


Review

Remote Eye Gaze Tracking Research: A Comparative Evaluation on Past and Recent Progress

Ibrahim Shehi Shehu ^{1,*} , Yafei Wang ¹, Athuman Mohamed Athuman ¹ and Xianping Fu ^{1,2,*}

¹ Information Science and Technology College, Dalian Maritime University, Dalian 116026, China; wangyafei@dmlu.edu.cn (Y.W.); realathu@dmlu.edu.cn (A.M.A.)

² Pengcheng Laboratory, Shenzhen 518055, China

* Correspondence: shehu@dmlu.edu.cn (I.S.S.); fxp@dmlu.edu.cn (X.F.)

Abstract: Several decades of eye related research has shown how valuable eye gaze data are for applications that are essential to human daily life. Eye gaze data in a broad sense has been used in research and systems for eye movements, eye tracking, and eye gaze tracking. Since early 2000, eye gaze tracking systems have emerged as interactive gaze-based systems that could be remotely deployed and operated, known as remote eye gaze tracking (REGT) systems. The drop point of visual attention known as point of gaze (PoG), and the direction of visual attention known as line of sight (LoS), are important tasks of REGT systems. In this paper, we present a comparative evaluation of REGT systems intended for the PoG and LoS estimation tasks regarding past to recent progress. Our literature evaluation presents promising insights on key concepts and changes recorded over time in hardware setup, software process, application, and deployment of REGT systems. In addition, we present current issues in REGT research for future attempts.

Keywords: remote eye gaze tracking; hardware setup; software process; application area



Citation: Shehu, I.S.; Wang, Y.; Athuman, A.M.; Fu, X. Remote Eye Gaze Tracking Research: A Comparative Evaluation on Past and Recent Progress. *Electronics* **2021**, *10*, 3165. <https://doi.org/10.3390/electronics10243165>

Academic Editor: Michael Wehner

Received: 13 November 2021

Accepted: 13 December 2021

Published: 19 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The speed of eye movements, regularity of blinks, lengths of fixations, and patterns of visual search behavior are all significant to how a person is responding to any kind of visual stimulus [1]. This is because our eyes automatically follow what interests or threatens us. The eyes are a vital part of human physiology that have continued to hold the attention of industry and academic researchers over several decades. Eye research and systems have progressed over four distinct periods distinguishable by the type of data they provided and the nature of their intrusiveness. The characteristics of these four periods are shown in Figure 1.

The first period was characterized by the basic study of the eye's structure (see Figure 1a) and the theories of eye movement (see Figure 1b) [2]. The relevance of parts of the eye, including the pupil, cornea, iris, and sclera; and eye movements, such as saccade, fixation, smooth pursuit, and blink, has been studied extensively for REGT systems in [3]. In the second period, eye oculography and tracking emerged, using intrusive instruments that are placed on the human body [4,5]; examples are shown in Figure 1c–f. The progress achieved in hardware processors and image processing techniques paved the way for non-intrusive gaze tracking applications in the third period [6]. These capabilities were explored even further in the fourth period for the remote deployments shown in Figure 1g–k, which were described as remote eye gaze tracking (REGT) systems in [7,8].

The four periods were characterized by specific concepts with similar terminologies, which are: eye oculography, eye tracking, and eye gaze tracking. Eye oculography identifies eye movements. The electro-oculogram (EOG), for example, was described in [9,10] for recording eye movement using electrodes around the eye to measure skin potentials. The electro-retinogram (ERG), which is another neurophysiological means for recording eye movement with a head-mounted eye monitor camera, was described

in [11]. Eye tracking locates the subject’s eye(s) in an image or video, as demonstrated using methods such as electro-encephalography (EEG) in [12], the scleral search coil in [13], psychophysical experimentation with foveal cone-target in [14], and the video-oculogram (VOC) in [15]. Eye gaze tracking estimates the drop point of visual attention known as the point of gaze (PoG) or point of regard (PoR); see early [8] and recent [16,17] research attempts. Eye gaze tracking is also used for estimating the direction of visual attention known as line of sight (LoS); see early [18] and recent [19–21] research attempts. In the literature eye gaze tracking and eye gaze estimation terminologies refer to the same thing. However, eye gaze estimation has been used as a broader term in [22], to refer not only to processing continuous data with time dependencies, but also to refer to static images in order to determine where the subject is gazing at on a device interface or in a real-world scene.

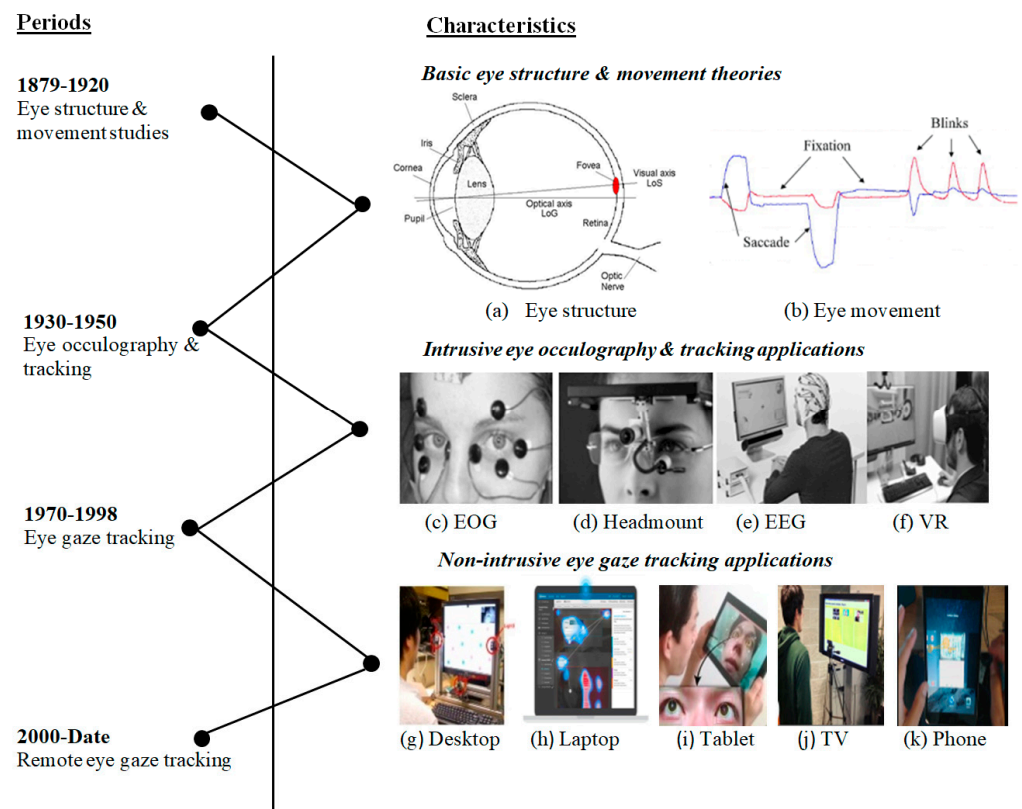


Figure 1. The four periods characterized by specific events.

The operations of REGT systems can be classified into two modules, hardware and software. In Figure 2, the hardware module comprising the display and camera is setup in apposition (i.e., remotely) to the subject in order to feed the required data to the software module for either PoG or LoS estimation.

To achieve PoG (2D) or LoS (3D) estimation, two components of the human physiology are used: the eyeballs and the head pose [23]. Several research attempts at estimating gaze using features of the eye have been reported in [24–28], and for the pose of the human head in [29–32]. To improve the accuracy of gaze estimation further, researchers measured and combined data from eye features and head pose in [33–36], and more recently in [37,38]. This method is popularly used to compensate for the errors caused by the head movements, as further discussed in Section 2.3.

We presented the highlight-based evolution of REGT systems (before/after 2015) in [39]. This paper proposes an extension by:

- Collecting more literature on REGT systems’ evolution (past periods before 2015, and recent periods after 2015).

- Comprehensively comparing the key concepts and changes recorded in the evolutionary periods of REGT's hardware setups, software processes, and applications.
- Presenting current issues in REGT systems' research for future attempts.

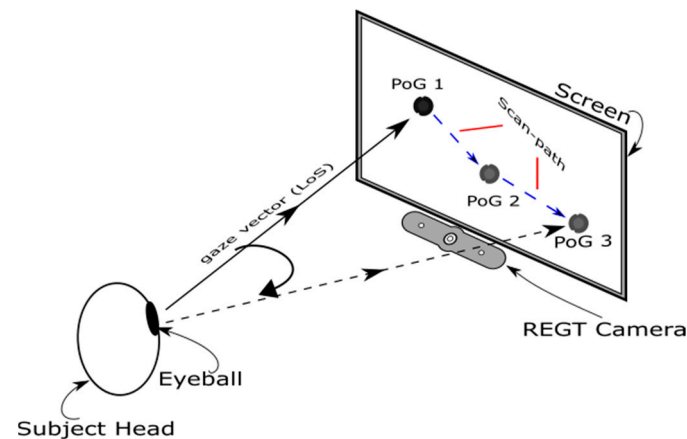


Figure 2. A description of a 3D direction of visual attention defined by LoS, with a corresponding drop point PoG1 determined as the 2D intersection of direction and point of visual attention. The scan-path shows the directional change in visual attention from PoG1 (subject's initial visual attention) to PoG 3.

To the best of our knowledge, this level of detail and coverage is not found in similar published literature on this topic, as earlier review papers only focused on past periods and recent review papers focused on recent periods.

The remainder of this paper is arranged and discussed in modules. Section 2 discusses the various components of REGT systems' hardware modules. Next, popular techniques and algorithms for eye feature detection, extraction, gaze mapping, and datasets for training and validating gaze estimation methods are reported in Section 3. REGT systems as solutions across deployment platforms are discussed in Section 4. In Section 5, our conclusion recapitulates on research changes in REGT systems across the different modules, where we also present current issues for future attempts.

2. Hardware Setup

The physical setup of an REGT system is the hardware module. To achieve the gaze estimation task, this module must provide data for the software module to process. The physical setup is usually made up of various parts consisting of the device interface, which provides the interaction point between the subject and the tracking software; an illuminator which creates patterns of near-infrared on the subject's eyes for the active light methods discussed in Section 2.2; the camera, which captures images of the subject's eyes, patterns of near-infrared for processing by the tracking software; and lastly, the subject whose eye gaze is being tracked.

It is imperative to note that there is no universal standard for setting up the hardware module with regard to the number of cameras, number of illuminators, and subject to camera positions. These factors are usually dependent on the purpose, method, and deployment requirements of the REGT system. More so, hardware requirements for an REGT setup have changed remarkably over time. In Figure 3, the traditional REGTs utilized in the past required more hardware components and sessions to achieve the gaze estimation task as compared to the more modern ones. Modern REGT systems have drastically reduced sessions required for the same task, owing to advancements in hardware technologies and software techniques. These advancements are highlighted and discussed in the subsequent sections describing the four hardware components necessary for an REGT setup.

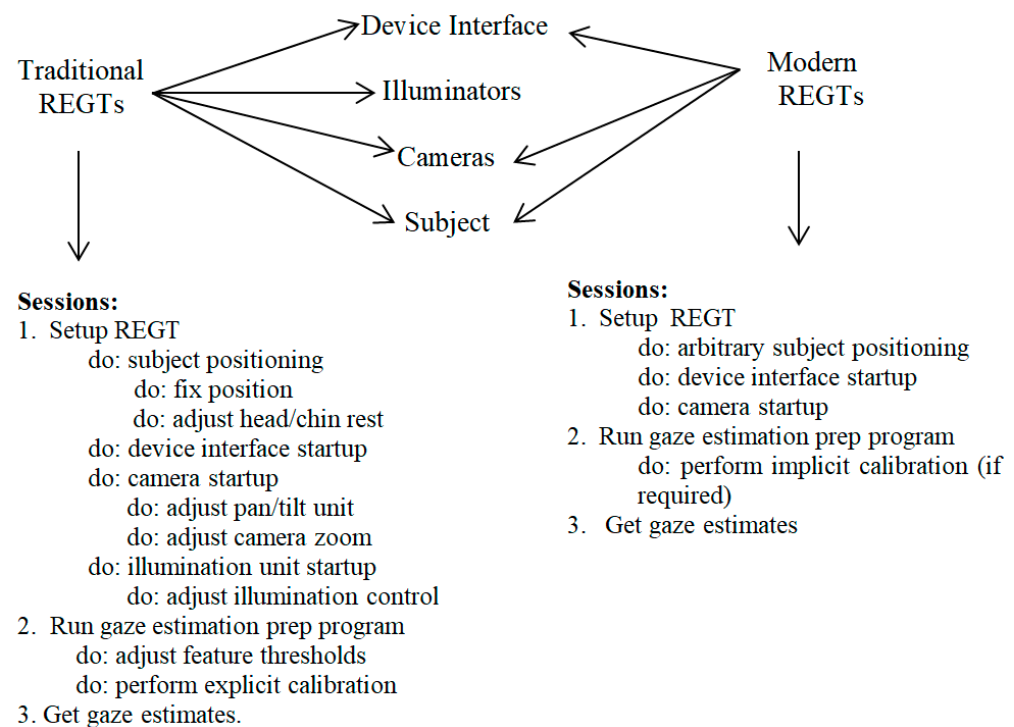


Figure 3. Changes in hardware components and session requirements for REGT over time.

