simulations and RGB datasets, researchers can develop comprehensive event-vision systems that capitalize on the strengths of both traditional RGB data.

D. NOISE REMOVAL

The functionality of ECs is contingent on pixel intensity changes resulting from motion. This reliance on motion-induced intensity changes can lead to the introduction of significant noise into the event data [62]. There are several key sources of this noise ranging from leakage and thermal noise from transistor switching to variations in ambient light resulting in random events been generated [8]. As pointed out in the literature [63], under sub-optimal lighting conditions, noise can easily exceed the sensor's available bandwidth, resulting in missed events, inaccurate event timestamps, decreased contrast sensitivity, and ultimately, poor application performance. The impact of noise is particularly pronounced in low-light conditions, where minor fluctuations in intensity generate numerous spurious events outside the region of interest that overwhelm the useful events in the scene. This poses a substantial challenge for applications such as eye tracking, where precise and accurate data is crucial.

Various denoising techniques have been developed to filter out noise events while preserving the signal from actual scene changes [64], [65]. However, only a few of these have incorporated denoising for tasks involving eye motion analysis. Table 3 summarises some of noise removal methods explored in event based vision systems. Li et al. [60] designed a Latent Denoising Neural Network that effectively cleans noise from feature maps. This network models noisy latent features as Gaussian distributions and uses a diffusion technique to remove the noise in a self-supervised manner. The denoised feature maps are then used to train a student network, ensuring accurate gaze estimation. This approach resulted in a gaze estimation performance with a Mean Angular Error of less than 2°, outperforming state-of-the-art methods in event-based gaze estimation.

Carredu [61] introduced a noise reduction filter to mitigate noise from eyelid movement and light reflection in event-based data. Their method employs an entropy-based filtering approach to improve the clarity of regions of interest. By calculating the density of active pixels within a specified area and applying an entropy calculation ($H(p1, p0) = -p1 \cdot \log(p1) - p0 \cdot \log(p0)$), the filter effectively identifies regions with significant activity, minimizing randomness. Subsequently, they evaluated the effects of noise reduction by comparing pupil localization with groundtruth utilizing a 50-pixel radius around estimations to match groundtruth. Additionally, the proposed method showed improved pupil detection accuracy in 8 out of 12 recordings which was tested. However, the reliance on the virtual circle method for evaluation, may obscure finer inaccuracies in pupil detection. There is a need for more precise synchronization techniques and robust evaluation metrics to address detector and tracker frequency discrepancies.

Similarly, E-gaze [22] incorporates a box filter applied to calculate the sum of events near specific coordinates. A high sum indicates the presence of movements and strong features, while sums below a specified threshold are flagged as potential noise events and removed. This method effectively reduces spatio-temporal noise but has several gaps. These include the static threshold value, lack of consideration for filter size and shape, edge cases, temporal consistency, and the impact on subsequent processing stages. An illustration of this approach is shown in Figure 9.

Some studies such as EV-eye [28], incorporates noise removal in their post processing pipeline to remove noise such as glint from the pupil area via morphological closings [66].

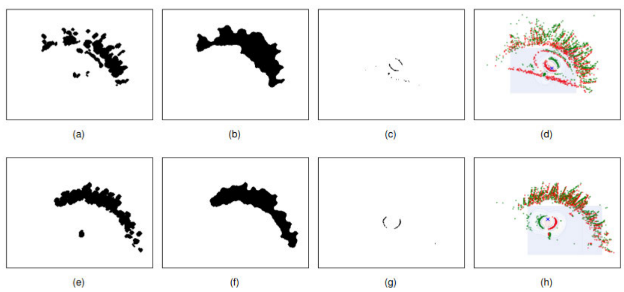

FIGURE 9. An example of noise removal during event processing for eye tracking, (c) and (g) represents the pupil mask before and after noise removal. Image source: [22].

Effective noise removal is essential for enhancing the robustness of eye tracking and gaze estimation algorithms. As event-based vision technology is still evolving, there is currently no standard method to benchmark denoising performance for eye motion analysis tasks due to the unknown and variable nature of noise distribution in eye regions captured by ECs. This variability is dependent on environmental factors, the specific scene being captured, and the characteristics of the sensor itself [67]. The primary method adopted for evaluating noise removal involves reconstructing event streams into frames, allowing for visual inspection of the effects of noise removal (as shown in

TABLE 3. Summary of noise removal methods in event-based vision systems.

Method	Key Features	Results
Latent Denoising Neural Network [60]	Self-supervised learning, feature map denoising.	Achieved Mean Angular Error of $< 2°$, outperforming other methods.
Entropy-Based Filtering [61]	Uses entropy calculation $(H(p1, p0))$, evaluates pupil localization accuracy.	Improved pupil detection accuracy in 8 out of 12 recordings.
Box Filter (E-gaze) [22]	Static threshold, detects movements, removes low-sum noise.	Reduces spatio-temporal noise but has limitations in threshold and consistency.
Morphological Closings (EV-eye) [28]	Focuses on removing glint and other noise artifacts.	Applied in post-processing pipeline, specific performance details not provided.

Figure 10). Other evaluation methods include metrics such as the Peak Signal-to-Noise Ratio (PSNR) and more recently established metric, the event structural ratio (ESR) [68].

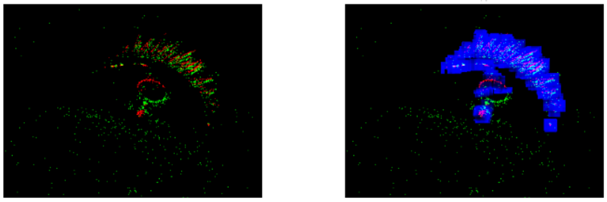

FIGURE 10. A frame-based evaluation of noise removal is conducted, with the initial image depicting an event frame, while the blue mask illustrates the effectiveness of the entropy filter in successfully identifying noise, such as eyelashes and glint(blue mask). Image source: [61].

Overall, the importance of noise removal in event-based vision cannot be understated, as it significantly contributes to the accuracy and reliability of the resulting task. Further research and development are needed to establish standardized noise removal procedures and to evaluate their effectiveness across different applications. Understanding the characteristics of noise in ECs especially for tasks in eye motion analysis is crucial to improve their robustness and reliability.

IV. EVENT REPRESENTATIONS

The data produced by Event Cameras (ECs) poses a unique challenge for traditional machine learning and deep learning frameworks due to its unique structure. This necessitates the development of novel algorithms tailored to handle such data or the adaptation of existing models into compatible formats. This challenge is particularly significant in the field of eye motion analysis. Conventionally, event representation through frames is straightforward and compatible with Convolutional Neural Networks (CNNs). However, due to the activation of ECs by pixel intensity, certain ocular regions might not be captured in the frames if no movement occurs in those areas, unlike traditional imaging where all components of the eye are consistently visible within the frame. This issue is illustrated in Figure 3.

An alternative and potentially more effective method involves extracting direct eye features from the event streams. Numerous studies have explored the identification of eye regions by leveraging specific eye movement characteristics, such as blinks [29], while others have developed techniques for directly filtering eye regions from event streams [30]. Various methodologies for event representation include 2d-frames, graph-based, voxel-grid, binary-spikes, and time-surface representations [8], [69], [70].

In eye motion analysis, the selection of an appropriate event representation is crucial for the accuracy of specific tasks, such as pupil tracking. A frame-based representation might fail to capture the pupil in certain frames, as depicted in Figure 3. Therefore, a more sophisticated approach to event representation is necessary.

Table 4 summarizes various methodologies used in eye motion analysis, highlighting the diversity of neural network models and representation techniques employed. For eye and gaze tracking tasks, both frame-based and voxel grid representations are commonly utilized, with frames being the predominant choice across multiple studies. Specific methods such as landmark detection, Region of Interest (ROI) detection, and neural network architectures like CNNs, U-Nets, and ConvLSTMs are applied. Blink detection primarily uses frame-based representations as well, although one instance employs voxel grids, and another uses event-based data. Pupil tracking consistently employs frame-based representations, utilizing various landmark detection techniques. This range of approaches reflects ongoing efforts to optimize the accuracy and efficiency of eye motion analysis technologies.

V. APPLICATIONS

This section provides a comprehensive review of present literature in tasks involving eye motion analysis. This section will classify related work primarily on the methodology used to solve a specific task. Our analysis focuses on 5 main tasks: Eye tracking, Gaze Estimation, Blink Estimation, Pupil Segmentation / Tracking, Saccades and Microsaccade Analysis.

A. EYE TRACKING

Eye tracking is one of the significant tasks in eye motion analysis. This task involves assessing where a person is looking (the gaze point) or measuring the movement of the eye relative to the head. Previous studies on eye tracking relied on the use of traditional cameras to detect eyes and other eye features for tracking. These methods have been efficient for downstream tasks such as activity recognition and attention estimation. However, despite considerable effort to solve this task using machine learning, problems

TABLE 4. Classification of methodologies used for eye motion analysis tasks with event cameras.

