



RESEARCH ARTICLE

OPEN ACCESS

EYE TRACKING AND GAZE ESTIMATION: A SYSTEMATIC REVIEW OF DEEP LEARNING METHODS AND EDUCATIONAL APPLICATIONS

Khalid Sami Yousif ^{*1}, Emad A. Mohammed ² and Mayyada Al- Hammoshi ³

^{1, 2}Technical Engineering college for computer and AI, Mosul - Iraq

³Washington University of Science and Technology, Washington-USA

¹<https://orcid.org/0009-0001-8715-0065>, ²<https://orcid.org/0000-0002-2575-6213>, ³<https://orcid.org/0009-0007-5022-1046>

E-mail: *khalid.sami@ntu.edu.iq, e.a.mohammed@ntu.edu.iq, mayyadah70@gmail.com

ARTICLE INFO

Article History

Received on: March 30, 2026

Reviewed on: April 30, 2026

Accepted on: July 01, 2026

Published on: July 31, 2026

Keywords:

Gaze Estimation,

Eye Tracking,

Head Pose Estimation,

Deep Learning,

CNN,

Transformer.

ABSTRACT

Tracking of the eyes and head have occupied a crucial role in human computer interaction, virtual and augmented reality, accessibility systems and cognitive studies. This paper gives a critical review on gaze estimation and head pose tracking algorithms, data sets, and performance determinants. This article compares classical and deep learning-based methods, such as convolutional neural networks, transformer models, geometric, hybrid, and multi-task learning methods. The article also tests the robustness in the external conditions like blinking, occlusion, and changes in illumination, and cross-subject individual differences. The most important publicly available datasets are compared regarding the type of input, diversity of participants, the size of data, recording distance, and environmental conditions. This article also contrasts new webcam and specialized eye-tracking systems, and, in doing so, point out trade-offs between accuracy, cost and applicability in real-time. The review also stresses on the role of domain adaptation, contrastive learning, and multi-modal inputs in generalizing across domains. In general, the article contains an attempt to present a comprehensive picture of the existing situation to researchers and practitioners to assist them in designing and choosing gaze and head pose estimation systems to be used in various applications.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Gaze estimation and eye tracking have become the fundamental technologies in comprehending the human visual attention, mental processes, and the behavior of interaction. Fundamentally speaking, the techniques seek to ascertain the location of a person with their gaze, either through direct eye tracking or by inferring the direction of the gaze based on visual images or video streams [1]. Traditionally, eye tracking used specialized hardware systems, which had infrared illumination and high-speed cameras, which had the ability to capture fine-grained ocular characteristics like pupil placement and corneal reflections. Although these systems are highly accurate, they tend to be costly, obtrusive as well as restricted to controlled laboratory conditions. Consequently, their application in real-life use, especially in education and large-scale behavioral research has been limited [2]. The past few years have seen a drastic change of gaze estimation because of the development of computer vision and deep learning. Appearance-based approaches, which are methods that are based on the direct recognition of the direction of gaze based on the images of the face or eyes, have been in the limelight as a viable alternative to the traditional model-based approaches. In these techniques, convolutional neural networks (CNNs), transformer networks and hybrid geometric-learning networks are used to extract meaningful features of typical RGB images. This has in turn made it possible to carry out gaze tracking with the use of normal cameras without having to have special sensors. This has led to increased scalability and availability of eye tracking solutions most particularly web camera versions installed on both desktop and mobile computers [3]. The advent of low cost, web-based eye tracking systems is an important step toward the democratization of eye tracking technology. In contrast to specialized eye trackers, webcams are everywhere, cheap, and already integrated into the majority of consumer gadgets such as laptops, tablets and smartphones.

This has enabled researchers and practitioners to implement gaze tracking applications in a variety of real-world scenarios without enormous financial and technical constraints [4]. Moreover, the current deep learning architectures have been shown to mitigate the natural drawbacks of webcams, including reduced resolution, fluctuating light conditions, and the inability to run infrared, by training expressive representations of gaze-sensitive features. Open-source systems and frameworks have further made the process go faster on offering pre-made tools of real-time gaze estimation, calibration and data collection [5]. These developments are especially significant in such sphere as education, where student attention and engagement are important factors that have to be understood to enhance learning outcomes. Conventional strategies of gauging the engagement, including self-reporting, survey, or observation research, are frequently based on subjectivity, time-consuming, and not easily scaled. On the contrary, eye tracking using webcams is a non-invasive, automated technique of monitoring visual attention of students in the learning process. An examination of the gaze patterns will enable educators and other researchers to understand the interaction of students with teaching material, understand non-understood areas, and the effectiveness of the teaching techniques. This comes in handy particularly in online and remote learning where students cannot be observed directly [6]. In addition, the role of gaze estimation in education technologies facilitates the larger paradigm of learning analytics and intelligent tutoring systems. An example is the presentation of content which can be optimized using gaze data, or generate customized feedback, or real-time disengagement. Eye tracking can be used in visualization of cognitive processes in such subject matters as mathematics and statistics where visual attention is the main tool in problem solving and understanding. Consequently, the use of inexpensive gaze tracking systems can improve research and practice in the education field by allowing the collection of information to make decisions on large scales [7]. These positive developments notwithstanding, there are a number of challenges. Webcam-based systems are less accurate than dedicated eye trackers by their nature, and often susceptible to variations in head position, occlusions, and variations in the appearance of the user.

Moreover, the process of creating a strong generalization among various devices, settings, and target audiences remains a research topic. These have been tackled in recent works with methods like domain adaptation, contrastive learning and multi task learning, whose goal is to enhance the robustness and transferability of models. Ethical issues, such as privacy, data security, and informed consent, should be, conversely, handled with care at all times, especially in the case of the implementation of gaze tracking technologies in sensitive settings, such as education [8]. Since this area is quickly developing and numerous methods are suggested in the recent literature, it is evident that a thorough and systematic review is necessary that would synthesize the existing knowledge and outline the perspectives in future research. Although a number of surveys have investigated eye-tracking and gaze estimation through particular lens, e.g. hardware-based systems, deep learning methods, or educational practices, a gap still exists in the combination of these aspects within the framework of affordable solutions based on web-cams. It is with the intention of filling that gap that this review attempts to present a coherent overview of the state of the art, specifically focusing on approaches drawing on the use of standard cameras and present-day machine learning methods. There are three main purposes of this review. First, it aims at giving a clear and systematic overview of the basic ideas and methodology behind eye tracking and gaze estimation both conventional and modern. Second, it will also seek to review new developments in webcam-based systems in a systematic manner, particularly deep learning models, system architecture, and performance assessment. Third, the review discusses how these technologies have been used in the learning environment and how they have the potential to assess student attention, improve learning analytics, and facilitate adaptive learning environments. Besides providing a summary of the existing works, the review also has several major contributions. It suggests a unified taxonomy of gaze estimation algorithms to fits the constraints of low-cost camera configurations, compares the state-of-the-art methods in reference to accuracy, robustness, and computational cost, and locates key difficulties, which should be solved to allow implementation in the reality. Moreover, it identifies the future trends of research, such as cross-domain generalization, multimodal integration, and privacy-saving methods, which will have an influence on the future of webcam-based eye tracking.