2.1. Interface

In the literature, several interfaces (e.g., phone, tablet, laptop, desktop, and TV) have been utilized as points of interaction between the subject and the tracking software [40]. These device interfaces may operate as standalone (unmodified) gaze trackers or operate with other components (modified) that make up the gaze tracker, as shown in various setups in Figure 1g–k. A modified desktop device with an external light and camera is presented in Figure 1g [41]. An unmodified laptop device embedded with web camera gaze tracker software is presented in Figure 1h [42]. In Figure 1i, an unmodified tablet device is embedded with web camera gaze tracker software [43]. In Figure 1j, a modified TV device has web camera gaze tracker software [44], and Figure 1k presents an unmodified mobile phone device with embedded web camera gaze tracker software [45].

The ideal setup for an REGT system's interface, also known as display or screen, has been of great concern, particularly when there are no standards. In a desk-based device interface, setup orientation in traditional REGT systems was largely determined by the subject's seating or standing position, screen size and positioning, and tracking distance, as illustrated in Figure 4a. In addition, the interface being parallel with the subject's face was essential; great importance was attached to subject's face being within a defined distance from the REGTs camera—e.g., 30 cm [46], 45–60 cm [47], 50–60 cm [48], 60 cm [49], 80 cm [50], and 81 cm [51]. Modern REGTs setup orientations, however, are more concerned about the subject being anywhere within the camera's field of view (FOV), from any arbitrary position away from the REGTs camera [52,53]. FOV is the amount of scene that a particular camera can capture. Wide FOV cameras or pan-tilt mechanisms are famously used in the modern REGT setup to extend the area of the tracking regardless of the subject's position or distance from the REGTs camera, as shown in Figure 4b.

The pan-tilt mechanism is more versatile than the traditional fix in place setup in Figure 4a. The mechanism is commonly based on the camera [54,55]; it gives the tracker the ability to focus on the subject from a range of angles (x , y , and z), without having to take the camera down for focus adjustments.

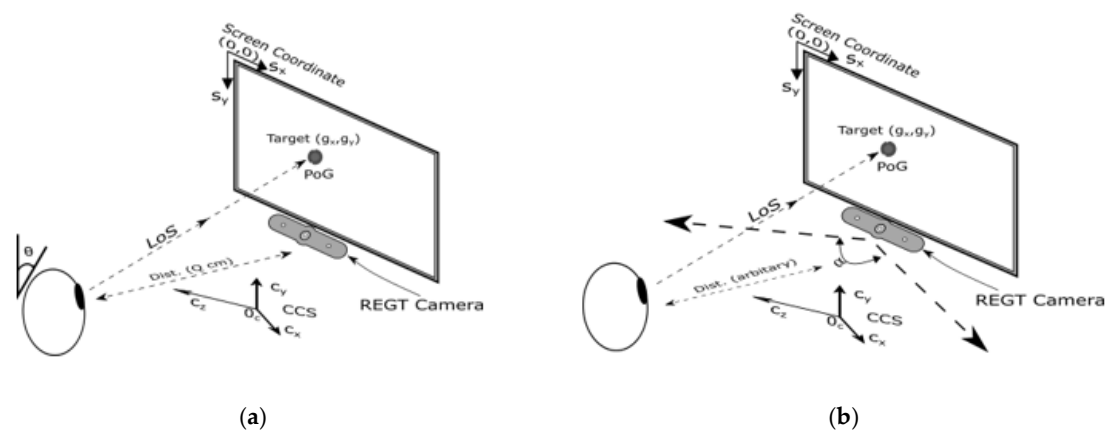


Figure 4. Setup scenarios for desk-based gaze estimation: (a) Traditional setup: shows a defined fixed head position of subject at angle θ , tracking distance in Q cm, screen coordinates (S_x, S_y), camera coordinates (C_x, C_y , and C_z), and the corresponding 2D target point on the screen (g_x, g_y). (b) Modern setup: camera FOV defined by angle α , tracking distance from arbitrary position and distance, camera coordinates (C_x, C_y , and C_z), and the corresponding 2D target point on the screen (g_x, g_y).

2.2. Illumination

The common sources of illumination for REGT systems in the literature are visible and infrared lights [50]. The visible light source referred to as passive light is natural illumination found indoors or outdoors. On the other hand, active light is light from an infrared source. Researchers in [56] demonstrated the use of single near infrared lights (NIR) for a traditional REGT setup. In order to improve facial and eye feature detection for more accurate gaze estimation, researchers further demonstrated the use of two or more NIR in [57–59] and [7,60], respectively. In recent times, the illumination requirements for REGTs have changed because active light methods do not work accurately outdoors (in the real world) due to the uneven lightening properties. They achieve better results under controlled or constant laboratory lightening.

Passive light methods have been explored by researchers [61]. Those authors demonstrated how they tracked the human face in visible light using a real-time video sequence to extract the eye regions for gaze estimation. Both the active and passive light methods have illumination properties that may affect their accuracies for gaze estimation [62]. We describe those properties in Figure 5 together with mitigation strategies.

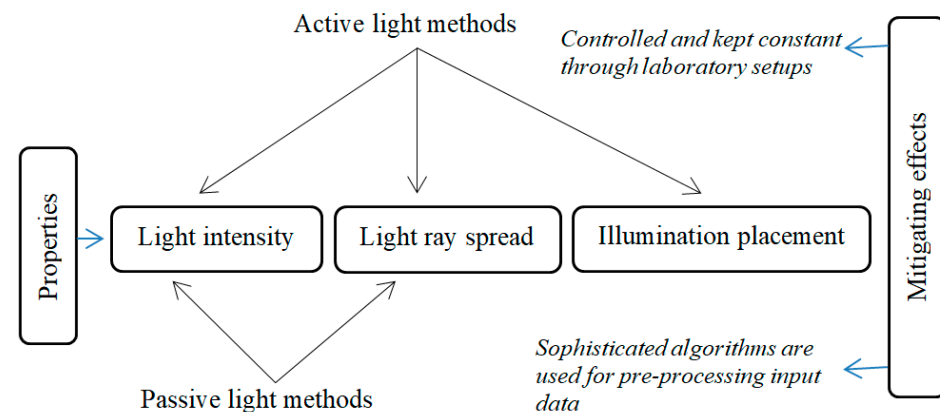


Figure 5. Various illumination properties that affect REGT accuracy and how they are mitigated by different methods.

The illumination properties that affect active light methods affect passive light methods as well. Effects common to both are: the light intensity, which is the amount of light produced by a specific IR source, or outdoor visibility intensity created by the sun; the light ray becomes a problem when it does not spread in a direction that is favorable to the gaze

estimation methods; an illumination source placement is more pertinent to the active light methods, which becomes a problem when the source of illumination is not placed relative to the subject's position.

2.3. Camera

A camera that is capable of providing high quality images and a reasonable trackable area is very crucial for setting up gaze estimation experiments that achieve good results for traditional REGT systems. A wide trackable area provided by wide-angle cameras tolerates the subject's free movements, whereas narrow-angle (zoom) cameras restrict such freedom, but on the other hand, provide more focus on the subject being tracked. Past attempts by researchers, such as [63], have demonstrated the use of two or more-camera systems to address the issue of continuous change in the head position. The use of narrow-angle cameras was demonstrated to allow huge head movement; for instance, Kim et al. [64] used one narrow-angle camera with mirrors. The mirrors rotated to follow head movements in order to keep the eyes within the view of the camera. Similarly, researchers in [51] used narrow-angle cameras supported by pan-tilt to extend the tracking area, and in [65–67] demonstrated the use of a wide-angle camera to estimate a rough location of the eye region and directed another active pan-tilt-zoom camera to focus on the eye region for eye feature extraction.

Selection of a suitable REGT camera will depend on the software process adopted. A typical setup for feature-based methods would include an RGB camera that extracts essential 2D local eye features, such as pupil location, pupil contours, cornea reflection (glints), and eye corner location information [68–70]. If 2D-regression is used to estimate the gaze, this information is enough to facilitate the calibration procedure and gaze mapping function [71,72]. To compensate for the errors caused by the head movements, the model-based methods that involve developing a 3D geometric model of the eye have to be employed. The 3D model-based gaze estimation method uses stereo [73–78] and depth [79,80] cameras that provide additional information, such as the subject's head or face rotation (h_x, h_y, h_z) information and distance from the camera, as described in Figure 6.

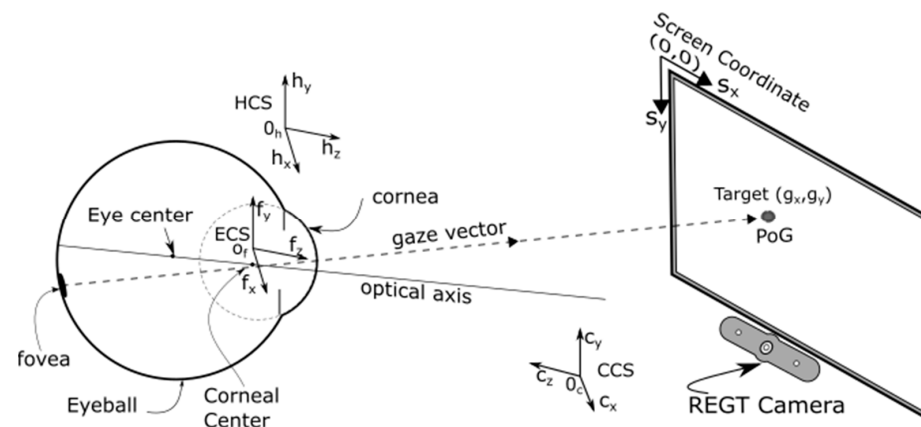


Figure 6. Camera is placed below the horizontal eye line to provide clear view of the eye below or above the tracker's interface, at a position that gives a clear view of the subject's frontal face relative to the source of illumination. The eye images are used to identify reflection patterns (glints), pupil, and other useful features which are then used by algorithms common to traditional REGTs to estimate the gaze vector. The head coordinate system (HCS) provides head information; the eye coordinate system (ECS) provides eye information, and the camera coordinate system (CSS) provides camera view information, where C_z points to the direction viewed by the camera.

Common to both feature-based and model-based methods is the use of NIR cameras. There have been other efforts to estimate PoG and 3D LoS using visible light RGB cameras; we could refer to these attempts as passive light feature or model-based methods. Some

authors in [81] introduced a passive light method that estimates the 3D visual axis using both an eye model and eye features that can work both indoors and outdoors.

Modern REGT based on conventional appearance-based methods requires RGB-D cameras equipped with a set of depth-sensing hardware (such as LiDAR, NIR projectors and detectors) for 3D face model reconstructions [82–85]. This 3D face model provides head pose and facial landmark information that can be used to estimate the 3D LoS. Recent appearance-based methods, in contrast, use images captured by visible light RGB cameras, and then employ machine or deep learning algorithms that directly regress on these images for gaze estimation.

2.4. Subject

The subject whose eye gaze is being tracked is central to the gaze estimation task. The subject's eye orientation is vital data, and in some cases the head pose is required. The trackable area of the REGT system is determined by the correlation between the camera's FOV and the subject's position vertically and horizontally.

In a traditional REGTs setup, the scope is established with a given distance. Figure 7 illustrates that a larger portion of the screen can be tracked if the eye tracker is placed farther away from the subject, and a smaller portion of the screen can be tracked if the eye tracker is placed closer to the subject. That was exemplified in [86] with Tobii's X series eye trackers: the subject was located within 65 cm from the eye tracker; they could do eye tracking at an angle up to 35–36° out from the center of the built-in camera.

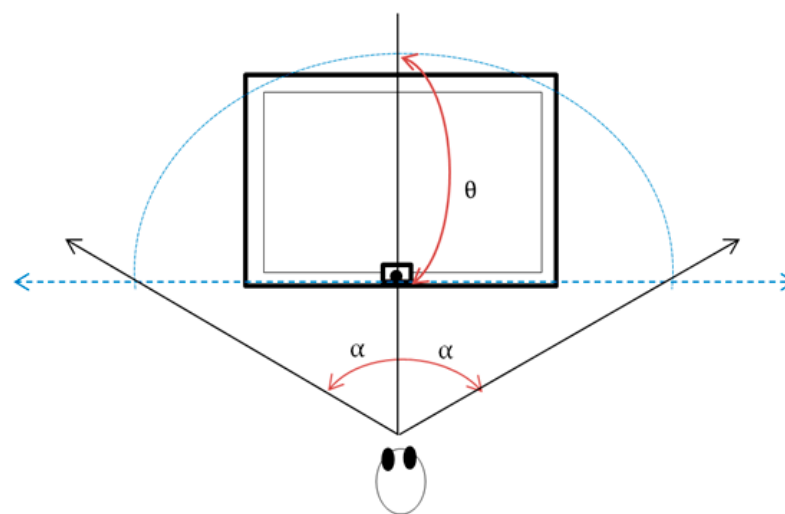


Figure 7. Scope of gaze tracking for vertical trackable area θ and horizontal trackable area α .