	Application	Neural Network	Methodology	Representation
Angelopous *et al*. [41]	Eye tracking	–	Landmark Detection	Frames
Ryan *et al*. [20]	Eye tracking	CNN	GR-YOLO	Voxel Grids
Zhao *et al*. [28]	Eye tracking	U-Net	Landmark detection(pupil)	Frames and Events
Li *et al*. [18]	Eye tracking	U-Net	ROI detection	Frames
Chen *et al*. [43]	Eye tracking	CB-ConvLSTM	Landmark detection	Voxel Grids
Bonnazi *et al*. [44]	Eye Tracking	SNN	Spike surface	Frames
Yang *et al*. [37]	Eye Tracking	U-Net	Landmark Detection	Frames
Ryan *et al*. [46]	Gaze tracking	Multi-task Neural Network	ROI detection	Time-surface
Zhao *et al*. [28]	Gaze Tracking	–	Polynomial Regression	Frames
Banerjee *et al*. [28]	Gaze Tracking	ResNet-50 + CNN	Landmark Detection	Frames
Feng *et al*. [30]	Gaze Tracking	Depthwise CNN (DWSConv)	ROI detection	Frames
Lenz *et al*. [71]	Blink detection	–	Landmark detection (face)	Frames
Lenz *et al*. [29]	Blink detection	–	Landmark detection (face)	Frames
Chen *et al*. [72]	Blink detection	SNN	Landmark detection	Frames
Chen *et al*. [73]	Blink detection	–	Landmark detection	events
Ryan *et al*. [20]	Blink detection	–	Statistical approach	Voxel grids
Kagemoto *et al*. [74]	Pupil tracking	–	Landmark detection	Frames
Kang *et al*. [74]	Pupil tracking	RetinaFace	Landmark detection	Frames
Kang *et al*. [27]	Pupil tracking	Feature Pyramid Network	Landmark detection	Frames

such as the low frequency and bandwidth (hertz) of these methods hinder its use in fine-grained applications such as the diagnosis of neurodegenerative disorders and high frequency applications such as gaming. To achieve eye tracking beyond Kilohertz (KHz) frequency, a high camera bandwidth is required.

Due to the distinct nature of data generated by ECs, conventional machine learning algorithms can not be used directly with the data they provide. Hence, research in event-based vision primarily depends on statistical approaches and the use of representations which makes it feasible for integration with deep learning algorithms. For each of these approaches, there are sub-classifications based on how data is used: by combining events and frames; by the use of only events; and by converting events into representations such as the ones highlighted in Section IV.

1) HYBRID STATISTICAL METHODS

In the domain of eye motion analysis, several studies relied on fusing event streams with RGB frames for enhanced accuracy. This common approach is a result of event-based frames lacking sufficient numbers of events to enable reliable representation of eye features [22], [43]. One of the earliest studies in the use of ECs for eye tracking (EBV-EYE) [41], demonstrates the use of ECs for high frequency eye tracking using this hybrid approach. The authors proposed a 2D parametric model representing the pupil from events and feed this to a parabolic model which maps events to gaze points. However, this approach is not robust to camera slippage relative to the face, requiring periodic re-calibration. Furthermore, the dataset utilised is not diverse and the proposed approach does not incorporate any mechanism to mitigate the effects of noise leading to an impact in robustness.

Similarly, Feng et al. [30] proposed an Auto ROI algorithm to continuously predict ROIs of near-eye images for eye tracking. This method leverages software-emulated events and temporal feedback to predict regions of interest (ROIs) in real-time. The key components include a lightweight eye segmentation neural network co-trained with ROI prediction, and a feedback-driven mechanism that continuously optimizes the gaze tracking process. By processing only the most relevant pixels within the predicted ROIs instead of full-resolution frames, the algorithm achieves significant efficiency gains, reducing computational overhead by 68-82% while maintaining high accuracy ($0.1° - 0.5°$ gaze error) at 30 Hz on mobile processors. However, a limitation of this method lies in the co-training of the segmentation network and ROI prediction which could potentially introduce additional complexity compared to training each component separately.

A more recent study (EV-EYE) performed eye tracking utilising both RGB frames and event streams by localizing the centroids of the pupil area and applying a polynomial regression to identify the corresponding Point of Gazes [28]. However, the matching-based method constrains the pupil shape until the next frame arrives, which may not accurately represent real-world conditions where the pupil can change shape continuously.

While the integration of frame and event data in eye tracking presents promising advancements, challenges related to dataset quality, real-time processing, noise sensitivity, and computational demands remain significant hurdles that need to be addressed for broader application and effectiveness. Combining frame and event data introduces complexity in processing. The algorithms must effectively synchronize and integrate the two types of data, which can be challenging. Any misalignment or inefficiency in this process can degrade the overall tracking performance and accuracy.

2) CNNs

The integration of event data with machine learning models such as Convolutional Neural Networks (CNNs), has shown potential especially for eye tracking and has been widely utilised. Most CNN methods for event-based eye tracking, rely on first reconstructing the event data into frames. This is because CNNs are primarily designed to operate on dense, grid-like image data, rather than the sparse, asynchronous event data produced by ECs. The process of reconstructing event data into frames can be computationally expensive and may result in the loss of temporal information present in the original event stream. Li et al. [18] proposed a 3-channel frame representation of events and utilize a low latency segmentation CNN model to predict pupil events from these frames. Eventually, an event-based RoI mechanism is employed to track the pupil with minimal CNN inferences by focusing on a specific region after the initial detection. While their approach significantly reduces latency, the accuracy of pupil detection in the presence of occlusions has not been reported. In a subsequent study [22], the authors propose a two-dimensional kernel density estimation with a donut kernel, followed by elliptical least square fitting to extract pupil features. To optimize latency, previously extracted pupil features define a region of interest for the current event set, streamlining the process and ensuring efficiency. Similarly, Chen et al. [43] convert events into a time bin representation and proposed a Convolutional Long Short-Term Memory (CB-ConvLSTM) model architecture to extract spatiotemporal features for pupil tracking from the event stream, outperforming conventional CNNs. Ryan et al. [20] proposed a modification to YoloV3 by incorporating a GRU architecture into the network to perform eye tracking on event streams represented as voxel grids. Figure 11 illustrates the results of this approach to eye tracking on a frame-based representation of events. Yang et al. [37] adopt a frame interpolation method that exploits the high temporal resolution of event-streams to generate high frame-rate videos so that the detailed profile of the eye movement can be presented in a slow-motion mode. Subsequently, a U-Net based network for pupil segmentation is used to extract pupil region from video frames.

Frame reconstruction methods essentially enable us to apply well-known computer vision algorithms. While these methods show promise in leveraging CNNs for event-based eye tracking, challenges remain in fully exploiting the sparse, asynchronous nature of event data while maintaining high accuracy and low latency. Continued research is needed to develop CNN architectures and event representations tailored to the unique properties of eye movement

Recently, the AIS 2024 Challenge on Event-Based Eye Tracking has spurred a significant increase in related research. The task was to predict the pupil center spatial coordinates from the raw event stream input. The participating teams proposed a variety of methods, including stateful models like GRUs, ConvLSTMs, and Mamba [75], as well as techniques for spatial-temporal processing of the event data. Different approaches were used for converting the raw event data into representations suitable for deep learning models, such as binary map representations and point-based networks. For readers seeking a more comprehensive overview of research methodologies employed by challenge participants, there is a comprehensive review specifically on the challenge by Wang et al. [45].

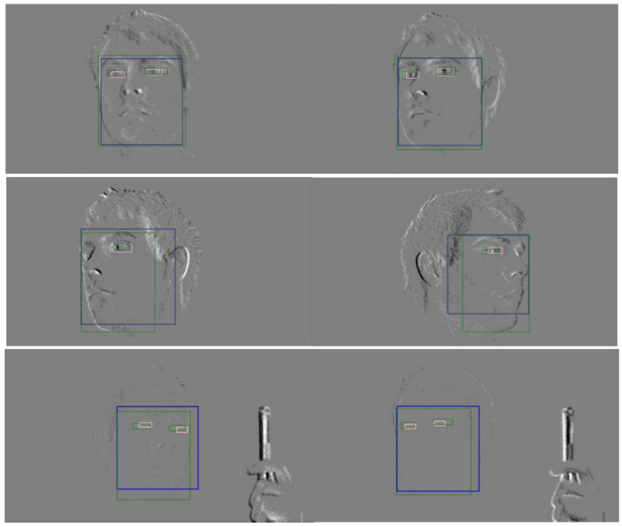

FIGURE 11. Illustration of eye tracking results with green bounding boxes representing groundtruth and red boxes indicating predictions. Image source: [20].

3) PURELY EVENT-BASED STATISTICAL METHODS

In contrast to Hybrid and CNN-based approaches, a few works have demonstrated eye tracking directly from event streams without the need to convert them into different representations. The nature of the data from ECs allows Spiking Neural Networks (SNNs) to utilize their memory characteristics effectively, leveraging membrane potential to extract temporal features. This capability enhances the tracking accuracy while maintaining lower computational costs compared to conventional Artificial Neural Networks (ANNs). Additionally ECs provide asynchronous incremental outputs, which align well with the operational characteristics of SNNs. This means that SNNs can process incoming data in real-time without needing to wait for a complete frame, enhancing the system's responsiveness and stability. A few works have utilised these architectures for eye tracking.

One study [44] trains directly a Spiking Neural Network (SNN) regression model based on Integrate And Fire (IAF) neurons, named "Retina," using a continuous regression output CNN with a non-spiking temporal 1D filter slid across the output spiking layer. The Retina model exhibits reduced precision when trained without the continuous resetting of neuron states. However, the frequent resetting of these states could lead to potential interruptions in continuous tracking on a neuromorphic chip. Investigating methods to address this issue and enhance model performance while considering these additional hardware limitations presents a valuable opportunity for future research.

Similarly, the work by Jiang et al. [39] process event data by encoding their information in spike trains. Event points are divided into subsets to introduce temporality and represented in a tensor format for convolutional layers. The SEW ResNet backbone network [76] extracts spatial and temporal features, while a tracker uses Non-Spiking PLIF [77] neurons to accumulate membrane potential and provide precise outputs. This approach demonstrates superior re-tracking performance and energy efficiency compared to traditional ANNs, offering a robust solution for real-time eye tracking with reduced computational costs. This approach may struggle to provide accurate tracking results in the presence of noise since the tensor representation used can be computationally intensive.

Other studies utilised events directly [38], with a novel illumination scheme called Coded Differential Lighting to enhance corneal glint detection for event-based eye tracking. By using pairs of pulsed light sources that turn on and off in a binary-encoded, differential manner, the system amplifies the specular reflections (glints) on the cornea while suppressing events from the rest of the diffuse scene. Combined with frequency-based filtering to remove unwanted noise, this approach enables a purely event-based corneal glint detection and tracking algorithm that can operate at high sampling rates (up to 1 kHz) on standard hardware, while consuming very low power (about 35 mW). While this method operates effectively at high sampling rates and low power, it is contingent on specific lighting conditions. Variability in ambient lighting can affect the reliability of glint detection, potentially leading to inaccuracies in tracking.