II. FUNDAMENTALS OF EYE TRACKING AND GAZE ESTIMATION

The eye-tracking technology is a collection of techniques and apparatuses that monitor the eye positions and movements with the goal of identifying the point of gaze (where one is looking) and the movement of the eyes with respect to time. It has found its way to be a mandatory device in areas like human-computer interaction, neuroscience, psychology, marketing, and assistive technologies. Eye-tracking systems can be of great use by providing insights on cognitive processes, user experience, and decision-making because they capture visual attention and gaze behavior. Throughout the decades, eye-tracking technology has experienced a massive evolution, which started as an intrusive and hardware-intensive system to highly accurate and non-invasive; non-hardware solutions that are software-driven. These improvements have been occasioned by sensor technology, computer vision, and artificial intelligence [9]. Figure (1) shows the evolution of eye-tracking technology over time, with the key milestones in technological advancement in the 1940s being the ones with high attention to date. The figure outlines the process of the shifting of the old method of electrical measurements to the current system, which is powered by AI and is sensor-based, showcasing how the entire system has been getting better and better, to where it is more precise, easy to use, and integrated [10].point

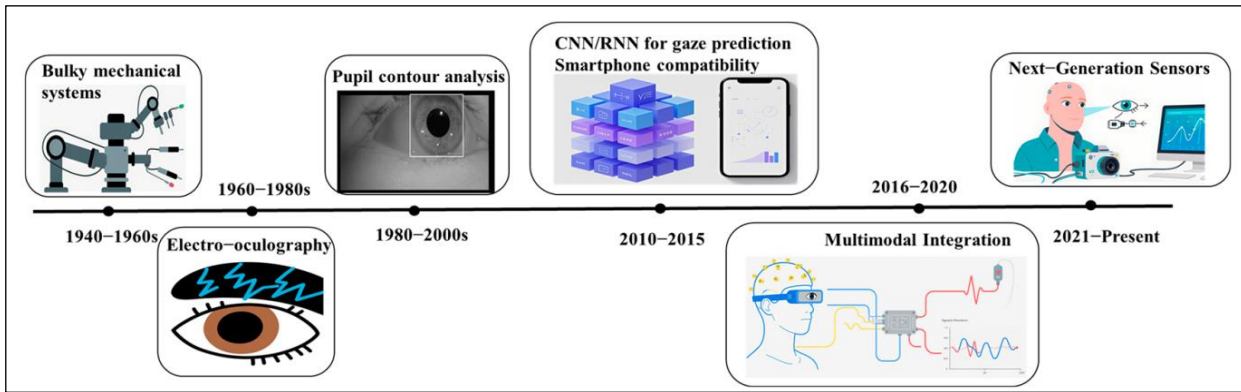


Figure 1: Timeline of Eye-Tracking Technology Evolution
Source: [10].

The history of eye-tracking technology development can be separated into several important stages each of which marked a significant breakthrough in the methodology and usage.

- The first stage (1940s-1960s) was based on the principle of electro-oculography (EOG) where electrical potential of the eye movements are measured with the help of electrodes put around the eyes. Although this technique was able to give fundamental propositions on eye movement, it lacked a high spatial resolution and involved direct physical contact with the subject and was not very useful in widespread practice [11].
- In the 1960s-1980s, the eye-tracking systems were replaced with bulky mechanical systems. These systems tended to be head-mounted systems or hard structures that limited the movements of the user. Despite their ability to increase the measurement abilities, they were inconvenient and could be mostly used in the controlled laboratory settings [12].
- A major improvement was made between the 1980s and 2000s when an analysis of the contour of the pupil was introduced. This method of non-invasive tracking through the use of the camera was made possible since the camera was able to identify and determine the shape and position of the pupil. It is the stage of the modern eye-tracking starting with better accuracy and increased comfort to the user [13].
- The years 2010-2015 were marked by the introduction of the methods of artificial intelligence, especially convolutional and recurrent neural networks. Those approaches increased the accuracy of estimating gaze and allowed tracking of the eye on the consumer devices of smartphones. The technology became easier and larger-scale [14].
- In 2016-2020, multimodal integration, the integration of gaze data with other physiological measures, started to become available in eye-tracking systems, including brain activity and heart rate. This has made it possible to gain a better understanding of human behavior and apply it in virtual reality, healthcare, and more sophisticated user interfaces [15].
- Since 2021, there is an additional development of next-generation sensors that revolutionized the field. New systems are small, efficient in energy use, and can track in real time and with high accuracy. Now they are present in the devices that people use in their daily lives, including the augmented reality headsets, laptops, and wearable technologies, with vast potentials in bringing them into practice [15].

II.1 BASIC CONCEPTS OF EYE TRACKING

Eye tracking is founded on the number of basic concepts that explain the motions of the eyes and the measurement of visual attention. These concepts are important to comprehend the eye-tracking data and utilize it in research and in the real practice. The eye of a human being does not plan through a scene in a smooth way, but it switches between quick motions as well as short periods of rest. These movements demonstrate information collection and processing in the form of visual processing. Eye-tracking systems are used to record such patterns and convert them into meaningful measures that indicate attention, interest and cognitive effort [16]. The fixations, saccades, and gaze points are the main concepts that the eye-tracking analysis is based on. More sophisticated features such as scan paths and areas of interest (AOIs) are also used to assist the researcher to learn more about viewing behavior in complex settings. The combination of the two notions enables qualitative and quantitative study of visual attention [17]. Table 1. Shows Basic Concepts of Eye Tracking.

Table 1: Basic Concepts of Eye Tracking.

Concept	Definition	Description / Importance
Gaze Point	The exact location where a person is looking at a given moment	Represents the fundamental output of eye-tracking systems
Fixation	A period when the eyes remain relatively still on a single point	Indicates focused attention and information processing
Saccade	A rapid movement of the eye between fixations	Helps shift attention from one point to another; no visual processing occurs during it
Scan Path	The sequence of fixations and saccades over time	Shows the overall viewing pattern and behavior
Area of Interest (AOI)	A predefined region in a visual scene selected for analysis	Used to measure attention on specific elements (e.g., buttons, images)
Pupil Dilation	Changes in pupil size	Can indicate cognitive load, emotional response, or lighting conditions

Blink Rate	Frequency of eye blinks	(fatigue and attention levels)
Dwell Time	Total time spent looking at a specific area	Reflects the level of interest or importance of that area
Heatmap	Visual representation of gaze data using color intensity	Highlights areas that attract the most attention
Gaze Plot	A visual diagram showing fixations and saccades in sequence	Useful for analyzing viewing order and navigation patterns

Source: [18].

II.2 GAZE ESTIMATION TECHNIQUES

Gaze point estimation is the calculation process of identifying the point where the observer is gazing. This calculation is carried out in the majority of eye-tracking systems, based on approximating the point of gaze (POG) on a two-dimensional surface, which could be a display monitor or a scene in front of the camera. This estimation is normally done by creating some mapping between the features obtained in the eye and the location of the gaze. In this case, the eye feature coordinates (X_e, Y_e) are transformed into the screen coordinates (X_s, Y_s) by a non-linear regression (polynomial) equation to represent the connection between these variables [19]. This mapping is typically represented as a 2-dimensional poly, in which the horizontal and vertical gaze coordinates are determined in terms of the combinations of powers of the eye feature coordinates. The equations may be expressed as follows [20]:

$$X_s = a_0 + \sum_{p=1}^n \sum_{i=0}^p a_{(i,p)} X_e^{p-i} Y_e^i \quad (1)$$

$$Y_s = a_0 + \sum_{p=1}^n \sum_{i=0}^p b_{(i,p)} X_e^{p-i} Y_e^i \quad (2)$$

Where, (X_e, Y_e) are the coordinates of the extracted eye features, (X_s, Y_s) are the estimated coordinates of the gaze on the screen and $a_{(i,p)}$ and $b_{(i,p)}$ are the coefficients that specify the mapping function. The order of the model, parameter (n), is what determines the complexity and flexibility of the model. A calibration procedure will be necessary to obtain the best values of these coefficients. In this process, the user is requested to concentrate on a number of predefined points that are shown on the screen. The system compares the position of these points with the estimated gaze position to calculate the difference between them and change the coefficients so as to reduce the error in estimation. The reason why this step of calibration is necessary is that it compensates the personal discrepancies among different eye geometry, and enhances the accuracy of the gaze estimation. Although calibration is used, there is still a possibility of error in the estimated gaze position to some extent. The difference in the angle between the optical and visual axis of the human eye is one of the main reasons of this disparity in vision otherwise known as the kappa angle. The line between the center of the cornea and the center of the pupil is referred to as the optical axis whereas the actual direction of body with respect to a target is referred to as the visual axis. Given that the two axes are not aligned, a systematic error is incorporated in the gaze estimation procedure. This bias is different in people and should be countered to allow more accuracy [21]. Moreover, an anatomical makeup of the eye is also significant in eye tracking [22]. The anterior part of the eye has the cornea and the pupil, which refract and regulate the entry of light to the eye respectively. The center of the pupil and the center of the cornea are considered to be the key features of the gaze estimation since it defines the optical axis and forms the foundation of many computation models. These anatomical and geometric relationships are important in understanding how to come up with proper and accurate eye-tracking systems. Figure (2) shows Schematic of the human eyeball, showing key anatomical parameters and reference points for gaze tracking [23].

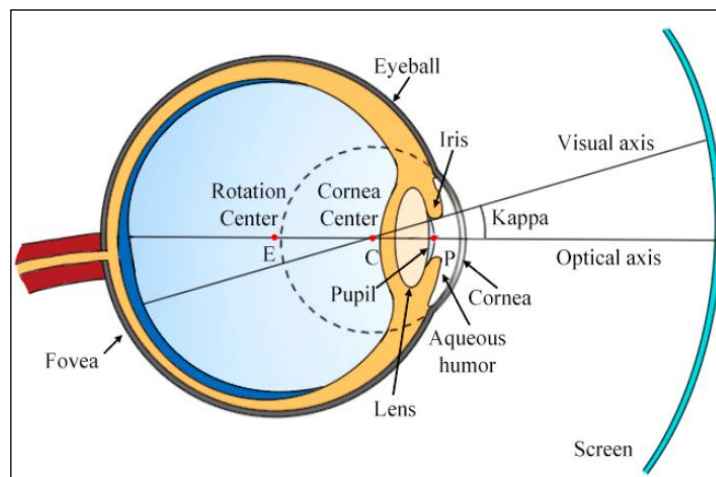


Figure 2: Schematic of Human Eyeball.

Source: Authors, (2026).

III. WEBCAM-BASED EYE TRACKING SYSTEMS

One of the low-cost and most commonly available gaze estimation methods is the eye tracking (Webcam-based) system, which uses the standard cameras, including embedded webcams on laptops or external ones. These systems are based on computer vision and machine learning algorithms to recognize facial features, extract information relating to the eyes, and approximate the direction of gaze of the user in real-time. In contrast to the conventional infrared based eye trackers, webcam solutions can work in the visible light conditions and this presents a challenge in form of sensitivity to light variations, poor image quality, and inconsistency in user placement.

Nevertheless, in spite of these limitations, recent progress in the deep learning and image processing methodologies have greatly improved their work with the examples of their application in remote usability testing, online education, psychological researches, and assistive technologies [24]. Figure (3) illustrates the general workflow of an eye tracking system that uses a webcam to capture the image and then it is scanned by gaze estimation. This starts with camera input, the video frames are captured continuously. These frames are identified by the face detection algorithms to identify the face of the user and approximate the head position, relevant in the context of dealing with natural head expression. After that, the extraction of eye features is made to determine the essential features like the pupil center and the corneal reflections. Such features are the input in the gaze estimation module in which a mapping function is used in estimating the point of gaze of the user on the screen. Lastly, the system does real-time processing to allow generation of gaze predictions with minimum latency hence allowing smoothness and interactive applications [25]. The engineering of this type of systems should take into consideration such aspects as hardware and environmental limitations, such as camera resolution, frame rate, lighting conditions, which directly influence tracking accuracy. The method of personalizing the system by each user is normally done through its calibration to be able to create a correspondence between the eye features that are detected and the known screen coordinates. In addition, real time processing is also necessary to stay responsive, which may demand optimized algorithms and light-weight models. There are tools like GazeFollower which are open-source frameworks, allowing the research and development in this area to be achieved by combining these elements into ready-to-use systems [26].

IV. DEEP LEARNING APPROACHES FOR GAZE ESTIMATION

Deep learning has made a great contribution to the study of gaze estimation by providing more robust and accurate models capable of learning more complicated associations between eye appearance, head pose, and gaze direction. Deep learning methods automatically learn significant features of raw input data, including images of the eyes or face, unlike the traditional model-based or feature-engineering models, and thus are very useful in unconstrained settings. These methods are especially appropriate to the applications based on the webcam and in real world, where the changes in the lighting effect, the head position, and the features of the users are a significant challenge. The most popular architectures in gaze estimation are convolutional neural networks (CNNs) [27]. They have the design to be able to handle image data by learning spatial hierarchies of features using convolutional layers. In gaze tracking, CNNs are normally performed on eye region images or full-face images so as to obtain discriminative features that are linked to gaze direction. Their capacity to achieve fine grained visual patterns e.g. pupil shape, eye contours, etc. makes them very useful in appearance-based gaze estimation tasks [28].

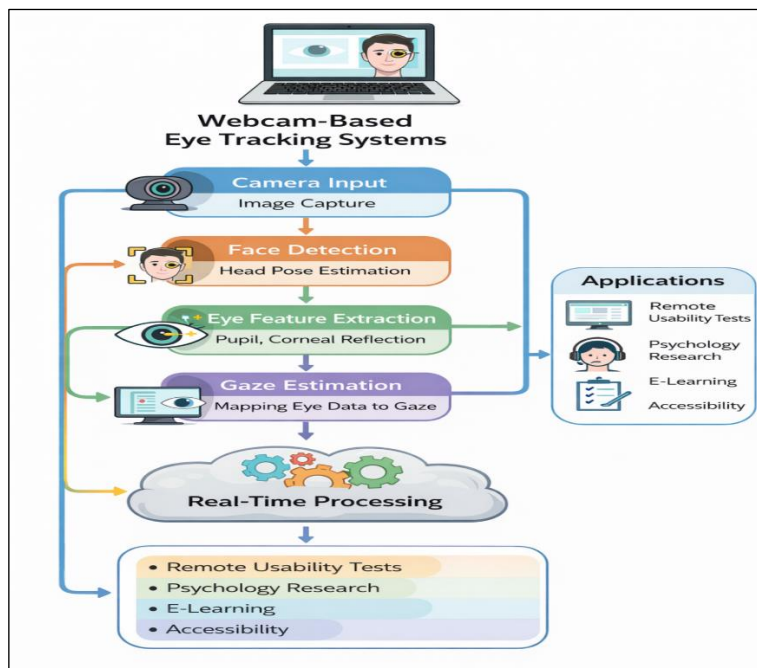


Figure 3: Webcam-Based Eye Tracking Systems.