3. Software Process

The software module of an REGT system includes algorithms that process data acquired through the hardware module for gaze estimation. In these reviews [68,71,72,87,88], the software process for gaze estimation is commonly classified by the source of light, kind of data acquired from the subject's face, and mapping techniques employed. In the existing literature, this classification is made with variety of naming schemes such as appearance-based versus model-based [52,89,90], appearance-based versus feature-based [88], shape-based versus model-based [91], and appearance-based versus geometry-based [16,87]. It is, however, apparent that some of these schemes represent alternative names for the same schemes: appearance-based was referred to as learning-based in [40,92], and as view-based in [71]; model-based was referred to as geometry-based in [16]. Classification into three broad categories was suggested by researchers in [71,93] to clear up the confusion surrounding these ambiguous naming schemes. We expanded on this in Table 1 to further describe the REGT software methods by mapping techniques, data, and light, along with their merits and demerits.

Table 1. Summary of REGT systems' methods by mapping techniques, data, and light, along with their merits and demerits. Ex. = examples of gaze mapping techniques from the literature.

Methods	Ex. Gaze Mapping Techniques	Data	Light	Merit	Demerit
Feature-based	2D Regression [94–96] Neural Network [36,97] Cross Ratio [24,98,99]	Vectors, anchor points	Active, passive	<ul style="list-style-type: none"> Creates mapping to gaze with relative high accuracy [71]. Discriminative in nature, because it focuses on extracting specific rich features [71,89]. Can be used for active light for PCCR, ICCR, and Purkinje image [28,100] and passive light using PC-EC, IC-EC [61,101]. 	<ul style="list-style-type: none"> Don't work well on low quality images [87]. Generally unsuitable for miniaturized, or calibration free gaze applications [50]
Model-based	2D/3D Geometry Fitting [43,70,102–107]	Eye model	Active, passive	<ul style="list-style-type: none"> Generative in nature, fits a geometry eye model to the eye image [52]. Commonly use the iris contours, eye ball centre to infer gaze based on model parameters [71]. Geometric models can be used for active light [98] and passive light [26]. 	<ul style="list-style-type: none"> Modelling features generatively to achieve fairly good accuracy may require more resources
Appearance-based	K-Nearest Neighbors [93,108], Artificial Neural Networks [109,110], Convolutional Neural Networks [111–114], Support Vector Machines [115,116], Random Forest [16,117,118], Local Linear Interpolation [119–123], Gaussian Process [124,125].	Pixel intensity, texture deference	Passive	<ul style="list-style-type: none"> Regress from eye images to gaze direction using machine learning [16,115] or deep learning [112,114] algorithms. Works well on low quality images captured from web cameras with visible lightning [89]. Widely used for miniaturized, or calibration free gaze applications [50]. 	<ul style="list-style-type: none"> Commonly use low quality images to deduce gaze instead of rich features thus suffer low accuracy [71].

REGTs based on active light methods rely on NIR to create reflection data known as glint (dark or bright), and techniques that find the relationship between these reflections and the screen coordinates [68]. Common examples include Pupil Centre Corneal Reflection (PCCR) [7,24,25,100,126], Iris Centre Corneal Reflection (ICCR) [28], Purkinje image reflection [27,127], and the eye model [98,99]. On the other hand, REGTs based on passive light methods rely on web cameras for visible data and techniques to map data to screen coordinates [68]. Pupil Centre Eye Corner (PC-EC) [128], Iris Centre Eye Corner (IC-EC) [61,101,118], the eye model [26,41,70], Appearance Image Pixel [129–131], and 3D Image Reconstruction [93] are common examples.

Real-time gaze estimation is achieved using any of these methods, when the eyes move and the image of the target object in a world scene or point in a device interface fixates (i.e., settle) on the fovea of the retina. At this point, a high acuity area of vision is achieved with approximately one degree of visual angle [132]. This fixation lasts for 100–1000 ms or 200–500 ms in a few and most situations, respectively, depending on the current cognitive load and quality of information being processed [132,133]. Based on this, REGT's accuracy is commonly determined against one degree (1°) of visual angle spanning approximately 1 cm at 57 cm from the subject's eye.

3.1. Feature-Based Versus Model-Based Methods

Active light feature-based and model-based methods are the foremost NIR methods, and have been demonstrated to be more effective than passive light methods for accurate gaze estimation [8]. In Figure 8, we describe the major components and procedures for feature-based and model-based methods.

3.1.1. Image Acquisition and Pre-Processing

NIR images captured from infrared cameras are required for active light methods, whereas visible RGB images are required for passive light methods. These images can be acquired from static or continuous data. In real-time deployment of REGTs, the use of continuous data is common for obtaining facial images of the subject via a video camera. Raw data from the acquisition process may require some pre-processing before they can be used for gaze estimation. When the images are input, most gaze estimation methods start by applying binarization and normalization to convert them into grayscale and scale them to having a reasonable threshold, resolution, and size. Noise reduction filters such as a Gaussian filter are then applied to compensate for any noise present due to any anomalies in the camera used.

3.1.2. Feature Detection

Feature detection is a popular Computer Vision (C.V) technique for identifying interest points, which closely defines a feature in an image. Interest points can be a corner, ridge, or edge. Traditional CV techniques capable of detecting rich features of interest are commonly used by the feature-based and model-based methods [134,135], such as circular Hough transform (CHT) [136–138], longest line detector (LLD) [139], subpixel detectors [140], and Haar detectors [141,142]. A pupil feature detector, the Haar detector [142], was applied to roughly estimate the pupil region and reduce the overall search space of the eye region algorithms. This is a widely used approach of isolating the pupil blob to assume that it is the darkest element in the image and then applying intensity thresholding to isolate it from the background. Pupil thresholding is applied in [143]. Then k-means clustering algorithms can be applied to the histogram image to obtain pupil and background pixels, respectively, as proposed in [144]. To remove unwanted edges and reflections, and mitigate problems of eye occlusion features, such as eyelashes, a series of open morphological operations are applied [144–148].

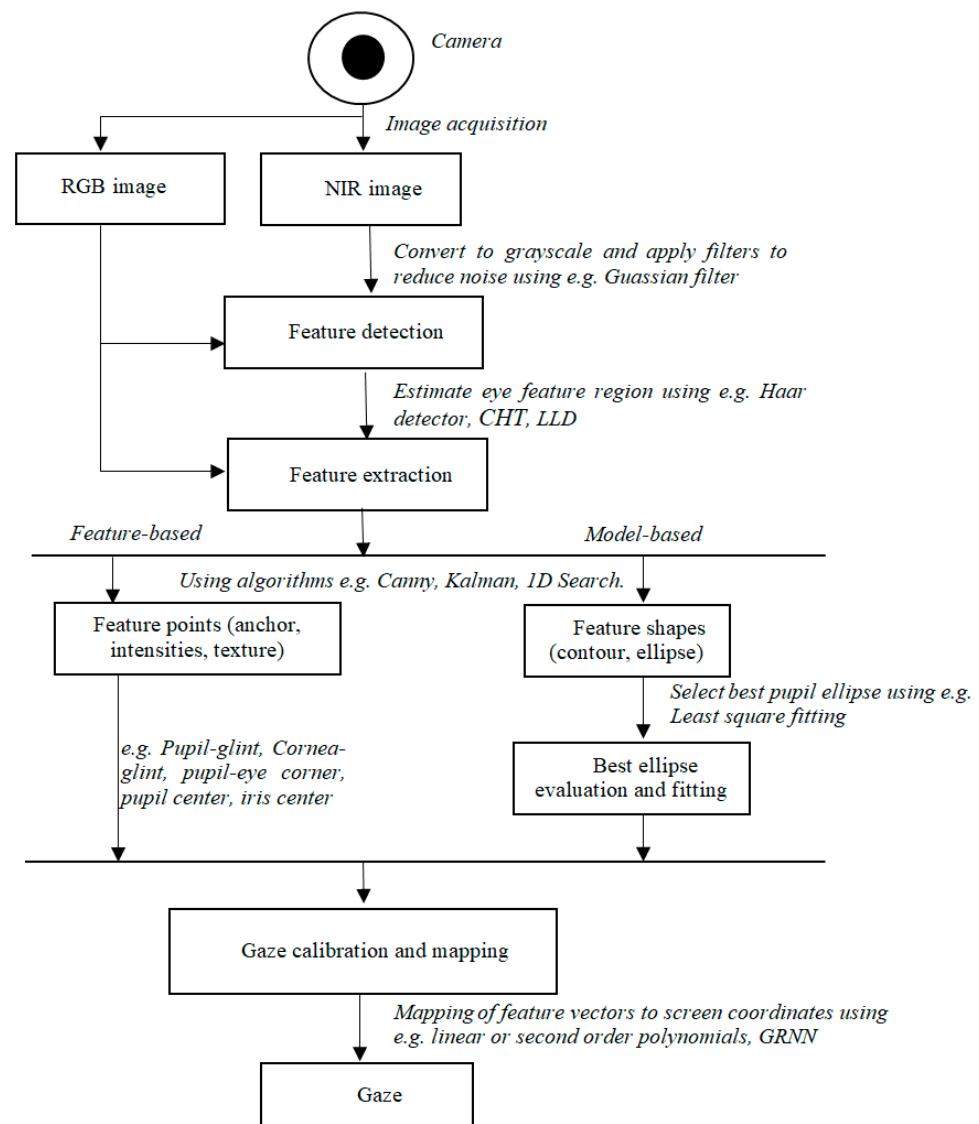


Figure 8. Major components and procedures for the feature-based and model-based gaze estimation methods.

3.1.3. Feature Extraction

Feature extraction for active light feature-based methods uses the feature points or positions. The authors of [7,24,126] extracted intensity points of the pupil and glint area. This happens when an IR light is shone into the subject's eye and a reflection occurs on the surface of the pupil/cornea. The reflection makes a bright spot (known as glint) on the pupil/cornea (see Figure 9a); the position of the glint varies according to the gaze direction. To estimate the point of gaze, Yoo and Chung [7] applied a projective invariant. Assuming the property of the projective space as a relation between the main display screen and the reflected screen, the projective invariant is constant under any projective transform. The projective invariant of the main display screen (Invariant(DScreen_x, DScreen_y)) is equal to that of the reflected screens (Invariant(RScreen₁, Invariant(RScreen₂, Invariant(RScreen₃)). As an alternative, given the detected glint and pupil points in [24,126], calibration procedures such as linear polynomial, second-order polynomials, homography matrix, and interpolation are commonly used to find the relationship between extracted feature parameters (e.g., glint and pupil points) and screen coordinates. These methods are discussed a little further on.

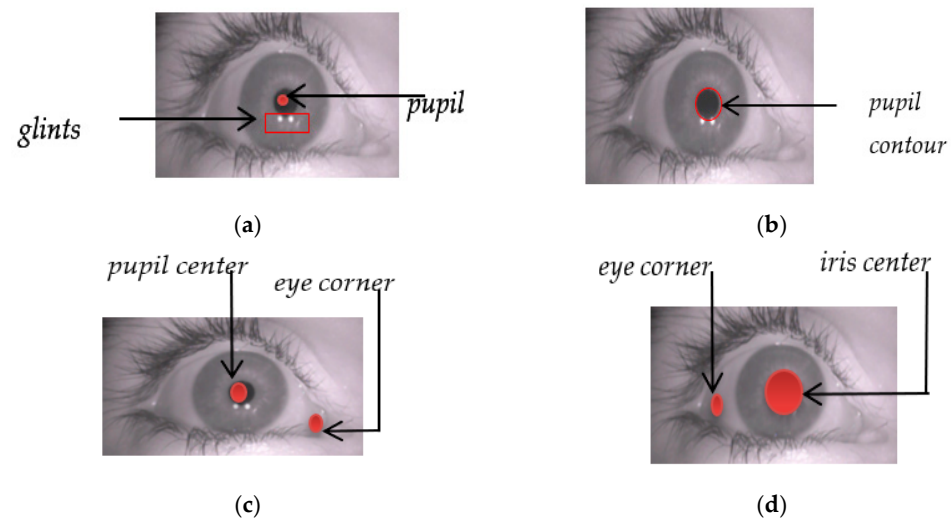


Figure 9. Descriptions of different eye features: (a) pupil and glints, (b) pupil contour, (c) pupil center and eye corner, (d) iris center and eye corner.

The active light model-based method in [98,99] uses edge detection methods (such as a canny edge detector) to get pupil contours (see Figure 9b), but these approaches can be computationally inefficient [145,149]. These pupil contours are extracted and evaluated using the ellipse fitting method, looking for the best candidate for the pupil contour [145,150]. Ellipse evaluation and fitting is the final stage in the pupil detection algorithms. The best method is to point to the exact pupil location by using the ellipse fitting approach. The commonly used method is the ellipse least-squares fitting method [150], but errors made in the pupil feature detection phase can highly influence the results. Instead, the random sample consensus (RANSAC) method is sometimes used, as it is effective in the presence of a large percentage of outliers of the pupil's ellipse feature points [145,151]. When this is done, a calibration procedure to find the relation of the ellipse feature to screen coordinates is performed using homography matrix.