B. GAZE ESTIMATION WITH TRACKING

Gaze estimation and tracking has gained popularity in the past decades due to its real-world applicability in areas such as Virtual Reality [78], Driver monitoring [79] and Diagnostics [80]. Gaze estimation, beyond eye tracking, focuses on determining where a person is looking, either on a screen or in the real environment [81]. Gaze estimation techniques can be broadly categorized into two main types: model-based and appearance-based methods.

1) MODEL-BASED

These methods rely on a geometric or physiological model of the eye. They often use features such as the pupil center and corneal reflections from an infrared light source to estimate the gaze direction [82]. Early studies have utilised model-based approaches for gaze estimation [83]. However, only a few studies have utilised these approaches with event-based vision. This is because these methods are typically sensitive to fluctuating ambient light conditions, necessitating the use of high-resolution images for accurate tracking.

Building on their proposed method for eye tracking, Angelopoulos et al. [41] fits parameters for an eye model from their eye tracking pipeline using polynomial regression to map the pupil center coordinates (x_c, y_c) to the screen coordinates (x_s, y_s). This approach involves two polynomial functions, one for each screen coordinate, which are trained using input-output pairs collected during calibration. The simplicity of these polynomial functions ensures fast and efficient computation, making the system suitable for high update rates.

Expanding on these foundational methods, Feng et al. [30] proposed a combination of ROI prediction and a segmentation model that produces an event maps, mimicking ECs function to facilitate current frame ROI prediction. To enhance robustness, the authors propose a feedback mechanism utilising two key cues from a previous frame to improve ROI prediction accuracy. This mechanism is particularly effective in scenarios of minimal eye movement, employing a strategy that adjusts tracking effort proportionally to observed activity levels. They then employ a U-NET-like segmentation network to segment the event maps to perform eye segmentation and subsequently gaze estimation. However, their approach demonstrates sensitivity in handling significant vertical eye movements, as evidenced by the slight drift in the predicted ROIs during such movements.

In addition to these advancements, Li et al. [22] proposed a method for gaze estimation leveraging a dual-phase, coarse-to-fine strategy integrating event-driven sensing and frame data. Initially, the dataset is partitioned into sub-regions, and local expert networks are trained using a transformer architecture to achieve high local accuracy. A latent-denoising distillation method then transfers knowledge from these local experts to a global student network, mitigating noise impact. The framework models gaze direction through differential estimation relative to anchor states, using convolutional layers and pointnet for event frame tokenization, and vision transformers for correlation modeling. This approach enhances gaze estimation by combining detailed spatio-temporal features from both modalities, ensuring robust and accurate predictions.

More recently, Menasti et al. [84] employ a movement prediction strategy that groups events into batches, tracks the event count, and calculates a running mean to determine when to process batches, optimizing resource usage. A pupil detector operating at 250 Hz, translates the event stream into event frames while using 2D convolution with directional templates to identify the pupil location and movement direction based on the edge patterns generated by the event camera. The initial pupil centre estimation is refined using two fitting strategies: the RANSAC algorithm iteratively selects random subsets of events to fit an ellipse, while the continuous update method uses events near the edge of the previous ellipse to continuously update the centre of the ellipse model.

2) APPEARANCE-BASED

they do not assume any specific eye model. Instead, they rely on the appearance of the eye region, sometimes utilizing deep learning algorithms to directly map features extracted from eye images to gaze points [85]. Banerjee et al. [86] leveraged such models for gaze estimation with ECs. They proposed a unique coding technique to represent event data into six-channel images which are then used as input for a lightweight 2D Convolutional Neural Network (CNN) designed for gaze detection. The primary task of this CNN is to predict the location of a red circle displayed on a screen, functioning as a regression challenge based solely on the encoded images.

In a subsequent study [40], they propose a dual pathway architecture which processes input sequentially with pre-trained encoders for feature extraction. This setup enables the prediction of gaze direction by pinpointing centroids in successive frames, facilitating precise, person-independent gaze tracking. The extracted features are combined to predict gaze centroids, with the final layer resulting in a gaze vector compared against a target vector for training. However this approach fails short in accurately determining gaze direction, attributed to rapid viewpoint changes within short intervals.

Ryan et al. [46] propose the first remote gaze tracking by developing a multi-task CNN that utilizes shared lower layer feature representations to simultaneously determine head pose and gaze direction. This network is divided into two distinct parts: the first part employs a LeakyIntegrator for initialization and sets up an event window iterator, which operates based on fixed-size, fixed-duration, or region-of-interest (ROI) based events. This part processes event windows to generate a time surface grid which is then integrated using the LeakyIntegrator. The processed input is subjected to a face and eye detection model, where non-maximum suppression is used to identify and localize faces and eyes. Subsequently, the All-In-One (AIO) network, takes the identified face and eye regions as input for gaze tracking and detection of occlusions. An example of this approach to gaze tracking is illustrated in Figure 12.

Zhao et al. [28] proposed a method to obtain pupil centres and apply a polynomial regression on this pupil centres to estimate point of gazes on a screen by mapping the pupil centres in image domain to the PoGs for gaze tracking. Subsequently, they measure the performance of this method by evaluating against an eye tracker calibration tool provided by Tobii. Li et al. [22] propose a gaze estimation architecture consisting of a recurrent neural network (RNN) and fully connected layers. Input pre-processing involves encoding the pupil ellipse parameters into a 21-element vector. The RNN, specifically an LSTM layer, processes a sequence of 10 vectors to capture motion and eye movement direction. Nonetheless, the pipeline proposed requires calibration for each usage as well as a measurement of distance from the screen for evaluation limiting usage for robust settings.

C. BLINK ESTIMATION

Eye blink parameters are sensitive indicators of various cognitive states and physiological conditions [87]. Slowed eyelid movements, prolonged blink duration, and frequent episodes of prolonged eye closure has been associated specifically with drowsiness [88]. Additionally, blink rate has

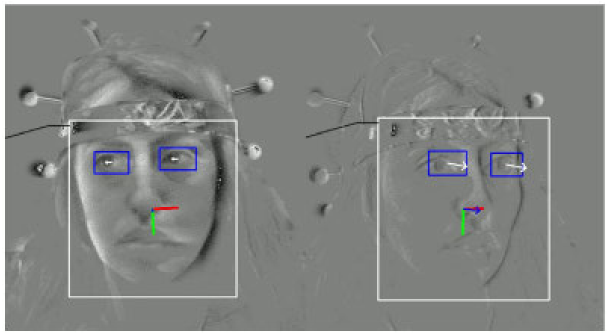

FIGURE 12. Illustration of gaze estimation with white and blue box indicating face and eye bounding box and arrows indicating eye gaze. Image source: [46].

been found to vary with attentional focus. For instance, while reading, adults may blink as few as 3 times per minute, but this rate can increase to as much as 30 times per minute during conversations [89]. During internally directed cognition tasks, increased pupil diameter, blink rate, and saccade rate predict susceptibility to visual distraction [90], confirming blink behaviour as a sensitive indicator of attentional focus.

Conventional methods for blink detection have relied on localization of the eyes and further classification of the eye state as open or closed. Some studies have utilised neural networks such as CNNs, RNNs and LSTMs [91], while Other studies have relied on video-based approaches which has shown promising results [92]. However to accurately capture blinks, a high frame rate (fps) camera is required. Blinks occur rapidly, with the eye closing and reopening within a fraction of a second [93]. Low frame rate cameras may fail to detect the complete blink cycle, leading to inaccuracies in blink detection

During a blink, the rapid changes in light intensity caused by the eyelid movement trigger a series of events which contains polarity information indicating the positive and negative changes in pixel intensity. Blinks generate a significant number of events within the eye regions. In particular, the downward movement of the eyelid over the eye typically generates an abnormal surge in the number of positive and negative events causing a reversal in the polarity during the eyelid downward movement as shown in Figure 13. This allows ECs to capture the unique temporal signature of a blink with high temporal resolution, on the order of microseconds, making them well-suited for real-time blink detection and tracking applications.

A few studies have leveraged the unique characteristics of ECs for this task. Lenz et al. [71] proposed the first purely event-based method for blink detection highlighting blinks as having a consistent duration across different lighting conditions and individuals, making them an ideal candidate for developing a universal blink signature. Their approach focuses on converting event streams into a temporal activity profile for each blink, updating this profile with a decay function whenever an ON or OFF event is detected. A canonical blink model distinguishes between ON and

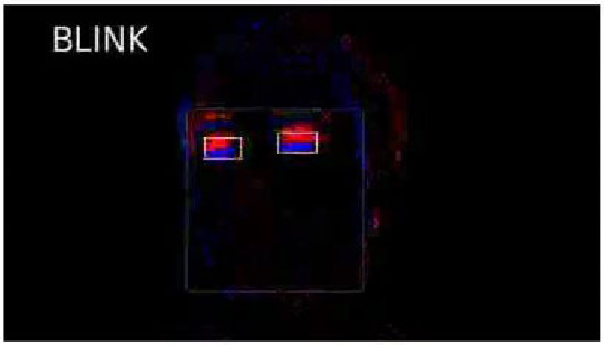

FIGURE 13. Illustration of a blink captured by an EC with polarities indicated in red and blue. Image source: [20].

OFF event activities, effectively capturing the natural activity pattern of a blink, including the minimal activity when the eye is closed and the surge of activity when the eyelid closes and opens. Events in a temporal window are then divided into grids and an activity filter reduces noise by verifying the spatio-temporal proximity of detected events. The blink model is then applied through sparse cross-correlation, comparing recent activity against the blink signature to identify potential blinks.

Lenz et al. [29] further explored this work, leveraging the dynamic stimulus of eye blinks as a temporal signature for face detection. The authors make use of the temporal activity acquired to initialize trackers and prevent drifts. Once tracking is initiated, the face's location is updated with microsecond precision, reflecting the camera's native temporal resolution. The system can track and re-detect faces over a significant range of scales, from 25 cm to 1.5 meters, and is robust to variations in lighting conditions and partial occlusions of the face, such as one eye being covered. The blink detection technique is simple yet effective, allowing for the simultaneous tracking of multiple faces.