Source: Authors, (2026).

Transformer-based models have also become a strong rival of CNNs, more recently. These models are based on the focus mechanisms to absorb global connections in the input data so that they can represent long-range relationships better. Gaze estimation Transformers can use the information of other parts of the face or even across multiple frames to be more robust to pose and occlusion changes. They can be extremely demanding in size of datasets and computation power, but have shown optimistic results in complicated cases. Appearance-based gaze estimation is a deep learning representation that directly predicts gaze direction from visual input, including images of the face or eyes, without a geometric representation of the geometric structure of the eye. This mode highly relies on deep neural networks which are able to acquire implicit knowledge of features of gazes. It is especially applicable to low-cost systems, including webcam-based tracking, where accurate geometric measurements are either unreliable or not available [29]. Contrary to that, geometric and hybrid deep models use both learned features and explicit eye geometries. Geometric techniques refer to anatomical and physical models, e.g. the correlation between the pupil and corneal reflections, to determine gaze direction.

Hybrid techniques are methods which combine these geometry constraints with deep learning models to enhance accuracy and generalization. This combination enables the system to combine the two data-driven learning and domain-specific knowledge. Multi-task learning has also found application in gaze estimation, with one model being trained to do several related tasks at the same time e.g. predict gaze direction and head pose. As the head position has a powerful effect on gaze direction, the combined learning of the tasks can enhance the performance. Multi-task models facilitate strong robustness by sharing the representations of various tasks and, therefore, lessening the use of distinct systems, which makes them effective and useful in real-life scenarios [30]. Table (2) shows Deep Learning Approaches for Gaze Estimation

Table 2: Deep Learning Approaches for Gaze Estimation.

Approach	Description	Advantages	Limitations
CNNs	Use convolutional layers to extract spatial features from eye/face images	High accuracy, strong feature extraction, widely used	May struggle with global context and large pose variations
Transformer-Based Models	Use attention mechanisms to model global relationships in data	Captures long-range dependencies, robust to occlusion	High computational cost, requires large datasets
Appearance-Based Methods	Direct mapping from image input to gaze direction	Simple pipeline, suitable for webcams and low-cost systems	Sensitive to lighting and appearance variations
Geometric Models	(pupil, cornea)	Interpretability, grounded in eye anatomy	Requires precise feature detection and calibration
Hybrid Deep Models	Combine deep learning with geometric constraints	Improved accuracy and generalization	Increased system complexity
Multi-Task Learning	Simultaneously predicts gaze and related tasks (e.g., head pose)	Better performance, shared learning, efficient	Requires careful training and balanced datasets

Source: Authors, (2026).

V. CROSS-DOMAIN GENERALIZATION AND ROBUSTNESS

One of the most significant problems of gaze estimation is cross-domain generalization since when a model is trained on a particular set of data or a specific environment it is not always capable of not deteriorating as soon as it is implemented into a different environment. Changes in lighting, quality of the camera, head posture, user attributes, and type of devices may significantly affect the model performance. To eliminate these issues, domain adaptation techniques have been developed, that are more likely to reduce the difference between the training domain (data in the training phase) and the target domain (where it is used). These methods make parameters or features representations of the models to optimize generalization with no requirement of a large quantity of labeled information in the target domain [31]. Contrastive ways of learning have also been used to enhance the robustness of gaze prediction. Contrastive approaches enable models to learn more consummating features of invariance that cross-domain by learning representations that decrease the distance between identical patterns of gaze and intensify the distance between nonidentical patterns. The technique particularly comes in handy in the management of variability that yields in a difference in appearance, light, or camera settings [3]. Variability is another important aspect during the process of enhancing cross-domain robustness. In reality, there can be cases when the lights are dynamic, the poses of the head are mixed, the devices are multiple and the features of the camera are mixed. The mechanisms applied in contemporary gaze estimation systems to ensure they are more resistant to these deviances are data augmentation, adaptive normalization, and multi-view learning. With these strategies included, models would be in a position to perform in a stable manner under unconstrained environments and therefore in the suitability in the real practice [32]. The evaluation measures and benchmark datasets play a crucial role in developing and testing strong gaze estimation models. Publicly available data is a normal system of test cross-domain generalization, but with different subjects, devices and conditions. Some of the measures that are used to gauge performance include Angular error, mean absolute error and accuracy of gaze estimation. This will be done by testing the models in a few datasets, to know that system is not overfit to a particular environment, which can generalize well to unknown conditions [33]. Table (3) shows Cross-Domain Generalization and Robustness Techniques

Table 3: Cross-Domain Generalization and Robustness Techniques.

Aspect	Description	Techniques / Approaches	Challenges Addressed
Domain Adaptation [34]	Aligns feature distributions between source and target domains	Fine-tuning, adversarial adaptation, feature alignment	Cross-dataset variability, unseen environments
Contrastive Learning [35]	Learns invariant features by contrasting similar and dissimilar samples	SimCLR, MoCo-style embeddings	Appearance and lighting variations
Handling Variability [36]	Manages differences in pose, lighting, devices, and user characteristics	Data augmentation, adaptive normalization, multi-view learning	Real-world unconstrained conditions
Benchmark Datasets [37]	Standardized datasets for training and evaluation	GazeCapture, MPIIFaceGaze, RT-GENE, ETH-XGaze	Ensuring fair comparison and model generalization
Evaluation Metrics [38]	Quantitative measures of gaze estimation performance	Angular error, mean absolute error, accuracy	Comparing robustness across models and datasets

Source: Authors, (2026).

VI. COMPARISON OF EXISTING METHODS

A comparison of current gaze estimation techniques will be able to give the information about their strengths, limitations, and appropriateness to various applications.

The accuracy, latency, and robustness are performance measures that are commonly applied to evaluate eye tracking systems. Eye trackers based on special hardware are normally more accurate as infrared lights, specific sensors, and careful calibration are used. Angular errors of such systems are capable of reaching 0.51° , which in scientific research and medical practice are the best as well as in high-precision human-computer interaction. They however, are expensive, are hardware intensive and not portable enough to be used widely [39]. Contrastingly, eye tracking systems using webcams are an inexpensive and readily available option that utilizes ordinary cameras. Though they offer poorer quality of accuracy, usually between $2-5^\circ$ angular error, they perform better with convenience, scaling and non-intrusiveness in those applications where precision is less significant than convenience and scaling. The latency can also be an important factor: although specialized trackers can provide high-speed tracking that can be used in real-time, systems based on webcams need to trade between computational efficiency and tracking quality, particularly when run on an ordinary consumer computer [40]. Cost against accuracy is a significant variable in the choice of a method. Specialized trackers with high end needs a lot of investment but provide accurate results whereas webcam-based needs cheaply but with low fineness of result. These two approaches have their choice determined by the needs of the application, the number of users and the limitations of the environment. Indicatively, research experiments that demand accuracy of gaze tracking can demand special trackers, whereas remote usability tests, e-learning, and accessibility applications can be efficiently executed through the use of webcam-based tracks [8]. In general, it is possible to classify existing methods depending on hardware requirements, complexity of computations, and applicability of the method. The current advances in the field of deep learning and the cross-domain generalization have reduced the gap in performance, now webcam based systems are capable of competitive accuracy in various cases. However, a trade off in webcam-based and specialized system persists as a choice between accuracy, affordability and installation [41]. Table (4) shows Comparison of Eye Tracking Methods.

Table 4: Comparison of Eye Tracking Methods.