Some examples of passive light featured-based applications in the existing literature can be found in [61,101,128]. The supervised descent method was used in [101] to localize the inner eye corners and perform the convolution of the integer-differential for the eye localization method. In two studies [61,101], the authors extracted the feature points of the iris center and eye corner (see Figure 9d). The pupil center and eye corner (see Figure 9c) were extracted in [128]. In [26,41,70], the authors extracted feature contours using passive light. In [101], the second-order polynomial was applied for the calibration procedure.

3.1.4. Gaze Calibration and Mapping

To map the extracted eye parameters to the screen coordinates, a relationship between locations in the scene image and the eye parameters must be estimated through a calibration procedure. The user-specific parameters are obtained by running calibration routines, such as screen marker calibration and natural features calibration. Using Figure 6 as a reference, some commonly used calibration procedures for both traditional feature-based and model-based REGTs are discussed as follows:

- **Five-point linear polynomial:** The linear polynomial calibration points are the simplest. The method presents a five-point marker on a screen for the subject to look at. By looking at these points and clicking on them, the mapping between screen coordinates and the extracted feature parameters is performed using following equation derived in [152]:

$$\begin{aligned} s_x &= a_0 + a_1 * f_x \\ s_y &= b_0 + b_1 * f_y \end{aligned} \quad (1)$$

where (s_x, s_y) are screen coordinates and (f_x, f_y) are the extracted feature parameters, e.g., pupil-glints vectors. Using the direct least squares method proposed in [150], the unknown coefficients a_0, a_1 and b_0, b_1 can be found during calibration. The problem with this simple linear method is that calibration mapping becomes inaccurate as the subject's head moves away from its original position [8].

- Nine or 25-point second-order polynomial: By fitting higher order polynomials, the second-order polynomial has been shown to increase the accuracy of this system compared to linear ones [8]. A second-order polynomial calibration function was used with a set of nine calibration points in [25,101] and 25 calibration points in [152]. The polynomial is defined as:

$$\begin{aligned} s_x &= a_0 + a_1 * f_x + a_2 * f_y + a_3 * f_x * f_y + a_4 * f_x^2 + a_5 * f_y^2 \\ s_y &= b_0 + b_1 * f_x + b_2 * f_y + b_3 * f_x * f_y + b_4 * f_x^2 + b_5 * f_y^2 \end{aligned} \quad (2)$$

where (s_x, s_y) are screen coordinates and (f_x, f_y) are the extracted feature parameters, e.g., pupil-glints vectors. During calibration, the coefficients $a_0 - a_5$ and $b_0 - b_5$ can be found using the least squares method proposed in [150].

- Homography matrix: Under homography, the calibration routines capture homogeneous coordinates as screen points $s = (s_x, s_y, 1)$, and their corresponding feature points $e = (f_x, f_y, 1)$ are captured (homogeneous coordinates). The transformation of point in 3D from screen points s to the feature points e is given by;

$$s = He \Leftrightarrow \begin{bmatrix} s_x \\ s_y \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} f_x \\ f_y \\ 1 \end{bmatrix} \quad (3)$$

where H is a (3×3) homography matrix, and He is a direct mapping of points in the screen. Once the matrix H is determined, the gaze in the screen can be estimated.

- Interpolation: The authors of [140] had the subject look at several points on a screen to record the corresponding eye feature points and positions. These points served as the calibration points. Then they computed the gaze coordinates by interpolation (a 2D linear mapping from the eye feature to the gaze on screen). The details of this mapping function are as follows:

$$s = s_1 + \frac{fx - fx_1}{fx_2 - fx_1}(s_2 - s_1), \quad f = f_1 + \frac{fy - fy_1}{fy_2 - fy_1}(f_2 - f_1) \quad (4)$$

where (f_x, f_y) are the eye feature vectors; the calibration points p_1 and p_2 for screen coordinates and eye features are $((s_1, f_1), (fx_1, fy_1))$ and $((s_2, f_2), (fx_2, fy_2))$, respectively.

3.1.5. Calibration Error Calculation

The averaged mapping error for the calibration methods we have described can be calculated by the following equations derived in [152]. Firstly, the mapping error of individual calibration point (i_{error}) is computed as:

$$i_{error} = \sqrt{(s_x - f_x)^2 + (s_y - f_y)^2} \quad (5)$$

where (s_x, s_y) are the actual screen coordinates, and (f_x, f_y) are the feature vectors. Further, the average calibration mapping error can be computed as:

$$calib_{error} = \frac{\sum_{i=1}^n i_{error}}{n} \quad (6)$$

where $calib_{error}$ is the calibration technique mapping error, and n is the number of calibration points.

3.2. Appearance-Based Methods

Old reflection-based methods face challenges when applied in real-world settings because they require dedicated NIR devices. These methods are majorly subject to eye occlusion, and other reflection irregularities, such as pupil dilation [19,153,154]. To solve these bottlenecks, recent attempts by researchers have been based on techniques which focus on image appearance to extract the characteristics of the entire eye region instead of specific features that require a dedicated device [155]. As described in Figure 10, the appearance-based methods are inspired by machine and deep learning algorithms and depend on the features learned.

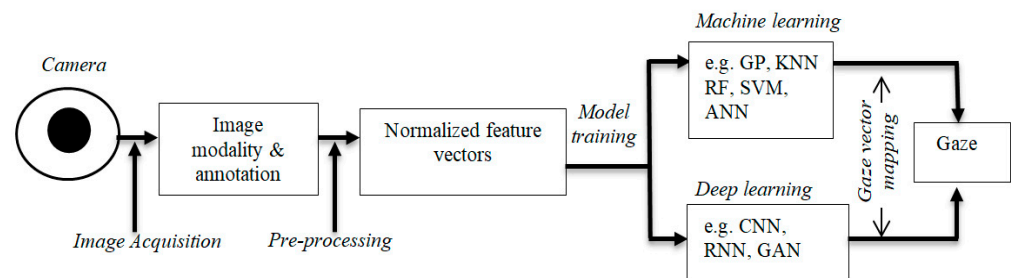


Figure 10. Major components and procedures for the appearance-based gaze estimation.

3.2.1. Image Acquisition and Pre-Processing

Regardless of the gaze estimation method, similar procedures for image acquisition and pre-processing are commonly used. The feature detection and extraction procedure by appearance-based methods is most often treated as a part of pre-processing. This procedure commonly uses modern C.V techniques based on machine or deep learning for extracting features from visible RGB or depth images. These images are obtained of a subject gazing at a known location on the screen to generate gaze coordinates as training data, which are used to train models that make predictions on the new data.

3.2.2. Model Training

The learning procedure for appearance-based methods depends on the characteristics (i.e., annotation) of the training data, such as the gaze-vector angles and the 3D location of the eye in the camera coordinate system, and the 2D coordinates of the gaze point on the screen in the screen coordinate system. The supervised appearance-based methods described in [53,155,156] rely on appropriately labelled or ground truth gaze data. This approach, however, is expensive and time consuming. Unsupervised appearance-based methods in [157–159] have demonstrated the effectiveness of learning on unlabeled data. Both procedures learn the mapping from a large set of data, and generalize this mapping to other subjects via training. Specifically, the task of feature extraction and feature vector to gaze point mapping have been demonstrated using common machine learning algorithms, such as the genetic algorithms (GA) [160], the Bayesian classifier [161], the Kalman and adaptive thresholding algorithm [162], AdaBoost [163,164], k-nearest neighbor (KNN) [93,108,117,130], adaptive linear regression (ALR) [37], random forest (RF) regression [16,117,118], gaussian process (GP) regression [124,125], linear ridge regression (LRR) [165], support vector machines (SVM) [36,115,116,166], artificial neural networks (ANNs) [109,110,167,168], and generalized regression neural networks (GRNNs) [36]. Instances of deep learning algorithms include region-based convolutional neural networks (RCNNs) [164,169,170] and You Only Look Once (YOLO) [171,172], convolutional neural networks (CNNs) [17,38,53,173–175], recurrent neural networks (RNNs) [20,176,177], and generative adversarial networks (GANs) [178–182]. Typically, using these learning-based algorithms for mapping requires a training period or implicit calibration [183]. The subject is asked to look at a number of predefined points on the screen while their gaze and eye locations are estimated and recorded for each of the points. The model is trained on the recorded points, depending on the size of training sample; the model is expected to

implicitly learn the mapping between the camera coordinate system and screen coordinate system.

The training complexity of machine and deep learning algorithms used for appearance-based gaze estimation in the literature is between $O(n^3)$ and $O(\log(n))$. ALR, SVM, and GP regression deal with sparse collections of training samples. These algorithms are not fit for operation on large datasets because their complexities can reach $O(n^3)$ for datasets with sample size n . RF and KNN are friendly to large datasets and could computationally train rapidly on such datasets at more moderate complexities, $O(n^2)$ and $O(n)$. An ANN outperformed other baseline machine learning methods for appearance-based gaze estimation, such as RF and KNN, for a large scale complex problem ($O(n)$, $O(\log(n))$). It is able to perform in complex computations involving much data, making it possible for researchers to explore deeper attributes of data—now popularized as deep learning (DL). Recent DL models utilized for gaze estimation include CNNs, RNNs, and GANs.

1. Convolutional Neural Networks

CNNs, also known as ConvNets, use perceptrons to analyze data [184]. Typical components of a CNN include the input and output layers, and various hidden layers. These hidden layers include convolutional layers, which detect patterns from images, a pooling layer, which reduces number of parameters and amount of computation in the network to control overfitting, and a fully connected (FC) layer, which gives the output. CNNs' successful use in image processing tasks, particularly 2D image classification, has inspired its use for gaze estimation in the recent literature [155,185]; see Figure 11.

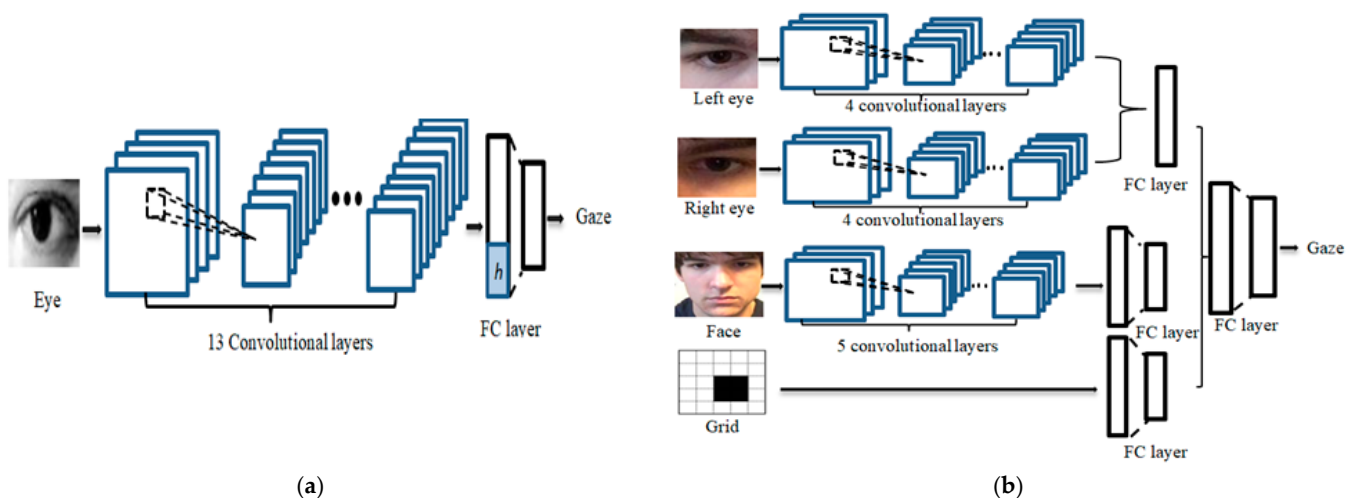


Figure 11. CNN frameworks for gaze estimation: (a) GazeNet framework based on a single-input, single-region CNN [38]; (b) Itracker framework based on a multiple-input, multi-region CNN [17].

Features are directly extracted from input data. These data include images of the eye, face, and others (described in Table 2) that need to be fed into the CNN. Using some mathematical functions, results are passed between successive layers [184,186]. A single-region CNN processes its inputs through a single network, as shown in Figure 11a, whereas a multi-region CNN processes its inputs in multiple separate networks for increased efficiency of the general network, as shown in Figure 11b. The outputs of a CNN for gaze estimation are classification results, i.e., discrete values suitable for gaze zone classification; or regression results, i.e., continuous values suitable for estimating more specific gaze angles and points of gaze.

One of the early attempts that demonstrated the use of CNNs for gaze estimation was reported in [53]. Their attempt was based on a LeNet inspired by MnistNet architecture [187] consisting of two convolutional layers, two connecting maxpooling layers, and a fully connected layer. To predict gaze angle vectors, linear regression training is performed on top of the fully connected layer. Motivated by this progress, the authors improved the

idea and proposed the GazeNet framework, the first deep appearance-based gaze estimation method based on a 13 convolutional layer VGG Network, in [38]. They combined the data from face pose and eye region and injected head angle into the first fully connected layer, as described in Figure 11a, and then trained a regression model on the output layer that predicted with angular accuracy of 10.8° for cross-dataset evaluation. Comparatively, Itracker [17] was based on AlexNet architecture. It uses a multi-region network with various inputs to provide more valuable information than using eye images alone and achieved a prediction error for a pixel distance of between 1.71 and 2.53 cm without calibration. Since then, several other works have applied CNNs and other deep learning models for gaze estimation using different architectures and structures, as summarized in Table 2.