A different perspective is offered by Chen et al. [72], who proposed a two-stage filtering process to preprocess data from a driver dataset, automatically recognizing eye and mouth movements. This technique emphasizes processing events closely linked to the eyes and mouth areas, as key indicators of drowsiness. They designed a Spiking Neural Network (SNN) filter to eliminate low-frequency signals from other facial movements and introduced a 3D filter for enhanced analysis. The study also utilizes a sliding buffer space to group events triggered by eye blinking and mouth movements, facilitating more accurate drowsiness detection.

Complementing this study, Chen et al. [73] detects blinks for a biometric authentication system. They proposed a method focused on event density, which is calculated by sliding a window along the time axis and counting the number of events within this window. Filtration and curve smoothing techniques are applied to the event streams to enhance the signal quality, removing noise caused by light fluctuations or sensor imperfections. Features are derived from the event density data, capturing various aspects of the blinking which are then used to describe the unique blinking patterns of individuals. Subsequently, the most relevant features for biometric authentication from eye blinks are extracted for further analysis.

In contrast to the studies above, Ryan et al. [20] adopt a statistical approach to blink detection focused on the inherent polarity output of ECs. They first filter events within a fixed duration of 5 ms time to identify moments when the mean polarity per pixel in the eye region surpasses a certain threshold. Subsequently, they filter the points in time identified in the first step based on the distribution of polarity along the vertical axis of each eye by examining the standard deviation of the event distribution vertically. Finally, within the same 5 ms event window, blinks are analyzed at a more granular level. The distribution of events typically shows a bimodal pattern, representing the eyelid's downward and upward movements. From this analysis, additional features such as blink duration, eyelid closing/opening duration, time the eyelid is closed, and the speed of the blink can be extracted. An example of the results of this approach is illustrated in Figure 14.

D. SEGMENTATION AND PUPIL TRACKING

Pupil tracking, often used interchangeably with eye tracking [39], [94], leverages eye movement data to provide a comprehensive understanding of visual and cognitive processes. This section focuses on pupil tracking as a distinct task from eye tracking, highlighting its unique applications while acknowledging its role as a baseline for eye tracking and related tasks, such as saccade estimation. The rationale behind this approach stems from the outcomes of reviewed studies, where eye tracking typically involves detecting eyes and indicating them with bounding boxes around the entire eye area. In contrast, pupil tracking restricts bounding boxes solely to the pupil areas, as illustrated in Figure 15.

Pupil tracking involves measuring changes in pupil size and reactivity [95]. These changes, including pupil dilation and constriction, can indicate cognitive load, emotion, arousal, and other mental states [96]. Pupil tracking is often applied in conjunction with other tasks such as eye tracking, which tracks the position of the eye and corneal reflections to determine where a person is looking.

Similar to event-based eye tracking, Several studies have used ECs for this task. Kagemoto and Takemura [74] present an approach to pupil tracking based on solving the challenge of sparsity of events leading to challenges in identifying pupil's center. The authors proposed the use of bright and dark pupil effects elicited by two near-infrared illumination sources; one aligned with the camera's optical axis and the other positioned at a distance to alternately illuminate the pupil, thereby facilitating the generation of events for detecting the pupil's center. For pupil center detection, a novel method is outlined that does not require subtracting consecutive frames, as in conventional techniques. Instead, the region around the pupil is predefined, and events within this area are categorised based on their polarity. The centre of

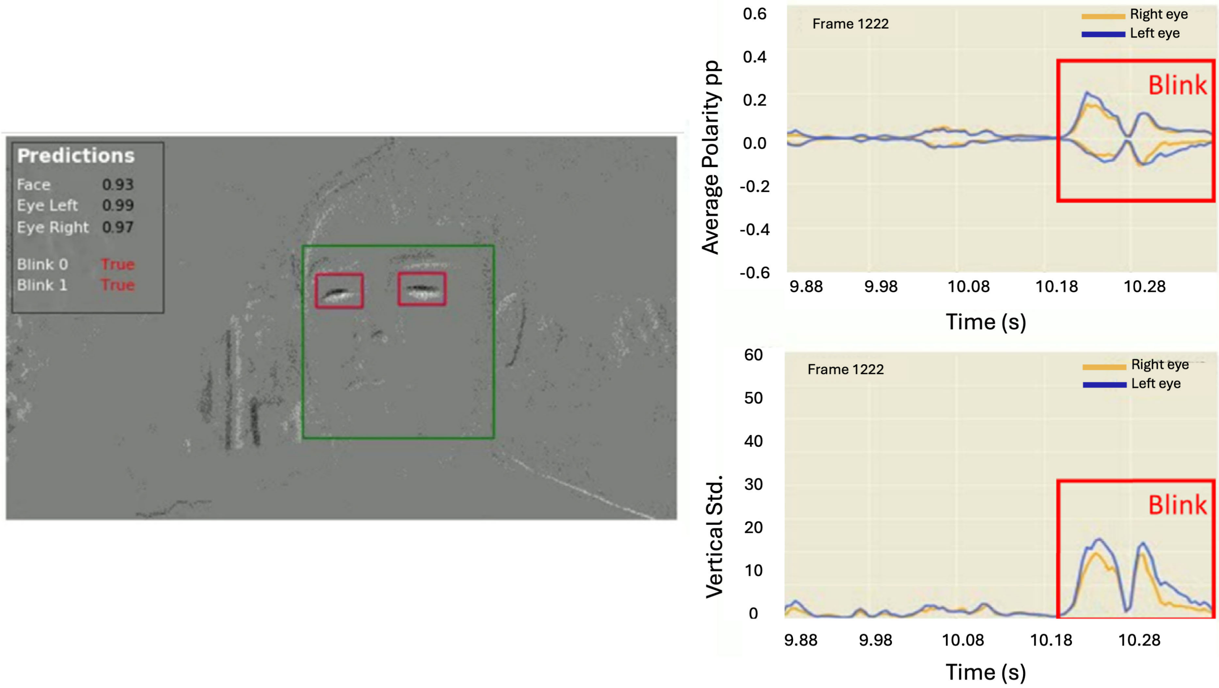

FIGURE 14. Results of the statistical approach to blink detection by Ryan et al. The figure shows the bimodal distribution of event polarities during a blink, highlighting the downward and upward movements of the eyelid. The right image shows real time detection on an event frame while the two graphs on the right indicates the average polarity per pixel over time and vertical standard deviation over time in the top and bottom graphs respectively. Image source: [20].

gravity of events from the dominant polarity is then calculated to determine the pupil's centre.

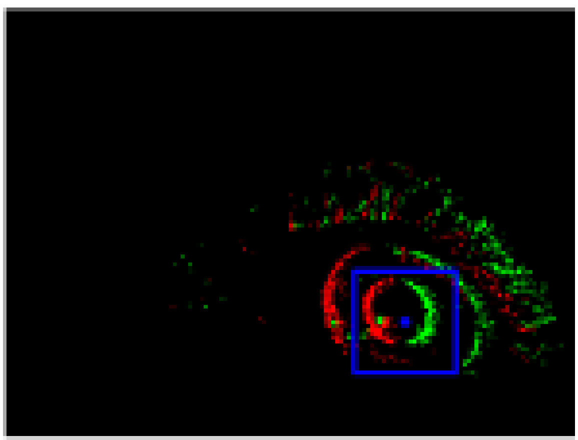

FIGURE 15. Illustration of pupil detection in a RGB event frame with polarity used as color channel. Image source: [61].

Kang and Kang [42] proposed a cross-modal learning pipeline which integrates both original RGB and newly transformed event-like images. Firstly, they generate event-like images through an RGB-to-event image domain translation technique using StyleFlow [97]. The StyleFlow algorithm for domain translation converts RGB images into event-like images that mimic the style of real event camera images while preserving the content's semantic information. Subsequently, they train the pupil localization model with these generated images and the RetinaFace algorithm, which is acclaimed for its joint face detection and keypoint alignment capabilities.

Kang et al. [27], extends their previous work on pupil localization by using cascaded Adaboost classifiers with multi-block local binary patterns (LBPs). The tracking system employs a coarse-to-fine strategy for inferring the pupil centre location, utilising Scale-Invariant Feature Transform (SIFT) features and a pupil-segmentation module for precise tracking. To optimise the performance of the eye-nose detection and pupil tracking, a specialised event camera image database (DB) was created, capturing real-world videos with varying motion levels. The training focused on images from both subtle and large motion categories to ensure the algorithms could handle different motion levels and eye shapes effectively.

Zhang et al. [94] proposed an anti-blink pupil estimation and tracking modules with a temporal feature fusion component and a rotated pupil region detector to accurately estimate and track the pupil through blinks. Additionally, an adaptation mechanism with an occlusion-ratio estimator is designed to exclude unreliable predictions, using a simple linear interpolation strategy for completing the eye movement trace. The Swift-Eye's process involves converting low frame-rate videos into high-speed videos through event-based frame interpolation with Time Lens [98]. This conversion is specifically fine-tuned for eye movement analysis. The framework then extracts and fuses multi-scale spatial features for pupil estimation and tracking, effectively dealing with the challenges posed by involuntary blinking.

E. SACCADES

Saccades are fast movements of both eyes simultaneously, typically observed when a person changes a point of focus of the eye from one object to another, for example during reading [99]. Saccades are some of the quickest eye movements produced by humans, with the potential to reach even higher velocities than blinks. The peak angular speed of the eye during a saccade can reach up to 700°/s in humans, especially during large saccades (covering a visual angle of 25°) [100]. Typically, saccades in response to an unexpected stimulus take around 200 milliseconds to initiate and last for about 20 - 200 milliseconds, with an average duration of 20 - 30 milliseconds during language reading [101]. Research in saccadic eye movement have investigated the mechanism of saccades and how some neurological states can be inferred from this type of eye movement such as concussions, traumatic brain injuries (TBI) [102], [103]. Early studies revealed the connection between saccades and Parkinson's Disease concluding that patients with Parkinson's Disease exhibited small saccade amplitude when compared to normal subjects [104], while more recent studies has highlighted the potential of saccadic eye movements as a reliable and non-invasive biomarker for detecting cognitive impairment in Parkinson's disease (PD) [105], [106], suggesting that saccade measurement could be a valuable tool for assessing cognitive function.