Method	Accuracy	Latency	Advantages	Limitations
Specialized Eye Trackers	$0.5^\circ-1^\circ$ angular error	Low latency	High precision, robust to lighting/pose, fast	Expensive, bulky, requires calibration
Webcam-Based Systems	$2^\circ-5^\circ$ angular error	Moderate latency	Affordable, accessible, non-intrusive	Lower precision, sensitive to lighting and pose
Hybrid Approaches	$1^\circ-3^\circ$ angular error	Low-moderate	Combines learned features with geometric models	More complex system design
Deep Learning Only Models	$1.5^\circ-3^\circ$ angular error	Depends on device	Can adapt to multiple domains, flexible	Requires training data, computational cost

Source: [42]

VII. RELATED WORKS

The methods that are used to estimate the head pose could be divided into three major categories, appearance-based, geometry-based, and clustering-based. All these categories are different in terms of the form of input data, the methodology used, and the accuracy. Appearance-based methods make use of visual representations derived on the face or the eye, and apply machine learning or deep architecture models to directly relate such representations to the angle of head orientation. Geometric approach techniques make use of explicit 3D geometric representations of the face and eyes, frequently with landmark points to compute pitch, yaw and roll. The combination of feature extraction and data-driven clustering methods into a single process, known as clustering, is aimed at making them more robust and increase their ability to generalize across data sets [43]. A representative comparison of representative methods, such as type of input, datasets, and accuracy measures are presented in Table (4).

Table 5: Comparison of Works on Head Pose Estimation Methods.

Method Category	Reference	Input Type	Dataset	Reported Accuracy	Details / Notes
Appearance-Based	[44]	Eye images	BiWi	4.66°	Uses eye-region visual features with regression models to estimate pitch and yaw.
	[23]	Full face	AFLW	5.77°	Employs facial landmarks and CNNs for robust head pose estimation in unconstrained settings.
	[45]	Eye images	CMU-PIE	7.36°	Focuses on illumination-invariant feature extraction from eye region.
		Full face	MIT-CBCL	1.31°	Uses controlled indoor images; very low angular error due to stable capture conditions.
		Full face	YaleB	7.58°	Handles varying illumination; includes expression variations in training.
Geometry-Based	[46]	Eye and face	Unknown	2.63°	Computes head pose using 3D facial landmarks and geometric transformations.
		Eye and face	Unknown	3.26°	Combines eye tracking with face geometry for robust 3D orientation estimation.
	[47]	Full face	MPIIGaze	4.8°	Applies explicit facial geometry modeling for gaze and pose estimation.
	[48]	Face landmarks	Unknown	7.65°	Focuses on outdoor and low-resolution images, highlighting method robustness.
	[49]	Full face	ColumbiaGaze	9.46°	Handles multiple subjects with variable head orientations.
		Full face	MPIIFaceGaze	14.93°	Generalization to unconstrained webcam data; larger error due to dataset diversity.

		Full face	GazeCapture	10.46°	Large-scale mobile gaze dataset; demonstrates model performance on consumer devices.
Clustering-Based	[50]	Face + landmarks	BiWi	4.9°	Combines facial feature clustering with regression for improved robustness.
	[51]	Full face	Unknown	1.57°	Uses hierarchical clustering on feature embeddings to reduce noise and outliers.
	[52]	Full face	AFLW2000	3.48°	Enhances accuracy by clustering similar head poses before regression.
		Face landmarks	300 W-LP	3.92°	Incorporates pose normalization and clustering to handle extreme angles.
		Full face	BIWI	2.24°	Robust to subject variability and lighting changes using data-driven grouping.

Source: Authors, (2026).

Table 5 provides a detailed comparison of appearance-based, geometry-based and clustering-based techniques of estimating head pose that indicate a difference in the type of input, datasets used, accuracy and methodological properties. Appearance-based methods are mainly based on visual features that are a result of eye or full-face images and use convolutional neural networks or regression models. Such methods are likely to be effective under controlled settings like the MIT-CBCL where there is fixed lighting and minimal variation of the head describing low angular error conditions. They may however degrade in unconstrained environments because of the variations in pose, lighting and occlusion. Methods that use geometry use explicit 3D face landmark and physical modeling of the head and eye. These methods are able to combine low errors with limited training data by exploiting geometric constraints and are also capable of making interpretable estimates. They can however in most cases need quality landmark detection and might not be as strong to low-resolution images or partial occlusions. The methods based on clustering provide hybrid approach, that is, similar head poses are clustered and then regression is run, which enhances resistance to noise, subject variability, and extreme angles. Such techniques show stable behavior with a variety of datasets and light environments. On the whole, Table 5 indicates that although appearance-based models are versatile and can be easily used, geometry and clustering models allow better resilience and explanatory capabilities, and thus it is necessary to balance computational complexity, data needs, and real-world implementation conditions. Possible external and individual factors that may influence gaze estimation performance greatly include blinking, occlusion, lighting conditions and user differences. This is important in understanding the effect of these factors on accuracy in the design of robust eye-tracking systems. Eye blinking has temporary signal loss and may interfere with feature extraction, and occlusions (that can be created by glasses, hair or hands) may partially obscure the eyes or face. The visibility of eye features is influenced by the variability of light (such as ambient variability or shadow variability) and the unique anatomical characteristics of the eyes (eye shape, pupil size, and facial geometry) can further influence the performance of the system. Table (6) presents the summary of typical studies assessing gaze estimation under these conditions, input types, data sets and reported performance measures, such as average precision (AP) and angular error in degrees.

Table 6: Evaluation of Gaze Estimation under Various Factors.

Factor	Reference	Input Type	Dataset	Performance Metric	Details / Notes
Blinking	[53]	Eye	RT-BENE, Eyeblink8	AP: 0.757 (RT-BENE), 0.997 (Eyeblink8)	Evaluates robustness to natural blinks using eye-focused features.
	[54]	Eye	RT-BENE	AP: 0.653	Shows reduced precision under frequent blink scenarios; highlights importance of temporal smoothing.
Occlusion	[55]	Head / Body	GAFA, MoDiPro	20.4° (GAFA 3D), 25.6° (MoDiPro 2D)	Evaluates partial occlusion due to head rotation and body posture in gaze estimation.
	[56]	Eye + Face	MPIIGaze, EYEDIAP, RT-Genie	3.8°, 5.4°, 7.9°	Combines multiple occlusion scenarios, including glasses and hair; evaluates both 2D and 3D gaze.
Illumination	[57]	Eye	MPIIGaze, EYEDIAP	5.20°, 7.36°	Tests performance under varying lighting conditions; sensitive to shadows and low light.
	[58]	Eye	MPIIGaze, EYEDIAP, UTMultiview	4.67°, 3.36°, 4.33°	Uses deep learning-based approaches for illumination-invariant feature extraction.
Individual Differences	[59]	Eye	EVE, MPIIGaze, XGaze	1.89°, 4.14° / 3.02°, 2.88°	Evaluates cross-subject variability; models trained on some users tested on unseen users.
	[60]	Eye	Custom / unspecified	0.9°	Demonstrates high accuracy for personalized calibration.
	[61]	Eye	Custom / unspecified	4°	Highlights challenges in generalizing gaze estimation across users without calibration.

Source: Authors, (2026).