Table 2. Recent (2015–2021) work on CNNs and other DL models for gaze estimation, presented by network type.

Deep Network Classification	Literature	Year	Input	Network Description	Output
Single-region CNN	[114]	2017	Full face	<ul style="list-style-type: none"> Spatial weighted CNN 	Point of gaze (2D)
	[17]	2016	Right & Left eye, face, and face grid	<ul style="list-style-type: none"> Four-region CNN model AlexNet backbone Dark knowledge method 	
Multi-region CNN	[188]	2017	Head pose, and eye	<ul style="list-style-type: none"> Two-region CNN model AlexNet backbone Gaze transform method 	
	[189]	2020	Right & Left eye, full face, and face depth	<ul style="list-style-type: none"> Two-region CNN model ResNet-18 backbone Facial landmarks global optimization 	
Single-region CNN	[53]	2015	Double eye, and head pose	<ul style="list-style-type: none"> LeNet backbone 	
	[38]	2019	Double eye, and head pose	<ul style="list-style-type: none"> VGG backbone 	
	[190]	2020	Full face	<ul style="list-style-type: none"> ResNet-50 backbone 	
Multi-region CNN	[191]	2016	Right & Left eye	<ul style="list-style-type: none"> Two-region CNN model Modified Viola-Jones algorithm 	
	[111]	2018	Right & Left eye, and head pose	<ul style="list-style-type: none"> Four-region CNN model Asymmetric Regression (AR-Net), and Evaluation Network (E-Net). 	
	[112]	2018	Right & Left eye, and face	<ul style="list-style-type: none"> Three-region CNN model Semantic image inpainting Net, landmark detection deep Net, and head pose estimation Net 	
	[173]	2019	Right & Left eye	<ul style="list-style-type: none"> Two-region CNN model Based on RT-GENE [51], and blink detection Net 	
CNN with RNN fusion	[177]	2018	Full-face, eye-region, facial landmarks	<ul style="list-style-type: none"> Two-region CNN for static feature extraction. VGG-16 backbone, and ADAM optimizer Many-to-one recurrent Net for temporal feature extraction and final prediction. 	
	[176]	2019	Left & Right eye, and face	<ul style="list-style-type: none"> Two-region CNN for static feature extraction. AlexNet backbone Many-to-one bi-LSTM for temporal feature extraction and final prediction. 	
	[20]	2019	Full face	<ul style="list-style-type: none"> Multi-frame bidirectional LSTM for temporal features and final prediction. ResNet-18 backbone 	
	[192]	2020	Right & Left eye	<ul style="list-style-type: none"> ResNet-18 architecture backbone GRU Cell for temporal model 	

Table 2. Cont.

Deep Network Classification	Literature	Year	Input	Network Description	Output
GAN	[193]	2020	Full face	<ul style="list-style-type: none"> Self-Transforming GAN (ST-ED) for gaze and head redirection 	
GAN	[194]	2021	Eye region, head pose	<ul style="list-style-type: none"> Multi-task conditional GAN (cGAN) for gaze redirection 	Gaze angle (3D)
CNN with GAN fusion	[195]	2017	Eye	<ul style="list-style-type: none"> SimGAN to improve realism of synthetic images 	
	[178,179]	2020	Eye	<ul style="list-style-type: none"> ResNet backbone EnlightenGAN to improve to lightening condition of dark input images. 	

The accuracy of gaze estimation methods largely depends on the effectiveness of feature detection and extraction [196,197]. Appearance-based gaze estimation using deep learning has recently demonstrated robustness against image input noise, blur, and localization errors [191,198]. Several works have primarily focused on appearance-based methods using CNN within the past five years. A few are described in Table 2, and others are reported in [52,68,89,155]. In addition, the recently published results shown in Table 3 suggest the appearance-based methods using CNNs work better with unconstrained setups than feature-based and model-based methods [38,52]. However, a feature-based or model-based method with constrained illumination (i.e., NIR) still achieves better accuracy for the gaze estimation than an appearance-based method under visible light [199]. For this reason, researchers have intensified efforts to improve the accuracy of unconstrained appearance-based gaze estimation [38,114], hence the exploration of other deep learning frameworks, such as RNNs and GANs.

2. Recurrent Neural Network

RNNs use long short-term memory (LSTM) to process temporal information in video data. The use of temporal (i.e., sequential) features was demonstrated in [20,176,177,192] to be effective at improving gaze estimation accuracy. This is because the video data contain more valuable information than image data. A framework in which fusion information from static features is obtained from the images and sequential features are obtained from video, is described in Figure 12.

The authors of [177] used a multi-region CNN to extract and process static features from the face, eye region, and face landmarks to estimate gaze. They then fed these learned features of all the frames in a sequence to an RNN that predicted the 3D gaze vector of the last frame. They achieved 14.6% superiority over the state-of-the-art method [53] on EyeDiap dataset using static feature only, adding that the temporal feature further improved the performance by 4%. Similarly, the authors of [176] enhanced the itracker network proposed in [17]. They removed face-grid, thereby reducing one network branch from the original, and then used a static feature from a concatenate of the two-eye region images to predict gaze. They further employed the bidirectional LSTM (bi-LSTM) to fit the temporal feature between frames to estimate the gaze vector for a video sequence. They achieved improved performance with the enhance itracker network—11.6% over the state-of-the-art methods, such as those in [17,114,199], on MPIIGaze dataset for static features, and a further 3% improvement with the bi-LSTM for temporal features on the EyeDiap dataset. Both works achieved better estimation accuracy by combining a static network and temporal network instead of only using a static network for gaze estimation.

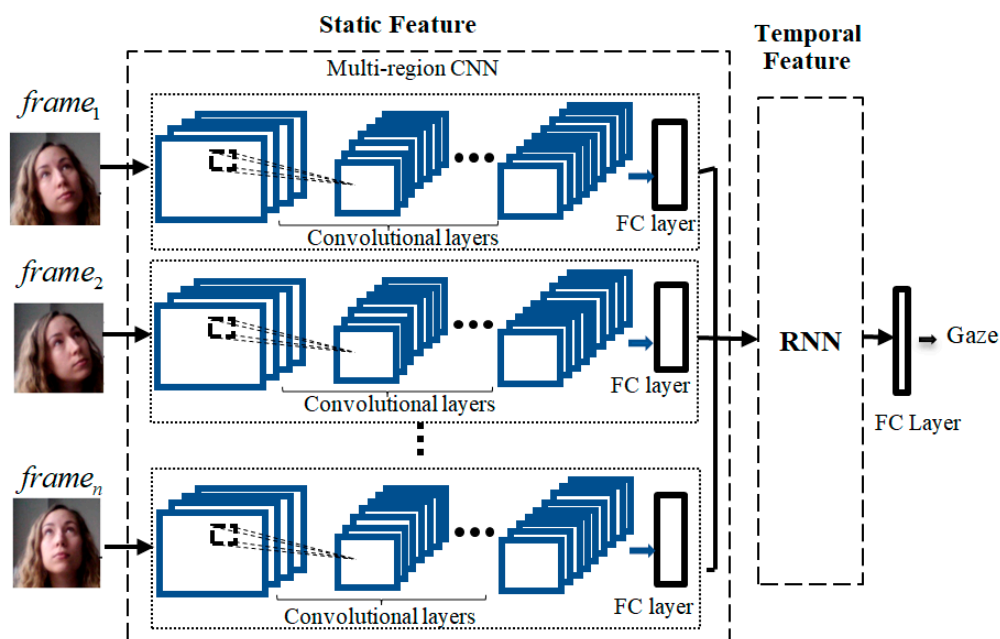


Figure 12. A CNN with RNN fusion for gaze estimation [177].

3. Generative Adversarial Networks

GANs are effective for data generation. The basic theory of the GAN and its generative models has been discussed in [200]. The use of GANs for gaze estimation was demonstrated recently in [178,179,193–195]. In [193], the authors used adversarial learning to manipulate the gaze of a given face image with respect to a desired direction. Similarly, the authors of [194] adopted the use of flow learning and adversarial learning; the network is described in Figure 13a.

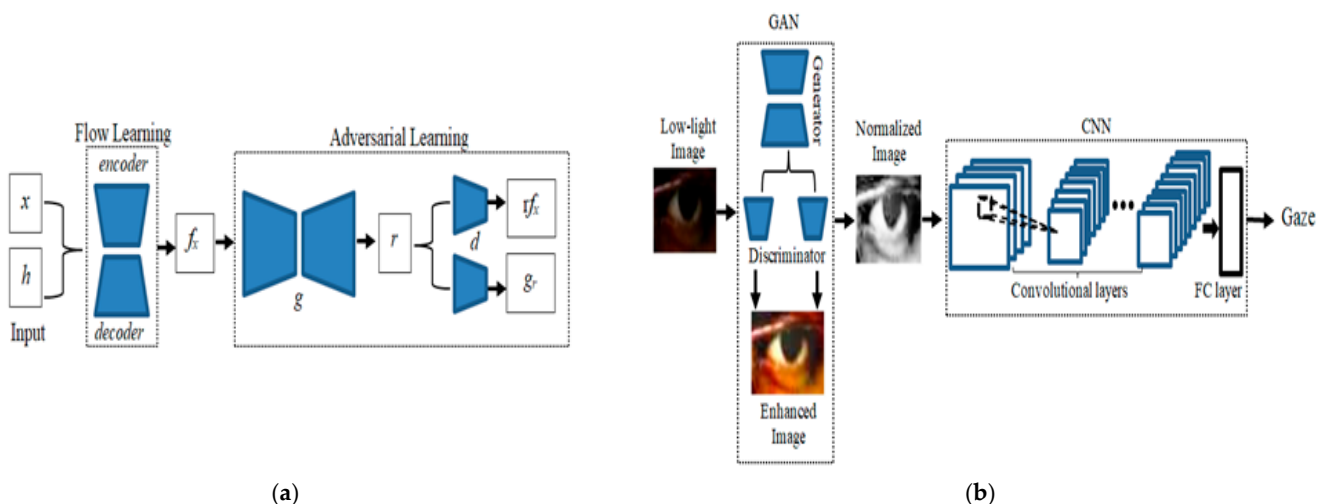


Figure 13. (a) GAN for gaze redirection [194]; (b) CNN with GAN fusion for gaze estimation [178].

Eye region x and the head pose h are taken in as encoder inputs. The decoder outputs fine-grained image f_x ; the generator g outputs the residual image r , which is added to f_x . The refined results r_f_x and the ground truth g_r are fed to the discriminator d . The discriminator network with gaze regression learning ensures that the refined results and the ground truth have the same distribution and the same gaze angles. Gaze redirection error achieved was about 5.15° .

The authors of [178] used a fusion of a GAN and a CNN for recovering the missing information of images captured under low-light conditions using the framework described

in Figure 13b. They first used the GAN to recover (i.e., separating the noise) near to original eye images from the low-light images, and then fed the recovered images into the CNN network proposed in [38] to estimate the gaze. Their experimental results on the GAN enhanced image dataset demonstrated an improved performance compared to GazeNet by 6.6% under various low light conditions.

3.3. Evaluation and Performance Metrics for REGTs

The performance of REGT is usually described by using three terms, which are: accuracy, robustness, and stability [22]. All three terms are important considerations for well performing REGT systems, but the accuracies and precisions of REGT systems hold greater implications for real applications [201].

3.3.1. Precision Evaluation of REGT Systems

Precision is the ability of the REGT system to reliably predict or reproduce relative gaze positions. It measures the variations in the recorded gaze positions using a confusion matrix. Precision $m_{precision}$ can be computed for classification problems as:

$$m_{precision} = \frac{(TP_{b1} + TP_{b2} + \dots + TP_{bn})}{((TP_{b1} + TP_{b2} + \dots + TP_{bn}) + (FP_{b1} + FP_{b2} + \dots + FP_{bn}))} \quad (7)$$

where TP is the true positive predictions; FP is the false positive predictions of successive samples b_1 – b_n as described in Figure 14b. The regression problem as the root mean square error a_{error} of successive samples b_1 – b_n , derived in [62,202], is:

$$a_{error} = \sqrt{\frac{1}{n} \sum_{i=1}^n \alpha_i^2} = \sqrt{\frac{\alpha_1^2 + \alpha_2^2 + \dots + \alpha_n^2}{n}} \quad (8)$$

where α is the visual angle in degrees and n is the number of recorded samples in the dataset. α_1 – α_n is the successive visual angle between samples b_1 – b_n . Precision is calculated for each eye individually and as a mean of both eyes.