A few studies have leveraged Machine Learning for Saccadic detection. Attre [107] presents an approach to saccade analysis with Machine Learning relying on pupil tracking via Hough Transform algorithm from RGB videos. Tirdad et al. [108] utilise ensemble machine learning approaches to train multiple random classifiers to analyze the non-linear patterns in saccadic eye movement, thus distinguishing between different phases of TBI. Similarly, Lutfullaeva [109] compare random forest with other classification algorithms to analyze saccadic movement from eye tracking signals. Bellet et al. [110] presents a deep learning approach by training a custom CNN to detect saccadic movement from a video based eye tracker.

ECs offer several advantages for detecting saccades compared to conventional frame-based cameras. Similar to the way ECs capture blinks, during saccadic eye movements, the rapid motion of the pupil results in a high density of events with positive and negative polarities reversing as the eye moves from one position to another as illustrated in Fig 16.

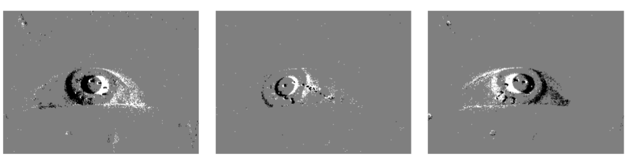

FIGURE 16. Illustration of event frames during a saccade.

This distinctive event pattern can be leveraged to identify saccades. One potential approach is to analyze the movement of events around the pupil area. By detecting these characteristic event patterns, saccades can be identified. Another promising method is to represent the event data into frames and apply conventional saccade detection algorithms. While not fully explored yet, this approach could yield good results by combining the strengths of ECs and existing saccade detection techniques. Despite the potential, current research has not yet fully exploited the application of ECs in saccade detection. This represents a significant opportunity for groundbreaking discoveries in both neuroscience and clinical diagnostics.

F. MICROSACCADES

Microsaccades are small, rapid, involuntary eye movements that occur during visual fixation, similar to miniature versions of voluntary saccadic eye movements but much smaller in amplitude (less than 1°) with high velocity and brief duration[[111]. They are generated by the same neuronal mechanisms that control larger saccades, as evidenced by their following the same main sequence relationship between peak velocity and amplitude [112].

The function of microsaccades is still debated, but they are thought to correct for small eye drifts to maintain fixation on a target, prevent fading of the retinal image, relate to cognitive processes like attention and reading [113], etc. Recent findings suggest that microsaccade rate increases with visual load, indicating that they reflect the visual complexity of a task rather than its mental demand [114]. This implies that monitoring microsaccade rate could be useful in applied settings where visual attention is critical, such as air traffic control or driver monitoring systems. Microsaccades can also be used as a diagnostic tool, as their characteristics change in various neurological and ophthalmic diseases, with automated detection methods used to identify them in eye movement recordings [115].

ECs offer a unique opportunity for detecting microsaccades, evidenced by their ability to perform very rapid eye tracking beyond 10000hz [41] and accurate blink detection Microsaccades typically last less than 2 milliseconds and cover an angular range of less than 2 degrees, making them much faster than regular saccades, blinks, etc. Conventional cameras are limited in their ability to capture these ultra-fast eye movements, as they are typically constrained to frame rates below 1000 Hz. Our review did not find any research to date that attempts to detect microsaccades, but as ECs can operate at equivalent frame rates greater than 10,000 Hz, they would appear to offer the capability required to more accurately detect and track microsaccades over today's video-based eye-trackers.

VI. CHALLENGES AND LIMITATIONS

The unique features of ECs have proven to be useful for eye motion analysis tasks, where capturing quick, subtle movements and operating across various lighting conditions are critical. However, there are several challenges and limitations associated with the use of ECs for such tasks.

In this section, we will discuss these challenges based on the tasks reviewed.

A. EYE TRACKING

Eye tracking accuracy can vary notably between users due to anatomical differences in the eyes, such as eyelid shape and eye size. External factors, including changes in lighting and reflections, can impede tracking accuracy, which can be challenging to maintain consistency across various environments. Previous research has made significant advancements in the use of ECs for high-frequency eye tracking. However, most of these studies have focused on bandwidth rather than accuracy. Inaccuracies can arise when other eye parts obstruct the pupil region, for instance during involuntary blinking, leading to erratic and intermittent pupil movement traces. Some studies, like Swift-Eye [94], have proposed methods to address this issue. However, there are still significant gaps in latency, making these solutions suitable only for offline tasks. This limitation restricts their applicability in real-time scenarios where low latency is required. Retina [44] proposes the use of SNNs, However the continuous resetting of neurons leads to inaccuracies in eye-tracking on neuromorphic hardware.

B. GAZE ESTIMATION

Accurately determining where a user is looking can be challenging, especially in dynamic settings or over long distances even in RGB cameras that contain colour information. The challenge of representations which best capture different eye parts are coupled with the need for rigorous calibration procedures which can be time-consuming and requires repetition if the setup or user position changes. E-gaze [22] proposed an approach to gaze estimation which meets the low latency target and does not require additional equipment for effective pupil localization, however there is the need to continuously re-calibrate whenever there is a change in position of the measuring device and distance.

C. BLINK ESTIMATION

Distinguishing between blinks and other eye-related movements (e.g., squinting) can lead to false positives or negatives in blink detection. Fast blinks might be missed by slower camera systems, while slow, deliberate blinks might be misinterpreted as gaze holds or look-aways.

D. PUPIL SEGMENTATION / TRACKING

In Applications such as pupil tracking, color information can be important, for example, in distinguishing different parts of the eye or in assessing blood flow and health conditions. ECs are highly sensitive to changes in the scene, which can include movement but also changes in lighting conditions. This sensitivity means they can generate a lot of noise, with events triggered by changes not relevant to the eye motion being studied. Filtering out this noise without losing relevant information can be challenging. Inadequate contrast between the pupil and the iris, particularly in individuals with dark eyes, can make it difficult to segment and track the pupil accurately. Changes in pupil size due to variations in emotional state, focus, or ambient light levels can complicate consistent tracking and require adaptive algorithms that can handle such dynamics.

E. SACCADE ESTIMATION

ECs are designed to excel in temporal resolution, capturing changes at very high frequencies. However, there is often a trade-off between temporal and spatial resolution. Improving one can sometimes come at the expense of the other. For applications requiring high spatial and temporal resolution such as saccades, low spatial resolution which can lead to less information in eye regions can result in less precise measurement of saccade amplitude, velocity and trajectory. Hence there is a need to develop algorithms and representations which can help mitigate these effects.

VII. POTENTIAL FUTURE DIRECTIONS

Building on the challenges and limitations identified in the preceding section, several promising avenues for future research and development can address these issues and advance the field of eye motion analysis utilizing event cameras (ECs). This section details potential directions for overcoming current limitations and enhancing the capabilities of ECs in eye tracking, gaze estimation, blink detection, pupil segmentation, and saccade estimation.

- **Enhanced Performance at Sub-Millisecond Temporal Resolution**: Focus on improving system performance to effectively utilize the full capabilities of event cameras with sub-millisecond temporal resolution, such as 1 millisecond. Research should aim to optimize algorithms and processing techniques that can fully leverage the high temporal resolution of ECs. This includes developing methods for efficiently handling and analyzing data at these high frequencies to ensure that the benefits of fine-grained temporal resolution are realized in practical applications.
- **Enhanced Calibration Techniques**: Develop more robust and automated calibration methods to improve tracking accuracy and reduce the need for frequent recalibration. Innovations could include adaptive calibration algorithms, real-time adjustment techniques, and AI-driven models to handle diverse anatomical and environmental conditions.
- **Improved (Micro-)Saccade Detection Algorithms**: To address the trade-off between temporal and spatial resolution in saccade estimation, future research should explore novel algorithms that effectively balance these dimensions. Approaches such as multi-scale event processing and the integration of high-resolution spatial data with high-frequency temporal data could offer more precise measurements of saccade amplitude, velocity, and trajectory. Additionally, the high-speed capabilities of event cameras present an opportunity to detect

microsaccades, which are often too subtle for traditional methods. Investigating hybrid event and frame-based systems may also provide complementary advantages in capturing both large saccades and microsaccades, thereby enhancing overall accuracy and sensitivity in eye movement analysis.

- **Context-Aware Blink and Eye Movement Detection**: To improve blink estimation and differentiate between blinks and other eye-related movements, future research should investigate context-aware detection methods. Incorporating contextual information, such as gaze direction and head pose, into blink detection algorithms may reduce false positives and negatives. Moreover, developing models that can accurately differentiate between fast and slow blinks could enhance the reliability of blink estimation across diverse applications.
- **Utilizing Event Camera Biases for Data Control**: To manage the flow of event camera data more effectively, future research should explore leveraging inherent biases in event camera systems. By understanding and utilizing these biases, it may be possible to implement strategies that control data flow and prioritize relevant information, thereby optimizing system performance and reducing computational overhead [116].

By addressing these key areas, future research can overcome the current limitations of ECs in eye motion analysis and unlock new possibilities for applications in fields such as human-computer interaction, neuroscience, and assistive technologies.