Table 6 presents the effect of external and individual forces on the estimation of gaze, such as blinking, occlusion, the change in illumination, and variation in inter-subject. Blinking of the eyes causes intermittent loss of signals which could have an immense impact on the accuracy of gaze as seen in data such as RT-BENE and Eyeblink8. The techniques that apply smoothness in the time or predictive modeling are in a position to deal with these transient events which minimizes the decline in performance. Another challenge is occlusion

which results due to head movements, glasses, or hand signs. Researchers have shown that the combination of head and body cues or both eye and face signal is better at enhancing robustness through the use of GAFA and MoDiPro. Correspondingly, the controlled datasets such as MPIIGaze, EYEDIAP and RT-Genie indicate that the multi-modal inputs can be used to reduce the effects of occlusions. Another important factor is the variability of illumination since discrepancies in the illumination of the eye features influence the visibility of face-related features. When trained on augmented or multi-condition datasets, the deep learning approaches demonstrate encouraging results at varying levels of illumination. Lastly, personal anatomy and facial features of the eye and face play a major role in estimating gaze. With cross-subject tests, it is seen that models do not generalize well without subject-specific calibration, as observed in EVE and XGaze datasets. In general, Table 6 has shown that it is necessary to design gaze estimation systems that should consider a combination of temporal modeling, multi-modal inputs, and methods to control the calibration to provide a reliable performance in the real-life conditions. The quality of datasets used in gaze estimation research depends greatly on the quality of participants used (varying), the nature of recording environments (varying) and the adequate sample sizes to generate and test models. These datasets vary according to the type of input (RGB or RGB-D), presence of full-face images, number of subjects, data magnitude, distance of recording and surrounding conditions (illumination and laboratory or natural conditions). Table (7) reflects a comparative summary of widely used gaze estimation datasets, including their essentials and the ability to be used in diverse experimental configurations.

Table 7: Overview of Gaze Estimation Datasets.

Dataset	Input Type	Participants	Full Face Available	Data Amount	Recording Distance	Environment / Illumination
MPIIGaze [62]	RGB	15	No	213,659 images	40–60 cm	Daily life
MPIIFaceGaze [63]	RGB	15	Yes	213,659 images	40–60 cm	Daily life
UnityEyes [64]	RGB	-	No	User-defined	User-defined	User-defined
ColumbiaGaze [30]	RGB	56	Yes	5,880 images	200 cm	Lab
Gaze360 [65]	RGB	238	Yes	~172,000 images	~200 cm	Daily life
ETH-XGaze [66]	RGB	110	Yes	1,083,492 images	100 cm	Lab
UT-Multiview [67]	RGB	50	No	1,216,000 images	60 cm	Lab
EYEDIAP [68]	RGB-D	16	Yes	94 videos	80–120 cm	Lab
RT-GENE [69]	RGB-D	15	Yes	277,286 images	~182 cm	-
NISLGaze [70]	RGB	21	Yes	2,079 videos	90 cm	-
GazeCapture [71]	RGB	1,474	Yes	>2.5 M images	Close	Daily life
TabletGaze [72]	RGB	51	Yes	816 videos	30–50 cm	Lab
GAFA [55]	RGB	-	Yes	882,000 videos	50 cm–7 m	Daily life
GFIE [73]	RGB-D	61	Yes	71,799 videos	1.04–6.48 m	Daily life

Source: Authors, (2026).

Table 7 gives a summary of commonly used gaze estimation datasets, contrasting input type, participant variety, size of data set, recording range and environmental circumstances. The table shows that there is a significant difference in the size and design of datasets, which indicates the versatility of the research in this area. Big datasets like GazeCapture or ETH-XGaze have millions of images or large amounts of video data, and can be used to train deep learning models that have huge demands in terms of diverse input. Smaller datasets, like ColumbiaGaze or NISLGaze, provide controlled conditions, as well as high-quality images, which come in handy especially when it comes to the testing of model performance in lab-based experiments. Another aspect of consideration is the inclusion of full-face pictures as opposed to eye-region pictures only. Head pose integration and multi-task learning methods have full-face datasets, whereas eye-only datasets specialize in finer ocular detail. Also, the availability of RGB-D information in such datasets as EYEDIAP and GFIE makes it easier to predict 3D gaze and consider the depth in modeling. The distance and environmental condition also affect the relevance of datasets, as natural head movement and natural illumination variability are only recorded in daily life datasets, but controlled lighting and constant distance are provided in lab-based datasets to evaluate them exactly. In general, Table 7 highlights the trade-offs between the size of dataset, environmental variability, and input type. The choice of the suitable dataset depends on the purpose of the model, the complexity of the model, and the need to be robust in uncontrolled situations or accurate in specific experiments.

VIII. CONCLUSION

This review gives a detailed discussion on the present situation and problems of gaze and head pose estimation technologies considering not only the methodological progress but also the practical aspect. Appearance-based methods, based largely on the visual cues picked off the eye areas or full-face image, have proved to be very flexible and data-driven to allow models to learn complicated images-to-gaze-direction linkages. Such techniques are especially useful when used with convolutional neural networks (CNNs) or transformer-based models, which are capable of learning spatial hierarchies and long-range dependence of visual information. Conversely, geometry-based techniques use explicit 3D facial landmark models to determine head orientation, which is interpretable and resistant to moderately unconstrained conditions. The clustering-based methods also improve the resilience of the result since the similar head poses are clustered together before regression, which would be better in the high variability situations (where the angle or the appearance of the subject varies a lot). Multi-task and hybrid learning models that combine gaze estimation and head pose prediction have shown that they have the potential to generalize to varying conditions by being able to utilize multiple correlated cues. In spite of these developments, the performance of gaze estimation is very sensitive to both external and individual variables. System accuracy is determined by the blinking, glasses or hair occlusion, inconsistent lighting and the difference in the anatomy among users.

Strong systems thus should include temporal modeling to deal with transient occurrences like eye blinks, multi-modal inputs to reduce occurrence of occlusions and adaptive calibration in consideration of inter-subject variability. To gain generalization to real-world settings, deep learning models that have been trained based on large-scale and diverse datasets are necessary. Smaller, manipulable datasets can be extremely useful in accuracy benchmarking and algorithm innovation validation, but they probably do not reflect the variability found in real-life situations. Gaze tracking systems based on webcams offer an easily accessible and inexpensive alternative to special infrared-based gaze tracking systems and can therefore be used in remote applications like e-learning, usability testing and mobile interfaces. Such systems, however, are usually more concerned with convenience and scalability at the cost of fine-grained accuracy, which requires more complicated modeling approaches to balance out noises, lighting variations, and free head rotations. Generalization strategies such as domain adaptation and contrastive learning are highly important in helping models trained in one setting or environment to be effective in unknown settings. In the future, the direction of research ought to be to create larger and more varied datasets encompassing a diverse set of lighting, pose, and subject variables, create models that are independent of illumination and occlusion, and create lightweight implementations that can be deployed on consumer devices in real-time. The combination of high-quality deep learning solutions with inexpensive and practical hardware will enhance the availability and usability of gaze and head pose estimation and open up new human-computer interaction solutions, as well as enabling real-world implementation of gaze and head pose estimation in fields of educational, medical, and interactive technology.

IX. AUTHOR'S CONTRIBUTION

Conceptualization: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Methodology: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Investigation: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Discussion of results: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Writing – Original Draft: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Writing – Review and Editing: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Resources: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Supervision: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

Approval of the final text: Khalid Sami Yousif, Emad A. Mohammed and Mayyada Al- Hammoshi.