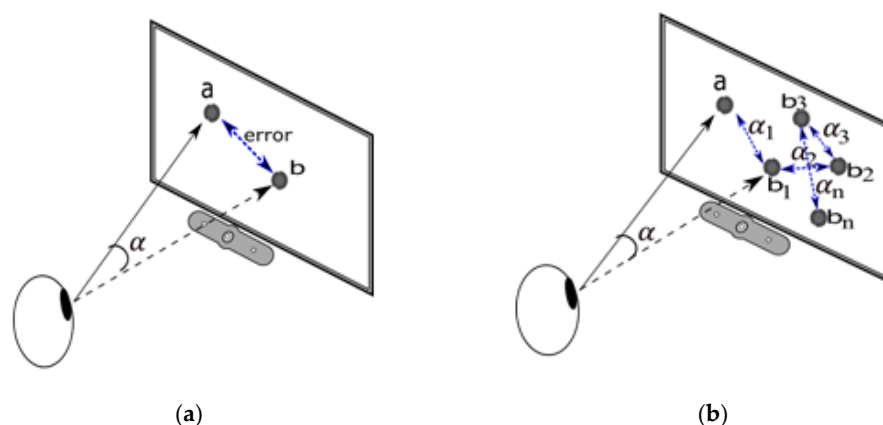


Figure 14. Illustration of the two common evaluation methods for REGT systems: (a) accuracy; (b) precision.

3.3.2. Accuracy Evaluation of REGT Systems

Described in Figure 14a, accuracy has been reported through various metrics in the literature as angular resolution, recognition rate, and pixel distance. The angular resolution describes the mean error of estimating the gaze angle α , as a deviation between real gaze position (ground truth) a and the estimated gaze position b , measured in degrees. The lesser the deviation in gaze angle between real gaze position and the estimated position, the better the accuracy of REGTs. The recognition rate is measured in percent, and it describes the average recognition rate of ground truth gaze positions; the pixel distance,

also known as the Euclidean distance, describes the mean error in distance between the real gaze position and estimated gaze position in millimeters or centimeters.

Measuring the classification accuracy of REGTs is simply done as the ratio of correct predictions (true positives TP and true negative TN) to total predictions $n_{predictions}$ of samples. Using a confusion matrix, the accuracy $m_{accuracy}$ can be computed for the recognition rate as:

$$m_{accuracy} = \frac{(TP + TN)}{n_{predictions}} \quad (9)$$

On the other hand, the measuring regression accuracy of REGT systems can be computed for angular resolution and pixel distances (as derived in [62]) as follows:

$$pixel_{error} = \sqrt{(act_x - est_x)^2 + (act_y - est_y)^2} \quad (10)$$

$pixel_{error}$ is derived for both eyes, act_x and act_y are the actual gaze positions, est_x and est_y are the estimated gaze positions.

$$angular_{error} = \frac{(pixel_{size} \times pixel_{error} \times \cos(gaze_{angle}))^2}{gaze_{point}} \quad (11)$$

$angular_{error}$ is derived for both eyes, $pixel_{error}$ is the mean value of $gaze_{point}$, $gaze_{angle}$ is the mean value of the gaze angle, and $gaze_{point}$ is the mean distance of the eyes from the tracker. Computations for $gaze_{point}$ and $gaze_{angle}$ are derived in [62].

In Table 3, cited accuracies are presented for various gaze estimation methods reported as angular resolution, gaze recognition rate and the pixel distance. The error for angular resolution and pixel distances are the commonly used metric to measure the accuracy of REGTs methods and systems in the literature. Deviating from this practice, the authors of [68] presented an argument against these metrics and their implications for inter-comparison of gaze estimation methods. They also stressed the need for a holistic consideration of sources of error that arise from the different components that went into the gaze estimation, such as the subject, device, and environment. Thus, they suggested that various parameters be evaluated within each component and across deployment platforms (for example, subject component: head pose variation, eye occlusion, human eye condition; device: camera quality, properties of the deployment device; environment: user distance, illumination changes, and motion caused by deployment device) to make fair inter-comparisons of gaze estimation methods possible.

Table 3. An accuracy comparison of different gaze estimation methods. Abbreviations: near infrared light (NIR), point interpolation (PI), linear interpolation (LI), eye corner (EC), ellipse shape (ES), iris center (IC).

Methods	Literature	Accuracy	Hardware Setup	Software Process
Active light feature-based	[203]	<1°	Desktop, stereo infrared camera, 3 NIR	Purkinje Image, 1 point
	[24]	0.9°	Desktop, 1 infrared camera, 2 NIR	PCCR, Multiple points
	[204]	96.71%	Desktop, 2 infrared camera, 4 NIR	PCCR, Multiple points
	[205]	10.3 mm	1 infrared camera, 4 NIR	PCCR, Multiple point
Active light model-based	[206]	<1°	Desktop, 1 infrared camera, 4 NIR	Eye model, 1 point
	[207]	1°	Desktop, stereo camera, Pan-tilt infrared camera, 1 NIR	Eye model, 2 points
	[98]	<1°	Desktop, 1 infrared camera, 2 NIR	Eye model, Multiple
Passive light feature-based	[208,209]	1.6°	Desktop, 1 web camera	PC-EC, GP, Grid
	[61,101,128]	1.2°–2.5°	Desktop, 1 web camera	PC-EC, PI, Grid
	[210]	>3°	Desktop, 1 web camera	EC, LI

Table 3. Cont.

Methods	Literature	Accuracy	Hardware Setup	Software Process
Passive light model-based	[26]	2.42°	Desktop, 1 web camera	ES-IC, Grid
	[70]	<1°	Desktop, 1 web camera	ES, Grid
	[41]	~500 Hz	Desktop, 2 web camera	Eye model, Grid
Passive light appearance-based with machine learning	[118]	1.53°	Desktop, 1 web camera	RF, 25 points
	[124]	2°	Desktop/Handheld, 1 web camera	GP, Grid
	[120–122]	2.2°–2.5°	Desktop, 1 web camera	LLI, Grid
	[110]	<3.68°	Desktop, 1 web camera	ANN, 50 points
	[119,123]	3.5°–4.3°	Desktop, 1 web camera	LLI, Saliency
Passive light appearance-based with deep learning	[108]	4.8°–7.5°	Desktop, 1 web camera	KNN, Calibration free
	[211]	7.74°	Handheld, 1 web camera	CNN, Calibration free
	[191]	81.37%	Desktop, 1 web camera	CNN, Calibration free
	[17]	1.71 cm and 2.53 cm	Handheld, 1web camera	CNN, Calibration free

To unite and standardize the diversity of evaluation protocols used in the literature, they proposed a framework in [212] to report system performance in formats that are quantitative and uniform for angular accuracy metrics, statistical metrics, sensitivity metrics, and a new metric based on receiver operating characteristic (ROC). This evaluation framework takes into consideration the characteristics of gaze estimation methods to fairly describe the quality of REGT systems. The ROC curve is plotted using Equation (12).

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (12)$$

In order to approximate the values of the true positive rate (TPR) and false positive rate (FPR), data from the gaze tracker are obtained to estimate the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Then, validating the ROC curve, the TPR and FPR values are computed with random error thresholds for TP , TN , FP , and FN . This metric, however, is yet to gain much recognition for evaluating REGT systems.

3.4. Dataset

Datasets are important data acquisition means for REGTs, built to store image or video data. Data stored in a dataset are usually labeled or annotated for head pose, gaze direction, gaze point, resolution, and illumination conditions. In early literature on appearance-based methods, datasets presented collected data under controlled laboratory conditions from synthesized images [117], and natural faces were referred to as realistic or naturalistic images [124,213–218]. However, because it was difficult to obtain well-labeled data in these datasets, the authors of [16,219,220] presented well-labeled datasets for naturalistic images, and dataset of synthesized images are presented in [93,221]. Recent methods for gaze estimation are focused on the real world; they use datasets containing data from real-world settings for validation. In Table 4, we present some datasets that have been used for real-world gaze estimation in the recent literature.

Table 4. Recent datasets (2015–2020) for real-world gaze estimation. The listed datasets contain varied data for gaze targets, head pose, and illumination conditions. Ex. = example of accuracies reported for training sets. # = number of samples.

Sample	Dataset	Year	Mode (Annotation)	# Samples	Resolution (Pixels)	Ex. Accuracy for Training
Image	MPIIGaze [53]	2015	RGB eye, 2D & 3D gaze	213,659	1280 × 720	7.74° [211], 4.3° [112], 4.8° [114]
	GazeCapture [17]	2016	RGB full face, 2D gaze	2,445,504	640 × 480	3.18° [185]
	RT-GENE [112]	2018	RGB-D full face, 3D gaze	277,286	1920 × 1080	7.7° [112], 24.2° [20], 8.4° [222]
	RT-BENE [173]	2019	RGB-D full face, 3D gaze	210,000	1920 × 1080	0.71° [173]
	Gaze360 [20]	2019	RGB full face, 3D gaze	172,000	4096 × 3382	2.9° [20]
	XGaze [190]	2020	RGB full face, 2D & 3D gaze	1,083,492	6000 × 4000	4.5° [190]
Video	EyeDiap [223]	2014	RGB-D full face, 2D & 3D gaze	94	1920 × 1080	5.71° [222], 5.84° [176], 5.3° [224]
	TabletGaze [16]	2017	RGB full face, 2D gaze	816	1280 × 720	3.63° [53], 3.17° [16], 2.58° [17]
	EVE [192]	2020	RGB full face, 2D & 3D gaze	161	1920 × 1080	2.49° [192]

Each of these datasets was composed under varied illumination conditions suitable for passive light gaze estimation. The XGaze [190] dataset was collected in the laboratory using 18 custom high definition SLR cameras with 16 adjustable illumination conditions. Participants included 110 subjects (63 male, 47 female) who were required to stare at an on-screen stimulus. EVE [192] was collected using a desktop setup. The study included 54 participants and four camera views, and natural eye movements as opposed to following specific instructions or smoothly moving targets. Gaze360 [20] captured videos of 238 subjects, and five indoor (53 subjects) and two outdoor (185 subjects) locations, over nine recording sessions using a Ladybug5 camera. RT-BENE [173] was released based on the RT-GENE [112] dataset collection. RT-GENE was captured using mobile eye tracking glasses, and contains recordings of 15 participants, 9 male, 6 female; two participants were recorded twice. TabletGaze [16] was captured on a mobile tablet setup using 51 subjects of varying race, gender, and glasses prescription. GazeCapture [17] contains data from over 1450 subjects, recorded using crowdsourcing: 1249 subjects used iPhones, whereas 225 used iPads. MPIIGaze [53] was collected with a laptop setup involving 15 participants during natural everyday laptop use. EyeDiap [223], was collected from 16 subjects (12 males and 4 females) using a Kinect sensor and an HD camera.

The characteristics of these datasets largely determine the kinds of applications they are suitable for. In Figure 15, we describe some important characteristics of the dataset and their application suitability.

Image samples, for example, would make a dataset with them more suitable for image-based gaze estimation and not video-based. If a dataset includes on screen gaze annotation, then it is most applicable for 2D gaze estimation. The resolution and quality of input images impacts the model's resources. Lower resolution would perform much better with lighter deployment resources, a shallower model architecture, and a lower model inference speed. Having a larger number of data samples is suitable for a deeper model, but could negatively impact resource-constrained deployment, for instance, on mobile devices. However, for training any model, error decreases considerably as the number of samples is increased, showing the significance of collecting much data.

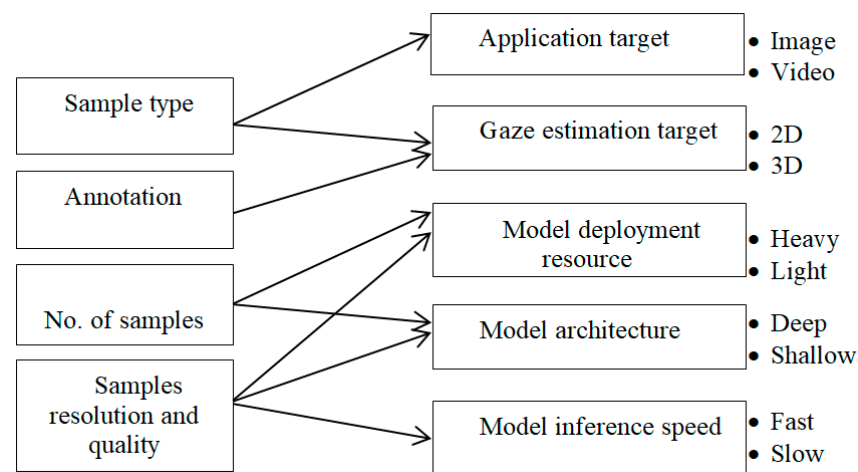


Figure 15. Dataset characteristics and factors that determine application suitability.

Important progress has been made for large-scale dataset collection in the recent literature. First, collecting large-scale images required by supervised gaze estimation methods through synthesis became popular because it is time consuming and tedious to label naturalistic images [225]. Even though learning-by-synthesis has the potential to save time and resource for data collection, the desired performance for highly accurate gaze estimation is still not achieved by this method due to the different illumination distribution. While early attempts in [226–228] trained models to improve the realism of the synthetic images, a different approach was presented to purify naturalistic images in [225,229,230]. The authors adopted the style transfer technique proposed in [227] to convert an outdoor naturalistic image distribution into indoor synthetic images. Their proposed method utilizes three separate networks: a coarse segmentation network, a feature extraction network, and a loss network. The coarse network takes in two images (naturalistic image and a synthetic reference style image) with their mask as the input for segmentation into the pupil (white) and the iris (red), followed by a feature extraction network that extracts gaze direction and pupil center position as the image content of the naturalistic image. Then, through a standard perception loss network, they retain the information of the naturalistic image and the distribution of the synthetic image (i.e., image color structure and semantic features) to the fullest extent in the output image. Using the proposed method, they purified the MPIIGaze [53] dataset. Experimental results have shown how purifying naturalistic images for training an appearance-based gaze estimator leads to improved performance compared to some state-of-the-art techniques.