VIII. CONCLUSION AND FUTURE DIRECTIONS

Eye Motion Analysis with Event Cameras (ECs) is an emerging field of promising research with demonstrated success in a variety of tasks. This review provides a thorough synthesis of the current landscape of event camera (EC)-based eye motion analysis, offering an in-depth examination of advancements, methodologies, and outcomes in the field. By categorizing and critically assessing existing research, it provides a clear overview of progress made and establishes a foundation for future exploration. The review also identifies significant task-specific challenges, limitations, and gaps in knowledge, offering a roadmap for addressing these issues through further investigation. In addition, it proposes strategic directions for future research, emphasizing how ECs can be better integrated into eye motion analysis. This includes exploring interdisciplinary collaborations, expanding into new application areas, and developing innovative methodologies. By focusing on these future directions, the review aims to enhance the accuracy and efficacy of ECs, thereby broadening their impact and utility in various domains.

ACKNOWLEDGMENT

The authors would like to thank the Time-Event Group, Fotonation-Tobii, for their support throughout the project.

REFERENCES

[1] R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*. London, U.K.: Oxford Univ. Press, Jun. 2015.

[2] N. Alsaeedi and D. Wloka, "Real-time eyeblink detector and eye state classifier for virtual reality (VR) headsets (head-mounted displays, HMDs)," *Sensors*, vol. 19, no. 5, p. 1121, Mar. 2019.

[3] M. Barz and D. Sonntag, "Automatic visual attention detection for mobile eye tracking using pre-trained computer vision models and human gaze," *Sensors*, vol. 21, no. 12, p. 4143, Jun. 2021.

[4] A. A. Bamidele, K. Kamardin, N. Syazarin, S. Mohd, I. Shafi, A. Azizan, N. Aini, and H. Mad, "Non-intrusive driver drowsiness detection based on face and eye tracking," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 1–21, 2019.

[5] N. I. Vargas-Cuentas, A. Roman-Gonzalez, R. H. Gilman, F. Barrientos, J. Ting, D. Hidalgo, K. Jensen, and M. Zimic, "Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0188826.

[6] T. T. Brunyé, T. Drew, D. L. Weaver, and J. G. Elmore, "A review of eye tracking for understanding and improving diagnostic interpretation," *Cognit. Res., Princ. Implications*, vol. 4, no. 1, pp. 1–16, Dec. 2019.

[7] D. Cazzato, M. Leo, C. Distante, and H. Voos, "When i look into your eyes: A survey on computer vision contributions for human gaze estimation and tracking," *Sensors*, vol. 20, no. 13, p. 3739, Jul. 2020.

[8] W. Shariff, M. S. Dilmaghani, P. Kielty, M. Moustafa, J. Lemley, and P. Corcoran, "Event cameras in automotive sensing: A review," *IEEE Access*, vol. 12, pp. 51275–51306, 2024.

[9] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, Jan. 2022.

[10] F. Becattini, L. Berlincioni, L. Cultrera, and A. Del Bimbo, "Neuromorphic face analysis: A survey," 2024, *arXiv:2402.11631*.

[11] Y. Shin, H. W. Lim, M. H. Kang, M. Seong, H. Cho, and J. H. Kim, "Normal range of eye movement and its relationship to age," *Acta Ophthalmologica*, vol. 94, no. S256, p. 747, Oct. 2016.

[12] A. Wong, "Listing's law: Clinical significance and implications for neural control," *Surv. Ophthalmol.*, vol. 49, no. 6, pp. 563–575, Nov. 2004.

[13] Y. Hirata, J. Nishiyama, and S. Kinoshita, "Detection and prediction of drowsiness by reflexive eye movements," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 4015–4018.

[14] P. Ebrahim, W. Stolzmann, and B. Yang, "Eye movement detection for assessing driver drowsiness by electrooculography," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 4142–4148.

[15] V. E. Wilkinson, M. L. Jackson, J. Westlake, B. Stevens, M. Barnes, P. Swann, S. M. W. Rajaratnam, and M. E. Howard, "The accuracy of eyelid movement parameters for drowsiness detection," *J. Clin. Sleep Med.*, vol. 9, no. 12, pp. 1315–1324, Dec. 2013.

[16] R. Schleicher, N. Galley, S. Briest, and L. Galley, "Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired?" *Ergonomics*, vol. 51, no. 7, pp. 982–1010, Jul. 2008.

[17] L. Spiering and O. Dimigen, "(Micro)saccade-related potentials during face recognition: A study combining EEG, eye-tracking, and deconvolution modeling," *Attention, Perception, Psychophys.*, pp. 1–22, 2024. [Online]. Available: https://link.springer.com/article/10.3758/s13414-024-02846-1#citeas

[18] N. Li, A. Bhat, and A. Raychowdhury, "E-track: Eye tracking with event camera for extended reality (XR) applications," in *Proc. IEEE 5th Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, Jun. 2023, pp. 1–5.

[19] R. Alessi, "Eye tracking con camera neuromorfica," Ph.D. thesis, Università degli studi di Genova, Genoa, Italy, 2023.

[20] C. Ryan, B. O'Sullivan, A. Elrasad, A. Cahill, J. Lemley, P. Kielty, C. Posch, and E. Perot, "Real-time face & eye tracking and blink detection using event cameras," *Neural Netw.*, vol. 141, pp. 87–97, Sep. 2021.

[21] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, "EKLT: Asynchronous photometric feature tracking using events and frames," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 601–618, Mar. 2020.

[22] N. Li, M. Chang, and A. Raychowdhury, "E-gaze: Gaze estimation with event camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 4796–4811, Jul. 2024.

[23] W. Porterfield, "An essay concerning the motions of our eyes. Part I. Of their external motions," *Edinburgh Med. Essays Observ.*, vol. 3, no. 1737, pp. 160–263, 1737.
[24] K. Harezlak and P. Kasprowski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Computerized Med. Imag. Graph.*, vol. 65, pp. 176–190, Apr. 2018.
[25] A. A. Akinyelu and P. Blignaut, "Convolutional neural network-based methods for eye gaze estimation: A survey," *IEEE Access*, vol. 8, pp. 142581–142605, 2020.
[26] A. F. Klaib, N. O. Alsrehin, W. Y. Melhem, H. O. Bashtawi, and A. A. Magableh, "Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies," *Expert Syst. Appl.*, vol. 166, Mar. 2021, Art. no. 114037.
[27] D. Kang, Y. K. Lee, and J. Jeong, "Exploring the potential of event camera imaging for advancing remote pupil-tracking techniques," *Appl. Sci.*, vol. 13, no. 18, p. 10357, Sep. 2023.
[28] G. Zhao, Y. Yang, J. Liu, N. Chen, Y. Shen, H. Wen, and G. Lan, "EV-eye: Rethinking high-frequency eye tracking through the lenses of event cameras," in *Proc. Adv. Neural Inf. Process. Syst.*, 2024, pp. 1–11.
[29] G. Lenz, S.-H. Ieng, and R. Benosman, "Event-based face detection and tracking using the dynamics of eye blinks," *Frontiers Neurosci.*, vol. 14, Jul. 2020, Art. no. 495936.
[30] Y. Feng, N. Goulding-Hotta, A. Khan, H. Reyserhove, and Y. Zhu, "Real-time gaze tracking with event-driven eye segmentation," in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Mar. 2022, pp. 399–408.
[31] W. Shariff, M. S. Dilmaghani, P. Kielty, J. Lemley, M. A. Farooq, F. Khan, and P. Corcoran, "Neuromorphic driver monitoring systems: A computationally efficient proof-of-concept for driver distraction detection," *IEEE Open J. Veh. Technol.*, vol. 4, pp. 836–848, 2023.
[32] J. Kim, Y. M. Kim, Y. Wu, R. Zahreddine, W. A. Welge, G. Krishnan, S. Ma, and J. Wang, "Privacy-preserving visual localization with event cameras," 2022, *arXiv:2212.03177*.
[33] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3852–3861.
[34] P. Zhang, S. Zhu, and E. Y. Lam, "Event encryption: Rethinking privacy exposure for neuromorphic imaging," *Neuromorphic Comput. Eng.*, vol. 4, no. 1, Mar. 2024, Art. no. 014002.
[35] S. Ahmad, P. Morerio, and A. Del Bue, "Person re-identification without identification via event anonymization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 11132–11141.
[36] S. Ahmad, P. Morerio, and A. Del Bue, "Event anonymization: Privacy-preserving person re-identification and pose estimation in event-based vision," *IEEE Access*, vol. 12, pp. 66964–66980, 2024.
[37] Y. Yang, G. Zhao, and Y. Shen, "High frequency event-based eye tracking towards mental health diagnosis," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2023, pp. 335–339.
[38] T. Stoffregen, H. Daraei, C. Robinson, and A. Fix, "Event-based kilohertz eye tracking using coded differential lighting," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 3937–3945.
[39] Y. Jiang, W. Wang, L. Yu, and C. He, "Eye tracking based on event camera and spiking neural network," *Electronics*, vol. 13, no. 14, p. 2879, 2024.
[40] A. Banerjee, N. K. Mehta, S. S. Prasad, S. Saurav, and S. Singh, "Gaze-vector estimation in the dark with temporally encoded event-driven neural networks," 2024, *arXiv:2403.02909*.
[41] A. N. Angelopoulos, J. N. P. Martel, A. P. S. Kohli, J. Conradt, and G. Wetzstein, "Event based, near eye gaze tracking beyond 10,000Hz," 2020, *arXiv:2004.03577*.
[42] D. Kang and D. Kang, "Event camera-based pupil localization: Facilitating training with event-style translation of RGB faces," *IEEE Access*, vol. 11, pp. 142304–142316, 2023.
[43] Q. Chen, Z. Wang, S.-C. Liu, and C. Gao, "3ET: Efficient event-based eye tracking using a change-based ConvLSTM network," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, vol. 28, Oct. 2023, pp. 1–5.
[44] P. Bonazzi, S. Bian, G. Lippolis, Y. Li, S. Sheik, and M. Magno, "Retina: Low-power eye tracking with event camera and spiking hardware," 2023, *arXiv:2312.00425*.
[45] Z. Wang, "Event-based eye tracking. AIS 2024 challenge survey," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 5810–5825.
[46] C. Ryan, A. Elrasad, W. Shariff, J. Lemley, P. Kielty, P. Hurney, and P. Corcoran, "Real-time multi-task facial analytics with event cameras," *IEEE Access*, vol. 11, pp. 76964–76976, 2023.
[47] R. Daza, A. Morales, J. Fierrez, and R. Tolosana, "MEBAL: A multimodal database for eye blink detection and attention level estimation," in *Proc. Companion Publication Int. Conf. Multimodal Interact.*, Oct. 2020, pp. 32–36.
[48] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4511–4520.
[49] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2176–2184.
[50] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges, "ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 365–381.
[51] U. Bissarinova, T. Rakhimzhanova, D. Kenzhebalin, and H. A. Varol, "Faces in event streams (FES): An annotated face dataset for event cameras," *Sensors*, vol. 24, no. 5, p. 1409, 2024, doi: 10.3390/s2405140.
[52] C. Scheerlinck, H. Rebecq, T. Stoffregen, N. Barnes, R. Mahony, and D. Scaramuzza, "CED: Color event camera dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1684–1693.
[53] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting static image datasets to spiking neuromorphic datasets using saccades," *Frontiers Neurosci.*, vol. 9, p. 437, Nov. 2015.
[54] J. Kaiser, J. C. Vasquez Tieck, C. Hubschneider, P. Wolf, M. Weber, M. Hoff, A. Friedrich, K. Wojtasik, A. Roennau, R. Kohlhaas, R. Dillmann, and J. M. Zöllner, "Towards a framework for end-to-end control of a simulated vehicle with spiking neural networks," in *Proc. IEEE Int. Conf. Simul., Modeling, Program. Auto. Robots (SIMPAR)*, Dec. 2016, pp. 127–134.
[55] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *Int. J. Robot. Res.*, vol. 36, no. 2, pp. 142–149, Feb. 2017.
[56] W. Li, S. Saeedi, J. McCormac, R. Clark, D. Tzoumanikas, Q. Ye, Y. Huang, R. Tang, and S. Leutenegger, "InteriorNet: Mega-scale multi-sensor photo-realistic indoor scenes dataset," 2018, *arXiv:1809.00716*.
[57] H. Rebecq, D. Gehrig, and D. Scaramuzza, "ESIM: An open event camera simulator," in *Proc. 2nd Conf. Robot Learn. (CoRL)*, 2018, pp. 969–982.
[58] M. J. H. Pantho, J. M. Mbongue, P. Bhowmik, and C. Bobda, "Event camera simulator design for modeling attention-based inference architectures," *J. Real-Time Image Process.*, vol. 19, no. 2, pp. 363–374, Apr. 2022.
[59] Y. Hu, S.-C. Liu, and T. Delbruck, "v2e: From video frames to realistic DVS events," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 1312–1321.
[60] J. Li, Z. Zhu, J. Hou, J. Hou, and J. Wu, "Modeling state shifting via local–global distillation for event-frame gaze tracking," 2024, *arXiv:2404.00548*.
[61] G. Careddu, "Event-based eye tracking for smart eyewear," Politechnico Milano, Italy, 2022. [Online]. Available: https://hdl.handle.net/10589/215580
[62] R. Cao, D. Galor, A. Kohli, J. L Yates, and L. Waller, "Noise2Image: Noise-enabled static scene recovery for event cameras," 2024, *arXiv:2404.01298*.
[63] L. Yuanyuan, L. Zhaohui, L. Zhou, Q. Wenlong, and Z. Haiyang, "Denoising algorithm based on event camera," in *Proc. 6th Conf. Frontiers Opt. Imag. Technol., Novel Detect. Technol.*, vol. 9, Apr. 2024, p. 14.
[64] Y. Duan, "LED: A large-scale real-world paired dataset for event camera denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 25637–25647.
[65] W. Lin, Y. Li, C. Xu, and L. Liu, "A motion denoising algorithm with Gaussian self-adjusting threshold for event camera," *Vis. Comput.*, vol. 40, no. 9, pp. 6567–6580, Sep. 2024.
[66] L. Vincent, "Morphological area openings and closings for grey-scale images," in *Shape in Picture*. Cham, Switzerland: Springer, 1994, pp. 197–208.
[67] R. W. Baldwin, M. Almatrafi, V. Asari, and K. Hirakawa, "Event probability mask (EPM) and event denoising convolutional neural network (EDnCNN) for neuromorphic cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1698–1707.