X. REFERENCES

- [1] M. Q. Khan and S. Lee, "Gaze and Eye Tracking: Techniques and Applications in ADAS," *Sensors*, vol. 19, no. 24, 2019.
- [2] G. R. Chhimpa et al., "A Comprehensive Framework for Eye Tracking: Methods, Tools, Applications, and Cross-Platform Evaluation.," *J. Eye Mov. Res.*, vol. 18, no. 5, Sep. 2025.
- [3] L. Huang, Y. Li, X. Wang, H. Wang, A. Bouridane, and A. Chaddad, "Gaze Estimation Approach Using Deep Differential Residual Network," *Sensors*, vol. 22, no. 14, 2022.
- [4] Y. Cheng, H. Wang, Y. Bao, and F. Lu, "Appearance-Based Gaze Estimation With Deep Learning: A Review and Benchmark.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 7509–7528, Dec. 2024.
- [5] S. Saxena, L. Fink, and E. Lange, "Deep learning models for webcam eye tracking in online experiments," *Behav. Res. Methods*, vol. 56, Aug. 2023.
- [6] Z. Hussain Sahito and Z. Hussain, "The Effectiveness of Active Learning Strategies in Enhancing Student Engagement and Academic Performance," *J. Soc. Sci. Rev.*, vol. 5, Mar. 2025.
- [7] Y. Wu, Z. Zhang, F. Aghazadeh, and B. Zheng, "Early Eye Disengagement Is Regulated by Task Complexity and Task Repetition in Visual Tracking Task," *Sensors*, vol. 24, no. 10, 2024.
- [8] T. Kaduk, C. Goeke, H. Finger, and P. König, "Webcam eye tracking close to laboratory standards: Comparing a new webcam-based system and the EyeLink 1000," *Behav. Res. Methods*, vol. 56, Oct. 2023.
- [9] D. Martinez-Marquez, S. Pingali, K. Panuwatwanich, R. A. Stewart, and S. Mohamed, "Application of Eye Tracking Technology in Aviation, Maritime, and Construction Industries: A Systematic Review," *Sensors*, vol. 21, no. 13, 2021.
- [10] L. Lin, Z. Wu, Y. Lu, and Z. Chen, "Recent Progress on Eye-Tracking and Gaze Estimation for AR / VR Applications : A Review," *Electronics*, 2025.
- [11] Y. Shi et al., "Eye tracking and eye expression decoding based on transparent, flexible and ultra-persistent electrostatic interface," *Nat. Commun.*, vol. 14, no. 1, p. 3315, 2023.
- [12] M. Yang, Y. Gao, L. Tang, J. Hou, and B. Hu, "Wearable Eye-Tracking System for Synchronized Multimodal Data Acquisition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 6, pp. 5146–5159, 2024.
- [13] R. Karmi, I. Rahmany, and N. Khelifa, "Gaze estimation using convolutional neural networks," *Signal, Image Video Process.*, vol. 18, no. 1, pp. 389–398, 2024.
- [14] B. Liu, S. W. Lye, and Z. Bin Zakaria, "An integrated framework for eye tracking-assisted task capability recognition of air traffic controllers with machine learning," *Adv. Eng. Informatics*, vol. 62, p. 102784, 2024.
- [15] D. Lohr, S. Aziz, L. Friedman, and O. V. Komogortsev, "GazeBaseVR, a large-scale, longitudinal, binocular eye-tracking dataset collected in virtual reality," *Sci. Data*, vol. 10, no. 1, p. 177, 2023.

- [16] S. R. L. R. and H. C., "Eye Tracking and Its Applications," IARJSET, vol. 8, Aug. 2021.
- [17] R. Hessels, A. Nuthmann, M. Nyström, R. Andersson, D. Niehorster, and I. Hooge, "The fundamentals of eye tracking part 1: The link between theory and research question," Behav. Res. Methods, vol. 57, Dec. 2024.
- [18] I. T. C. Hooge et al., "The fundamentals of eye tracking part 6: Working with areas of interest," Behav. Res. Methods, vol. 58, no. 3, p. 65, Feb. 2026.
- [19] X. Zhang, L. Wang, Y. He, Z. Mou, and Y. Cao, "High-speed eye tracking based on a synchronized imaging mechanism by a dual-ring infrared lighting source," Appl. Opt., vol. 63, no. 16, pp. 4293–4302, Jun. 2024.
- [20] B. R. Severitt, T. C. Kübler, and E. Kasneci, "Testing Different Function Fitting Methods for Mobile Eye-Tracker Calibration," J. Eye Mov. Res., vol. 16, no. 4, pp. 1–17, 2023.
- [21] J. Zhou, G. Li, F. Shi, X. Guo, P. Wan, and M. Wang, "EM-Gaze: eye context correlation and metric learning for gaze estimation," Vis. Comput. Ind. Biomed. Art, vol. 6, no. 1, p. 8, 2023.
- [22] K. Krafa, A. Khosla, and P. Kellnhöfer, "Eye Tracking for Everyone," 2024.
- [23] F. Gao, S. Li, and S. Lu, "How frontal is a face? Quantitative estimation of face pose based on CNN and geometric projection," Neural Comput. Appl., vol. 33, no. 8, pp. 3035–3051, 2021.
- [24] A. Cantero, C. Lebrato, J. Castro, M. Merino Monge, F. Biscarri-Triviño, and J. I. Escudero, "A review on visible-light eye-tracking methods based on a low-cost camera," J. Ambient Intell. Humaniz. Comput., vol. 15, Mar. 2024.
- [25] L. Falch and K. S. Lohan, "Webcam-based gaze estimation for computer screen interaction," Front. Robot. AI, vol. 11, p. 1369566, 2024.
- [26] A. J. Molina-Cantero, C. Lebrato-Vázquez, J. A. Castro-García, M. Merino-Monge, F. Biscarri-Triviño, and J. I. Escudero-Fombuena, "A review on visible-light eye-tracking methods based on a low-cost camera," J. Ambient Intell. Humaniz. Comput., vol. 15, no. 4, pp. 2381–2397, 2024.
- [27] T. Ge, "Deep Learning-Based Appearance-Based Gaze Estimation," J. Comput. Electron. Inf. Manag., vol. 19, pp. 1–12, Dec. 2025.
- [28] A. Bendimered, R. Iguernaissi, M. M. Nawaf, R. Cherif, S. Dubuisson, and D. Merad, "Dual Focus-3D: A Hybrid Deep Learning Approach for Robust 3D Gaze Estimation," Sensors, vol. 25, no. 13, 2025.
- [29] M. A. Pfeffer, J. K. W. Wong, and S. H. Ling, "Trends and Limitations in Transformer-Based BCI Research," Appl. Sci., vol. 15, no. 20, 2025.
- [30] V. Emile Tatinyuy, N. Woguia Auguste Vigny, M. Ngono Joseph, F. L. Aimé, and W. Pountianus Berinyuy, "Eye Gaze Detection Using a Hybrid Multimodal Deep Learning Model for Assistive Technology," Appl. Sci., vol. 16, no. 2, 2026.
- [31] H.-R. Yang, X. Chen, and C.-X. Ren, "A Generalized Label Shift Perspective for Cross-Domain Gaze Estimation," 2025.
- [32] M. F. Ansari, P. Kasprowski, and P. Peer, "Person-Specific Gaze Estimation from Low-Quality Webcam Images," Sensors (Basel), vol. 23, no. 8, Apr. 2023.
- [33] H. Kaur, S. Jindal, and R. Manduchi, "Rethinking Model-Based Gaze Estimation," Proc. ACM Comput. Graph. Interact. Tech., vol. 5, pp. 1–17, May 2022.
- [34] P. Siva Prasad and S. Agniraj, "Cross-Domain Adaptation Techniques for Robust Plant Disease Detection: A DANN-CORAL Hybrid Approach," Int. J. Exp. Res. Rev., vol. 42, pp. 68–84, Aug. 2024.
- [35] Y. Wang et al., "Enhancing Cross-Domain Remote Sensing Scene Classification by Multi-Source Subdomain Distribution Alignment Network," Remote Sens., vol. 17, no. 7, 2025.
- [36] A. G. Khoei, Y. Yu, and R. Feldt, "Domain generalization through meta-learning: a survey," Artif. Intell. Rev., vol. 57, no. 10, p. 285, 2024.
- [37] C. Chen, Y. Yang, M. Liu, Z. Rong, and S. Shu, "Regularized joint self-training: A cross-domain generalization method for image classification," Eng. Appl. Artif. Intell., vol. 134, p. 108707, 2024.
- [38] L. Lin, Z. Wu, Y. Lu, Z. Chen, and W. Guo, "Recent Progress on Eye-Tracking and Gaze Estimation for AR/VR Applications: A Review," Electronics, vol. 14, no. 17, 2025.
- [39] A. Kar and P. Corcoran, "Performance Evaluation Strategies for Eye Gaze Estimation Systems with Quantitative Metrics and Visualizations," Sensors (Basel), vol. 18, no. 9, Sep. 2018.
- [40] H. Zhang and L. Yin, "Robust Camera-Based Eye-Tracking Method Allowing Head Movements and Its Application in User Experience Research," J. Eye Mov. Res., vol. 18, p. 71, Dec. 2025.
- [41] I. T. C. Hooge, G. A. Holleman, N. C. Haukes, and R. S. Hessels, "Gaze tracking accuracy in humans: One eye is sometimes better than two," Behav. Res. Methods, vol. 51, no. 6, pp. 2712–2721, Dec. 2019.
- [42] R. Alexander and S. Martinez-Conde, "Fixational Eye Movements," 2019, pp. 73–115.
- [43] K. Khan, M. Attique, I. Syed, G. Sarwar, M. A. Irfan, and R. U. Khan, "A Unified Framework for Head Pose, Age and Gender Classification through End-to-End Face Segmentation," Entropy, vol. 21, no. 7, 2019.