Second, a broader range of settings for datasets was addressed. The authors of [20,190] proposed a broader range of settings for datasets in order to increase robustness to a larger variety of environmental conditions, as previous datasets are limited to relatively narrow ranges of head poses and gaze directions, such as the frontal face setting. In [190], the authors collected and made public the ETH-XGaze dataset of varied viewpoints, lighting, extreme gaze angles, resolutions, and occluders, such as glasses. Their dataset provided maximum head poses and gaze in horizontal (around yaw axis) and vertical (around pitch axis) directions in the camera coordinate system: $\pm 80^\circ$, $\pm 80^\circ$, and $\pm 120^\circ$, $\pm 70^\circ$ respectively. The image resolution is 6000×4000 . Similarly, the dataset [20] in provides $\pm 90^\circ$, unknown, and $\pm 140^\circ$, -50° for maximum head poses and gaze in the same axis with image resolution of up to 4096×3382 . These attempts have paved the way towards robust gaze estimation for unconstrained environmental conditions, particularly with respect to lighting and coverage of extreme head pose and gaze, which are critical for emerging gaze-based applications.

Benchmarks for Evaluating REGT Performance

At present, five evaluation protocols are available in the literature, two of which are popular:

1. Cross-dataset Evaluation

A given gaze estimation method is trained on one dataset (e.g., dataset A) and tested on another dataset (e.g., dataset B). These data samples may have the same or different characteristics, which correlates with the purpose of this method, which is to show the generalization capabilities of a given gaze estimation method. Examples of gaze estimation research that have utilized this evaluation method can be found in [20,38,53,112,231].

2. Within-dataset Evaluation

Examples of this evaluation can be found in [117]. The gaze estimation method is trained and tested with the same dataset, which is randomly separated into training and test samples. These samples may also have the same or different data characteristics. However, the generalization capability is verified less with this method compared to the cross-data evaluation method.

Recent studies have proposed subject-specific, cross-device, and robustness evaluations. These new evaluation methods have focused on specific data and device characteristics for evaluating emerging gaze estimation methods as follows:

3. Subject-specific Evaluation

This evaluation technique became popular due to the recent attention paid to subject-specific gaze estimation [185]. The evaluation is based on sampling a little subject-specific training data to improve the adaptation of a generic gaze estimator to a specific person from a few samples due to insufficient data samples. Other works [16,19,232] have also used this method.

4. Cross-device Evaluation

Emergence of cross-device training necessitated cross-device evaluation. This method evaluates a given gaze estimation method across multiple devices to measure cross-device performance. Researchers in [40] demonstrated the significance of this evaluation method on five common devices (mobile phone, tablet, laptop, desktop computer, smart TV).

5. Robustness Evaluation

Researchers in [190] proposed assessing the robustness of gaze estimation methods across head poses and gaze directions. The authors have stressed its importance to previous gaze estimation methods that only report the mean gaze estimation errors for cross-dataset, within-dataset, and person-specific evaluations on insufficient data samples that do not cover a wide range of head pose and gaze direction. They also argued that knowing the performance of a gaze estimation method with respect to robustness across head poses and gaze directions is important, since a method with a higher overall error might have a lower error within a specific range of interest. They further demonstrated and compared recent methods to assess their robustness and reported gaze estimation error distribution across head poses and gaze directions in horizontal and vertical directions to facilitate research into robust gaze estimation methods.

4. Applications

The use of gaze applications as a solution has expanded over time. We present how gaze application usage has evolved over time across fields of human endeavors and deployment platforms in Figure 16. The authors of [233] presented an early review on gaze applications. They broadly classified gaze application usage into two fields, diagnostic (see examples in [234–236]) and interactive (see examples in [237,238]). Later reviews in [63,91,239] presented wider usage for gaze applications, which included that by REGT systems, augmented reality systems [240–244], human behavior analysis [245–250], and

bio-metric authentication [251–255]. Recently, the usage of REGTs has gotten even wider with emerging usage in privacy-aware interactions.

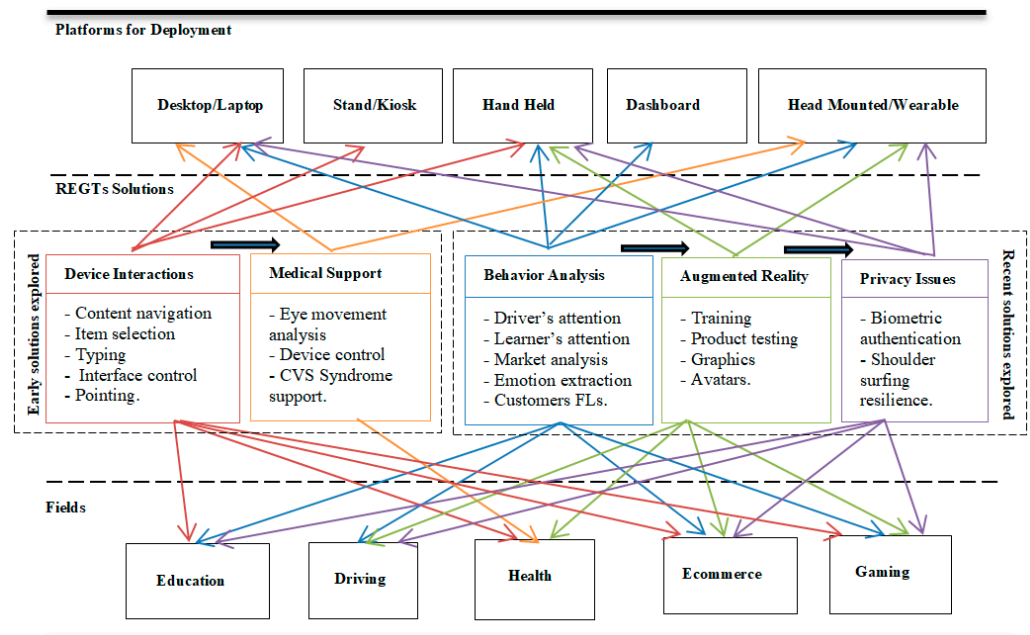


Figure 16. Classification of eye solutions across fields and deployment platforms.

We have identified emerging eye solutions and classified them under the existing categories, such as device interactions (IoT smart home control [256], semi-autonomous driving [257,258], artistic drawing in robots [259–262]), human behavior analysis (confusion prediction [263–265], intention extraction [266,267], driver’s attention [30,32,268–270], detecting personality traits [271]), medical support (medical image interpretation [272–275], patients support [276]), augmented reality (social games [277,278], virtual space control [279]), and privacy issues (privacy-aware eye tracking [280–283], gaze-touch authentication [284–288]). The arrow pointers indicate a solution application in a field and across deployment platforms.

The emerging solutions are responsible for the recent shift in requirements for REGT system deployment on desktop platforms as a decades old practice [68] to those for dynamic platforms such as handheld devices and wearables [155]. In addition to applications, the large number of REGT systems produced today is research based, most of which are fabricated to suit research objectives. Several commercial REGT systems are also available on the market. Recently, ref. [289] ranked the top 12 REGT companies based on the number of patents and equipment testing. Tobii [290], SensoMotoric Instruments [291], now acquired by Apple Inc, and EyeLink [292], were ranked as leading companies. Several open-source REGT systems were released for free online to give new researchers a starting point, ensuring the sustainability of this research field. We have classified and described a few from academia and industry in Table 5.

Table 5. Available open-source REGT software.

Method	Provider	Language	Description
Passive-light	Itracker [17]	Python, Matlab	A CNN based eye tracker, which runs in real time (10–15 fps) on a modern mobile device
	RecurrentGaze [177]	Python	Based on a fusion of CNN-RNN
	NNET [293]	-	ANN based eye tracker implementation for iPad devices
	EyeTab [294]	Pthon, C++	Webcam model-based approach for binocular gaze estimation
	Opengazer [295]	C++, C	Based on the Viola-Jones face detector, that locates the largest face in the video stream capture from PC webcam
	TurkerGaze [296]	JavaScript, HTML	A webcam-based eye tracking game for collecting large-scale eye tracking data via crowdsourcing
	Camgaze [297]	Python	Binocular gaze estimation for webcam
	ITU gaze tracker [298]	-	Based on remote webcam setup
	CVC ET [299]	C++	Enhanced Opengazer with head repositioning feature which allows users to correct their head pose during eye tracker usage in order to improve accuracy.
	xLabs [300]	-	Webcam-based eye tracker, built as a browser extension for Google Chrome.
	Gazepointer [301]	C#, HTML	Windows-based web camera gaze estimation
	MyEye [302]	-	Gaze-based input designed for use by people with amyotrophic lateral sclerosis (ALS), a neuromuscular disease.
	NetGazer [303]	C++	Port of Opengazer for the Windows platform
	Active-light	OpenEyes [304]	Matlab
Ogama [305]		C#.NET	Uses infrared ready webcams
GazeParser [306]		Python	Based on infrared illumination. Python.
Pygaze [307]		Python	Wrapper for EyeLink, SMI, and Tobii systems.

It is likely that most of the passive light open-source applications will be supported and remain active much longer than the active light ones, because the active light open-source applications have been made available for a long time with no further support. Recently, ref. [308] has been providing implementation files (i.e., code, models, and datasets) for open-access academic papers.

5. Summary

We have discussed REGT systems that process both continuous and static data for eye gaze estimation. We focused on REGT's remote applications for PCs, laptops, and mobile devices, and have presented a thorough evaluation of the past and recent research. We have classified and discussed our evaluation in three parts: hardware, software, and applications. In our evaluation, we have compared key research trends in REGT to show how its research has changed over time. These key research trends have been summarized in Table 6. We mean by past progress, earlier information provided in the literature before 2015, and by recent progress, new information provided in the literature after 2015. Although our

literature search expanded over seven publishers (i.e., IEEE, ACM, Springer, MDPI, Elsevier, BOP, and PLOS) of academic journals and conference papers, we only cited some, not all literature relevant to the components discussed in this paper. The search for relevant literature was done using the keywords: remote eye gaze tracking; hardware setup for eye gaze tracking; software process (techniques and datasets) for eye gaze tracking; application for eye gaze tracking. In some cases, eye gaze tracking was replaced with gaze estimation.

Table 6. Summary of research trends in REGT. The cited works present examples of the points stated.

Component	Past Research Trend (<2015)	Recent Research Trend (>2015)
Hardware setup	<ul style="list-style-type: none"> Gaze tracking on modified devices was a common practice [41,48,267] Relied on NIR illumination, quality cameras, high computational power, and a defined subject position [26,28,41,98,100,127] Multiple cameras used in device setup to address the change in head position [63,64] 	<ul style="list-style-type: none"> Unmodified devices utilized for gaze tracking [16,43] Gaze interaction from arbitrary positions and orientations [52], multiple displays [40] Utilization of low-resolution cameras for visible light gaze tracking in both indoor and outdoor [17,61,92,112]
Image acquisition	<ul style="list-style-type: none"> Gaze data was acquired via video camera [61,110], and from dataset e.g., video [213] or image [215,218] 	<ul style="list-style-type: none"> The acquisition of large quantity images required for training deep learning gaze estimators are generated by synthesis [93,117,221]
Feature extraction	<ul style="list-style-type: none"> Largely depended on high quality images to extract rich features e.g., CHT [136–138], LLD [139], Subpixel [140] Extraction of features from low quality images commonly used hand-engineered classifiers e.g., haar-cascade [141,142] 	<ul style="list-style-type: none"> Scalable classifiers was used to extract visible and deep features based on Machine learning, e.g., Genetic algorithms [160], SVM [36], and Deep learning e.g., R-CNN [169,170], and YOLO [171,172]
Gaze mapping	<ul style="list-style-type: none"> Naming schemes for gaze estimation methods were broadly categorized into two with so much ambiguity [87]. Supervised learning was used for gaze mapping by early appearance-based methods [53,155,156] Explicit calibration was largely required for gaze mapping [24,41] Gaze mapping techniques were driven by face or eye model [61] 	<ul style="list-style-type: none"> Naming schemes for gaze estimation methods regrouped into 3 broad categories [71,93] Directly regress from eye image to gaze direction using machine, or deep learning algorithms [89,108,112,211] Unsupervised learning for appearance-based methods demonstrated [157–159] Calibration requirement became implicit (i.e., less required), researchers achieved gaze mapping through automatic calibration procedures [119,183] Gaze mapping techniques were driven by large and well annotated data [17,93] Proposal to have a standardized evaluation metric for gaze estimation methods that ensure accurate validation and comparison [68,190]
Dataset	<ul style="list-style-type: none"> Data collection was largely done under controlled laboratory conditions [124,213–218] Datasets were limited to the frontal face setting that provides only narrow range of head poses and gaze directions [215,218,223] 	<ul style="list-style-type: none"> Presentation of datasets with data from the real-world [53,223] Attempts to purify naturalistic images through style transfer [225,229,230] Cross-dataset evaluation for utilizing several datasets and domain adaptation for a more generalized, robust gaze model validation [20,38,231,232] Attempts to create a broader range of settings for datasets [20,190] to ensure robustness to a larger variety of conditions such as varied viewpoints, extreme gaze angles, lighting variation, input image resolutions
Application	<ul style="list-style-type: none"> Early gaze applications were used as diagnostic [233–235], and interactive [236,237] tools Desktop-based platforms for deployment was a commonly used [68] 	<ul style="list-style-type: none"> REGTs have been used recently in Social games [271,272], and Privacy-aware interactions [273–276] Gaze application deployed on dynamic platforms e.g., handheld devices, and wearables [155]