[68] S. Ding, J. Chen, Y. Wang, Y. Kang, W. Song, J. Cheng, and Y. Cao, "E-MLB: Multilevel benchmark for event-based camera denoising," *IEEE Trans. Multimedia*, vol. 26, pp. 65–76, 2023.

[69] X. Zheng, Y. Liu, Y. Lu, T. Hua, T. Pan, W. Zhang, D. Tao, and L. Wang, "Deep learning for event-based vision: A comprehensive survey and benchmarks," 2023, *arXiv:2302.08890*.

[70] N. Zubić, D. Gehrig, M. Gehrig, and D. Scaramuzza, "From chaos comes order: Ordering event representations for object recognition and detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12800–12810.

[71] G. Lenz, S.-H. Ieng, and R. Benosman, "Event-based face detection and tracking in the blink of an eye," 2018, *arXiv:1803.10106*.

[72] G. Chen, L. Hong, J. Dong, P. Liu, J. Conradt, and A. Knoll, "EDDD: Event-based drowsiness driving detection through facial motion analysis with neuromorphic vision sensor," *IEEE Sensors J.*, vol. 20, no. 11, pp. 6170–6181, Jun. 2020.

[73] G. Chen, F. Wang, X. Yuan, Z. Li, Z. Liang, and A. Knoll, "NeuroBiometric: An eye blink based biometric authentication system using an event-based neuromorphic vision sensor," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 1, pp. 206–218, Jan. 2021.

[74] T. Kagemoto and K. Takemura, "Event-based pupil tracking using bright and dark pupil effect," in *Adjunct Proc. 36th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2023, pp. 1–3.

[75] Z. Wang, Z. Wan, H. Han, B. Liao, Y. Wu, W. Zhai, Y. Cao, and Z.-J. Zha, "MambaPupil: Bidirectional selective recurrent model for event-based eye tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 5762–5770.

[76] W. Fang, Z. Yu, T. Masquelier, Y. Chen, and Y. Tian, "Spike-based residual blocks," in *Proc. Adv. Neural Inform. Process. Syst.*, vol. 34, 2021, pp. 21056–21069.

[77] W. Fang, Z. Yu, Y. Chen, T. Masquelier, T. Huang, and Y. Tian, "Incorporating learnable membrane time constant to enhance learning of spiking neural networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2641–2651.

[78] A. M. Soccini, "Gaze estimation based on head movements in virtual reality applications using deep learning," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2017, pp. 413–414.

[79] S. M. Shah, Z. Sun, K. Zaman, A. Hussain, M. Shoaib, and L. Pei, "A driver gaze estimation method based on deep learning," *Sensors*, vol. 22, no. 10, p. 3959, May 2022.

[80] J. Li, Z. Chen, Y. Zhong, H.-K. Lam, J. Han, G. Ouyang, X. Li, and H. Liu, "Appearance-based gaze estimation for ASD diagnosis," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6504–6517, Jul. 2022.

[81] S. Ghosh, A. Dhall, M. Hayat, J. Knibbe, and Q. Ji, "Automatic gaze analysis: A survey of deep learning based approaches," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 1, pp. 61–84, Jan. 2024.

[82] S. Cristina and K. P. Camilleri, "Model-based head pose-free gaze estimation for assistive communication," *Comput. Vis. Image Understand.*, vol. 149, pp. 157–170, Aug. 2016.

[83] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16495–16519, 2017.

[84] S. Mentasti, F. Lattari, R. Santambrogio, G. Careddu, and M. Matteucci, "Event-based eye tracking for smart eyewear," in *Proc. Symp. Eye Tracking Res. Appl.*, 2024, pp. 1–7.

[85] Y. Cheng, H. Wang, Y. Bao, and F. Lu, "Appearance-based gaze estimation with deep learning: A review and benchmark," 2021, *arXiv:2104.12668*.

[86] A. Banerjee, S. S. Prasad, N. K. Mehta, H. Kumar, S. Saurav, and S. Singh, "Gaze detection using encoded retinomorphic events," in *Proc. Int. Conf. Intell. Hum. Comput. Interact.* Cham, Switzerland: Springer, 2022, pp. 442–453.

[87] R. Paprocki and A. Lenskiy, "What does eye-blink rate variability dynamics tell us about cognitive performance?" *Frontiers Hum. Neurosci.*, vol. 11, p. 620, Dec. 2017.

[88] S. Shekari Soleimanloo, V. E. Wilkinson, J. M. Cori, J. Westlake, B. Stevens, L. A. Downey, B. A. Shiferaw, S. M. W. Rajaratnam, and M. E. Howard, "Eye-blink parameters detect on-road track-driving impairment following severe sleep deprivation," *J. Clin. Sleep Med.*, vol. 15, no. 9, pp. 1271–1284, Sep. 2019.

[89] A. R. Bentivoglio, S. B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese, "Analysis of blink rate patterns in normal subjects," *Movement Disorders*, vol. 12, no. 6, pp. 1028–1034, Nov. 1997.

[90] J. Hollander and S. Huette, "Extracting blinks from continuous eye-tracking data in a mind wandering paradigm," *Consciousness Cognition*, vol. 100, Apr. 2022, Art. no. 103303.

[91] R. Daza, D. DeAlcala, A. Morales, R. Tolosana, R. Cobos, and J. Fierrez, "ALEBk: Feasibility study of attention level estimation via blink detection applied to e-learning," 2021, *arXiv:2112.09165*.

[92] P. A. D. L. Medeiros, G. V. S. D. Silva, F. R. D. S. Fernandes, I. Sánchez-Gendriz, H. W. C. Lins, D. M. D. S. Barros, D. A. P. Nagem, and R. A. D. M. Valentim, "Efficient machine learning approach for volunteer eye-blink detection in real-time using webcam," *Expert Syst. Appl.*, vol. 188, Feb. 2022, Art. no. 116073.