- [44] H. Liu, C. Zhang, Y. Deng, T. Liu, Z. Zhang, and Y.-F. Li, "Orientation Cues-Aware Facial Relationship Representation for Head Pose Estimation via Transformer," *IEEE Trans. Image Process.*, vol. 32, pp. 6289–6302, 2023.
- [45] H. Zhao, Z. Ding, and Y. Fu, "Pose-Dependent Low-Rank Embedding for Head Pose Estimation," *Proc. AAAI Conf. Artif. Intell.*, vol. 30, no. 1, 2016.
- [46] H. Bao, W. Fang, B. Guo, and J. Wang, "Real-time wide-view eye tracking based on resolving the spatial depth," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 14633–14655, 2019.
- [47] J. Hu, Y. Lu, J. Zhang, J. Xu, and H. Yang, "Monocular free-head gaze tracking method for driving electric sickbed," *Meas. Sci. Technol.*, vol. 34, no. 12, p. 124002, 2023.
- [48] A. Yang, Z. Jin, S. Guo, D. Wu, and L. Chen, "Unconstrained human gaze estimation approach for medium-distance scene based on monocular vision," *Vis. Comput.*, vol. 40, no. 1, pp. 73–85, 2024.
- [49] A. Ruzzi, X. Shi, X. Wang, G. Li, and S. De Mello, "GazeNeRF : 3D-Aware Gaze Redirection with Neural Radiance Fields," *arXiv:2212.04823v2*, 2023.
- [50] C. Bisogni, L. Cascone, M. Nappi, and C. Pero, "IoT-enabled Biometric Security: Enhancing Smart Car Safety with Depth-based Head Pose Estimation," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 20, no. 6, Mar. 2024.
- [51] J. Tian, L. Cong, and H. Qin, "Mixed-Pose Positioning in Smartphone-Based Pedestrian Dead Reckoning Using Hierarchical Clustering," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
- [52] M. Xin, S. Mo, and Y. Lin, *EVA-GCN: Head Pose Estimation Based on Graph Convolutional Networks*. 2021.
- [53] K. Cortacero, T. Fischer, and Y. Demiris, *RT-BENE: A Dataset and Baselines for Real-Time Blink Estimation in Natural Environments*. 2019.
- [54] L. Ding et al., "CLERA: A Unified Model for Joint Cognitive Load and Eye Region Analysis in the Wild," *ACM Trans. Comput. Interact.*, vol. 30, no. 6, Sep. 2023.
- [55] S. Nonaka and S. Nobuhara, *Dynamic 3D Gaze from Afar: Deep Gaze Estimation from Temporal Eye-Head-Body Coordination*. 2022.
- [56] Y. Luo, J. Chen, and J. Chen, "CI-Net: Appearance-Based Gaze Estimation via Cooperative Network," *IEEE Access*, vol. 10, pp. 78739–78746, 2022.
- [57] Y. Cheng, Y. Bao, and F. Lu, *PureGaze: Purifying Gaze Feature for Generalizable Gaze Estimation*. 2021.
- [58] G. Liu, Y. Yu, K. A. F. Mora, and J.-M. Odobez, "A Differential Approach for Gaze Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 1092–1099, 2021.
- [59] J. Bao, B. Liu, and J. Yu, "An Individual-Difference-Aware Model for Cross-Person Gaze Estimation," *IEEE Trans. Image Process.*, vol. 31, pp. 3322–3333, 2022.
- [60] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, 2006.
- [61] D. W. Hansen and A. E. C. Pece, "Eye tracking in the wild," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 155–181, 2005.
- [62] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 162–175, 2019.
- [63] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation," *Nov.* 2016.
- [64] E. Wood, T. Baltrusaitis, L.-P. Morency, P. Robinson, and A. Bulling, *Learning an appearance-based gaze estimator from one million synthesised images*. 2016.
- [65] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik, and A. Torralba, *Gaze360: Physically Unconstrained Gaze Estimation in the Wild*. 2019.
- [66] 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," 2020, pp. 365–381.
- [67] Y. Sugano, Y. Matsushita, and Y. Sato, *Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation*. 2014.
- [68] K. Funes Mora, F. Monay, and J.-M. Odobez, *EYEDIAP: a database for the development and evaluation of gaze estimation algorithms from RGB and RGB-D cameras*. 2014.
- [69] T. Fischer, H. Chang, and Y. Demiris, *RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments*. 2018.
- [70] Z. Chen and B. E. Shi, "Towards High Performance Low Complexity Calibration in Appearance Based Gaze Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1174–1188, 2023.
- [71] K. Krafa, A. Khosla, and P. Kellnhofer, "Eye Tracking for Everyone," *arXiv:1606.05814v1*, 2016.
- [72] Q. Huang, A. Veeraraghavan, and A. Sabharwal, "TabletGaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets," *Mach. Vis. Appl.*, vol. 28, no. 5, pp. 445–461, 2017.
- [73] Z. Hu, Y. Yang, X. Zhai, D. Yang, B. Zhou, and J. Liu, "GFIE : A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments," *Natl. Key Res. De-velopment Proj. China under Grant 2019YFB1310604*, 2019.