[93] M. Susman and V. Kimmelman, "Eye blink detection in sign language data using CNNs and rule-based methods," in *Proc. LREC-COLING 11th Workshop Represent. Process. Sign Lang., Eval. Sign Lang. Resour.*, 2024, pp. 361–369.

[94] T. Zhang, Y. Shen, G. Zhao, L. Wang, X. Chen, L. Bai, and Y. Zhou, "Swift-eye: Towards anti-blink pupil tracking for precise and robust high-frequency near-eye movement analysis with event cameras," *IEEE Trans. Vis. Comput. Graphics*, vol. 30, no. 5, pp. 2077–2086, May 2024.

[95] J. T.-Y. Wang, "Pupil dilation and eye tracking," in *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide*. New York, NY, USA: Psychology Press, 2011, pp. 185–204.

[96] B. Mahanama, Y. Jayawardana, S. Rengarajan, G. Jayawardena, L. Chukoskie, J. Snider, and S. Jayarathna, "Eye movement and pupil measures: A review," *Frontiers Comput. Sci.*, vol. 3, Jan. 2022, Art. no. 733531.

[97] R. Abdal, P. Zhu, N. J. Mitra, and P. Wonka, "StyleFlow: Attribute-conditioned exploration of StyleGAN-generated images using conditional continuous normalizing flows," *ACM Trans. Graph.*, vol. 40, no. 3, pp. 1–21, Jun. 2021.

[98] S. Tulyakov, D. Gehrig, S. Georgoulis, J. Erbach, M. Gehrig, Y. Li, and D. Scaramuzza, "Time lens: Event-based video frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16150–16159.

[99] N. R. Carlson, *Psychology: The Science of Behaviour. Study Guide*. London, U.K.: Pearson, 2010.

[100] A. F. Fuchs, "Saccadic and smooth pursuit eye movements in the monkey," *J. Physiol.*, vol. 191, no. 3, pp. 609–631, Aug. 1967.

[101] R. Schweitzer and M. Rolfs, "Definition, modeling, and detection of saccades in the face of post-saccadic oscillations," in *Neuromethods*. Cham, Switzerland: Springer, 2022, pp. 69–95.

[102] R. Mani, L. Asper, and S. K. Khuu, "Deficits in saccades and smooth-pursuit eye movements in adults with traumatic brain injury: A systematic review and meta-analysis," *Brain Injury*, vol. 32, no. 11, pp. 1315–1336, Sep. 2018.

[103] A. Cade and P. R. K. Turnbull, "Classification of short and long term mild traumatic brain injury using computerized eye tracking," *Sci. Rep.*, vol. 14, no. 1, p. 12686, Jun. 2024.

[104] H. Matsumoto, Y. Terao, T. Furubayashi, A. Yugeta, H. Fukuda, M. Emoto, R. Hanajima, and Y. Ugawa, "Small saccades restrict visual scanning area in Parkinson's disease," *Movement Disorders*, vol. 26, no. 9, pp. 1619–1626, Aug. 2011.

[105] X. Liao, J. Yao, H. Tang, Y. Xing, X. Zhao, D. Nie, P. Luan, and G. Li, "Deciphering Parkinson's disease through eye movements: A promising tool for early diagnosis in the face of cognitive impairment," *Int. J. Clin. Pract.*, vol. 2024, no. 1, Jan. 2024, Art. no. 5579238.

[106] Y. Yu, J. Wang, L. Si, H. Sun, X. Liu, X. Li, and W. Yan, "Smooth pursuit and reflexive saccade in discriminating multiple-system atrophy with predominant parkinsonism from Parkinson's disease," *J. Clin. Neurol.*, vol. 20, no. 2, p. 194, 2024.

[107] M. Attre, "Seeing eye to eye: A machine learning approach to automated saccade analysis," Honors Scholar Theses. 613, Univ. Connecticut, Mansfield, CT, USA, 2019. [Online]. Available: https://digitalcommons.lib.uconn.edu/srhonors_theses/613

[108] K. Tirdad, A. D. Cruz, C. Austin, A. Sadeghian, S. M. Nia, and M. Cusimano, "Machine learning-based approach to analyze saccadic eye movement in patients with mild traumatic brain injury," *Comput. Methods Programs Biomed. Update*, vol. 1, Jan. 2021, Art. no. 100026.

[109] M. J. Lutfullaeva, "Application of machine learning methods for saccades and fixation detection from eye-tracking signal," *AIP Conf. Proc.*, vol. 2849, no. 1, 2023, Art. no. 090016.

[110] M. E. Bellet, J. Bellet, H. Nienborg, Z. M. Hafed, and P. Berens, "Human-level saccade detection performance using deep neural networks," *J. Neurophysiol.*, vol. 121, no. 2, pp. 646–661, Feb. 2019.

[111] E. Krueger, A. Schneider, B. D. Sawyer, A. Chavaillaz, A. Sonderegger, R. Groner, and P. A. Hancock, ''Microsaccades distinguish looking from seeing,'' *J. Eye Movement Res.*, vol. 12, no. 6, 2019.

[112] X. Troncoso, J. Otero-Millan, S. Macknik, I. Serrano-Pedraza, and S. Martinez-Conde, ''Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator,'' *J. Vis.*, vol. 9, no. 8, p. 447, Mar. 2010.

[113] K. Krejtz, J. Żurawska, A. T. Duchowski, and S. Wichary, ''Pupillary and microsaccadic responses to cognitive effort and emotional arousal during complex decision making,'' *J. Eye Movement Res.*, vol. 13, no. 5, 2020.

[114] A. Schneider, A. Sonderegger, E. Krueger, Q. Meteier, P. Luethold, and A. Chavaillaz, ''The interplay between task difficulty and microsaccade rate: Evidence for the critical role of visual load,'' *J. Eye Movement Res.*, vol. 13, no. 5, 2020.

[115] P. Essig, A. Leube, K. Rifai, and S. Wahl, ''Microsaccadic rate signatures correlate under monocular and binocular stimulation conditions,'' *J. Eye Movement Res.*, vol. 11, no. 4, 2020.

[116] M. S. Dilmaghani, W. Shariff, M. A. Farooq, J. Lemley, and P. Corcoran, ''Optimization of event camera bias settings for a neuromorphic driver monitoring system,'' *IEEE Access*, vol. 12, pp. 32959–32970, 2024.

KHADIJA IDDRISU received the B.Sc. degree in computer science from the University for Development Studies, Ghana, in 2020, and the M.Sc. degree in mathematical sciences from the African Institute for Mathematical Sciences (AIMS), Ghana, in 2021. She is currently pursuing the Ph.D. degree with Dublin City University. She is a Ph.D. Research Intern with Fotonation-Tobii. Her research interests include the applications of machine learning and computer vision for eye motion analysis, specifically in the use of event cameras for saccadic eye movement detection and analysis. She was a recipient of the Google and Facebook-sponsored African Masters of Machine Intelligence (AMMI) Program at AIMS Senegal, from 2022 to 2023. She was a recipient of the SFI Ph.D. Program funded by the Insight SFI Research Centre for Data Analytics.

WASEEM SHARIFF (Graduate Student Member, IEEE) received the B.E. degree in computer science from the Nagarjuna College of Engineering and Technology (NCET), in 2019, and the M.Sc. degree in computer science (artificial intelligence) from the National University of Ireland Galway (NUIG), in 2020. He is currently pursuing the Ph.D. degree with the University of Galway, under the IRC Employment Ph.D. Program. He is a Research and Development Engineer with Foto-Nation (Tobii), Galway. His research interests include machine learning for computer vision applications, with a particular emphasis on automotive driver monitoring applications and neuromorphic event-based vision.

PETER CORCORAN (Fellow, IEEE) was the Co-Founder in several start-up companies, notably FotoNation (currently the Imaging Division, Xperi Corporation). He is currently the Personal Chair of electronic engineering with the School of Engineering, University of Galway (formerly known as the National University of Ireland Galway). He has over 600 technical publications and patents, over 100 peer-reviewed journal articles, and 120 international conference papers. He was the co-inventor of more than 300 granted U.S. patents. He is a member of the IEEE Consumer Electronics Society, for over 25 years. He is an IEEE Fellow recognized for his contributions to digital camera technologies, notably in-camera redeye correction, and facial detection. He is the Editor-in-Chief and the Founding Editor of *IEEE Consumer Electronics Magazine*.

NOEL E. O'CONNOR (Member, IEEE) is currently a Full Professor with the School of Electronic Engineering, Dublin City University, Ireland, and the CEO of the Insight SFI Research Centre for Data Analytics. His research interests include multimedia content analysis, computer vision, machine learning, information fusion, and multimodal analysis for diverse applications. He is an Area Editor of *Signal Processing: Image Communication* (Elsevier) and an Associate Editor of *Image and Video Processing* journal (Springer).

JOE LEMLEY (Member, IEEE) received the B.S. degree in computer science and the master's degree in computational science from Central Washington University, in 2006 and 2016, respectively, and the Ph.D. degree from the National University of Ireland Galway. He is currently a Principal Research and Development Engineer and the Manager of FotoNation (tobii), Galway. His field of work is machine learning using deep neural networks for tasks related to computer vision. His current research interests include computer vision and signal processing for driver monitoring systems.

SUZANNE LITTLE received the Ph.D. degree from The University of Queensland, Australia, in 2006. She is currently a Professor with the School of Computing, Dublin City University. She is also a Principal Investigator of the Science Foundation Ireland (SFI), Insight SFI Research Centre for Data Analytics, working in the areas of media analytics, information access, and retrieval across a variety of application domains. She has worked on several EU projects in the areas of multimedia, technology-enhanced education, security, autonomous vehicles, and big data. She is a Funded Investigator of the i-Form SFI Advanced Manufacturing Research Centre and the Co-Director of the SFI Centre for Research Training in Artificial Intelligence. Her research interests include video analysis, semantic search, and data integration