In the past five (5) years, efforts by researchers have been made to eliminate the inherent challenges that have made highly accurate gaze estimation results difficult. Various attempts at appearance-based gaze estimation were focused mainly on addressing:

- Feature detection and extraction: The accuracy of gaze estimation depends largely on effective feature detection and extraction. Recent learning-based feature detection and extraction have brought remarkable progress for appearance-based gaze estimation, as researchers have continued to demonstrate its effectiveness [170–172]. However, the issue of deviations in location and degree of eye openness remains a challenge for this method.
- Variability issues in data collection: Collection of relevant data of varied characteristics has been a challenge for appearance-based gaze estimation. Recent attempts by researchers to create and make available robust datasets have been presented in [20,112,190]. However, the reported accuracies on these datasets are still not satisfactory, and thus require more attention.
- Subject calibration: Gaze estimation methods require a subject calibration process for different individuals. Appearance-based methods have demonstrated less stringent processes of calibration, where only a few samples of calibration data are required before they work for other individuals [185,232]. Even with this development, an appearance-based gaze estimation method that tries to either make the calibration process less stringent or absolutely remove it does so at the expense of estimation accuracy. Thus, the need to develop absolute appearance-based calibration-free methods with very good estimation accuracy remains a challenge.
- Head fixation: Although several appearance-based gaze estimation methods have been able to perform well with considerable accuracies, without requiring fixed head poses [93,108,111,113,211], most of them still can only handle small head movements to achieve high accuracy [110,116]. As such, more robust methods that freely allow for head movement are still sought.

Notwithstanding, more recent attempts in the past 2–3 years have focused on appearance-based deep learning methods. The use of CNNs, RNNs, and GANs has shown greater promise in addressing quite effectively some of the challenges mentioned for appearance-based methods. However, there are still factors negatively influencing overall performance of appearance-based deep learning methods. These factors include eye blink, occlusion, and dark scenes. Eye blink and occlusion are caused by obstructions which affect eye feature detection. Eye blinking is a temporary closure of both eyes, involving movements of the upper and lower eyelids. Occlusions may occur from eye illness, aging, and objects such as lenses and glasses. Dark scenes produce dark images, which reduce the sensitivity of local features, thereby affecting effective eye feature detection. Other concerns generated by appearance-based using deep learning that can lower accuracy are:

- Model input: Researchers are trying to determine if the choice of model input has any effect on the performance of the model. For instance, does a model that uses both eyes perform better than one that uses one eye in terms of accuracy? We have seen several attempts with models that use one (single) eye [188,191], both (double) eyes [17,53,111,112], or full face images [114,190], augmented with other inputs, such as head pose and face grid information, to estimate gaze. At the moment, there is not a clear understanding of how to determine the best or standard model input for deep learning gaze estimation methods. It is done at the discretion of researchers, and still remains a choice of convenience for the models proposed.
- Resolution of input images: Does training and testing a model with different input image sizes improve performance? In [38,53], the model trained with images of one size achieved much worse results than the one trained on images of multiple sizes. On the other hand, Zhang et al. [190] demonstrated an improved performance when training and testing with a single image size at high resolutions. Based on this, researchers are suggesting methods to handle training of cross-resolution input images.

- **Model augmentation:** Are there working step-up techniques employed for improving gaze estimation errors on public datasets from recent years? Early attempts used a single-region CNN model for gaze estimation, as demonstrated in [53,114], but recent step-up attempts are opting for multi-region CNNs [17,112,173,188], where each input is processed through a separate network to provide each network with a higher resolution of the input image and improve the processing capabilities of the networks for a better performance of the overall model. Another step-up technique is to improve early CNNs' poor generalization performances due to appearance and head pose variations using adversarial learning approaches. As recently attempted by researchers [179–182], the basic idea is to improve the generalization performance of a traditional CNN-based gaze estimator by incorporating adversarial nets with ConvNet. The adversarial nets are commonly used to improve the input image fed into the ConvNet to estimate gaze.
- **Data annotation:** Training a deep learning gaze estimation model by supervision requires data annotation for the model to learn the task. Grid-based annotation has been widely adopted for this method. How can its effectiveness be judged? It divides the screen into horizontal w , vertical h , and grids g , which produces $(w * h) / g$ grids. This usually causes more calculations and affects the accuracy of the model's output. Considering these drawbacks, bin-based annotation was proposed in [174] to control the number of labels on the gaze image. It divides the screen into horizontal and vertical bins and set the bin size as b , making $(w + h) / b$ bins. The grid-based annotation yields more annotation data, whereas the bin-based annotation yields less annotation data but requires more processing steps. Due to lack of sufficient reports on alternative annotation techniques, it is difficult to judge the effectiveness of the grid-base annotation.

6. Conclusions

This paper has presented a thorough comparative evaluation on REGT research. Our evaluation has discussed key components and research changes over time. In addition, issues inherent in the recent state of REGT research have been outlined. The prevalence of the issues outlined negatively impacts the performance and accuracy of the recently proposed gaze estimation methods which utilize deep learning. This has raised new concerns about how to further enhance the recent appearance-based deep learning methods through continued research efforts for developing new algorithms and augmentation techniques that will ensure REGT systems become more useful throughout fields of human endeavor in the very near future.

Author Contributions: Conceptualization and manuscript writing, I.S.S.; resources and manuscript editing, Y.W. and A.M.A.; supervision, X.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China Grant 62002043 and Grant 62176037, by the Research Project of China Disabled Persons' Federation on Assistive Technology Grant 2021CDPFAT-09, by the Liaoning Revitalization Talents Program Grant XLYC1908007, by the Dalian Science and Technology Innovation Fund Grant 2019J11CY001.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. What Eye-Tracking Can and Can't Tell You about Attention. Available online: <https://www.nmsba.com/buying-neuromarketing/neuromarketing-techniques/what-eye-tracking-can-and-cant-tell-you-about-attention> (accessed on 7 October 2019).
2. Judd, C.H.; McAllister, C.N.; Steele, W.M. General introduction to a series of studies of eye movements by means of kinetoscopic photographs. *Psychol. Rev. Monogr.* **1905**, *7*, 1–16.
3. Duchowski, A. *Eye Tracking Methodology: Theory and Practice*, 2nd ed.; Springer: London, UK, 2007. [CrossRef]
4. Mowrer, O.H.; Theodore, C.R.; Miller, N.E. The corneo-retinal potential difference as the basis of the galvanometric method of recording eye movements. *Am. J. Physiol. Leg. Content* **1935**, *114*, 423–428. [CrossRef]

5. Marge, E. Development of electro-oculography; standing potential of the eye in registration of eye movement. *AMA Arch. Ophthalmol.* **1951**, *45*, 169–185. [[CrossRef](#)]
6. Glenstrup, A.; Engell-Nielsen, T. Eye Controlled Media: Present and Future State. Master's Thesis, University of Copenhagen, Copenhagen, Denmark, 1995.
7. Yoo, D.H.; Chung, M.J. Non-intrusive eye gaze estimation without knowledge of eye pose. In Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 19 May 2004; pp. 785–790. [[CrossRef](#)]
8. Morimoto, C.H.; Mimica, M.R. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.* **2005**, *98*, 4–24. [[CrossRef](#)]
9. Rayner, K. Eye movements in reading and information processing: 20 Years of research. *Psychol. Bull.* **1998**, *124*, 372–422. [[CrossRef](#)] [[PubMed](#)]
10. Young, L.R.; Sheena, D. Survey of eye movement recording methods. *Behav. Res. Methods Instrum.* **1975**, *7*, 397–429. [[CrossRef](#)]
11. Eggert, T. Eye movement recordings: Methods. *Dev. Ophthalmol.* **2007**, *40*, 15–34. [[CrossRef](#)]
12. Joyce, C.A.; Gorodnitsky, I.F.; King, J.W.; Kutas, M. Tracking eye fixations with electroocular and electroencephalographic recordings. *Psychophysiology* **2002**, *39*, 607–618. [[CrossRef](#)]
13. Oeltermann, A.; Ku, S.; Logothetis, N.K. A novel functional magnetic resonance imaging compatible search-coil eye-tracking system. *Magn. Reson. Imaging* **2007**, *25*, 913–922. [[CrossRef](#)]
14. Domdei, N.; Linden, M.; Reiniger, J.L.; Holz, F.G.; Harmening, W.M. Eye tracking-based estimation and compensation of chromatic offsets for multi-wavelength retinal microstimulation with foveal cone precision. *Biomed. Opt. Express* **2019**, *10*, 4126–4141. [[CrossRef](#)]
15. Reingold, E.M. Eye Tracking Research and Technology: Towards Objective Measurement of Data Quality. *Vis. Cogn.* **2014**, *22*, 635–652. [[CrossRef](#)]
16. Huang, Q.; Veeraraghavan, A.; Sabharwal, A. TabletGaze: Dataset and analysis for unconstrained appearance based gaze estimation in mobile tablets. *Mach. Vis. Appl.* **2017**, *28*, 445–461. [[CrossRef](#)]
17. Krafka, K.; Khosla, A.; Kellnhofer, P.; Kannan, H.; Bhandarkar, S.; Matusik, W.; Torralba, A. Eye Tracking for Everyone. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2176–2184. [[CrossRef](#)]
18. Carlin, J.D.; Calder, A.J. The neural basis of eye gaze processing. *Curr. Opin. Neurobiol.* **2013**, *23*, 450–455. [[CrossRef](#)] [[PubMed](#)]
19. Liu, G.; Yu, Y.; Funes-Mora, K.A.; Odobez, J. A Differential Approach for Gaze Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1092–1099. [[CrossRef](#)]
20. Kellnhofer, P.; Recasens, A.; Stent, S.; Matusik, W.; Torralba, A. Gaze360: Physically Unconstrained Gaze Estimation in the Wild. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 6911–6920. [[CrossRef](#)]
21. Hoshino, K.; Shimano, S.; Nakai, Y.; Noguchi, Y.; Nakamura, M. Estimation of the Line of Sight from Eye Images with Eyelashes. In Proceedings of the 5th International Conference on Intelligent Information Technology (ICIIT 2020), Hanoi, Vietnam, 19–22 February 2020; pp. 116–120. [[CrossRef](#)]
22. Strupczewski, A. Commodity Camera Eye Gaze Tracking. Ph.D. Dissertation, Warsaw University of Technology, Warsaw, Poland, 2016.
23. Wang, J.; Sung, E. Study on eye gaze estimation. *IEEE Trans. Syst. Man Cybern.* **2002**, *32*, 332–350. [[CrossRef](#)]
24. Guestrin, E.D.; Eizenman, M. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1124–1133. [[CrossRef](#)]
25. Morimoto, C.H.; Koons, D.; Amir, A.; Flickner, M. Pupil detection and tracking using multiple light sources. *Image Vis. Comput.* **2000**, *18*, 331–335. [[CrossRef](#)]
26. Baek, S.; Choi, K.; Ma, C.; Kim, Y.; Ko, S. Eyeball model-based iris center localization for visible image-based eye-gaze tracking systems. *IEEE Trans. Consum. Electron.* **2013**, *59*, 415–421. [[CrossRef](#)]
27. Lee, J.W.; Cho, C.W.; Shin, K.Y.; Lee, E.C.; Park, K.R. 3D gaze tracking method using Purkinje images on eye optical model and pupil. *Opt. Lasers Eng.* **2012**, *50*, 736–751. [[CrossRef](#)]
28. Sigut, J.; Sidha, S. Iris Center Corneal Reflection Method for Gaze Tracking Using Visible Light. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 411–419. [[CrossRef](#)] [[PubMed](#)]
29. Murphy-Chutorian, E.; Doshi, A.; Trivedi, M.M. Head Pose Estimation for Driver Assistance Systems: A Robust Algorithm and Experimental Evaluation. In Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Washington, DC, USA, 30 September–3 October 2007; pp. 709–714. [[CrossRef](#)]
30. Fu, X.; Guan, X.; Peli, E.; Liu, H.; Luo, G. Automatic Calibration Method for Driver's Head Orientation in Natural Driving Environment. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 303–312. [[CrossRef](#)]
31. Lee, S.J.; Jo, J.; Jung, H.G.; Park, K.R.; Kim, J. Real-Time Gaze Estimator Based on Driver's Head Orientation for Forward Collision Warning System. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 254–267. [[CrossRef](#)]
32. Wang, Y.; Yuan, G.; Mi, Z.; Peng, J.; Ding, X.; Liang, Z.; Fu, X. Continuous Driver's Gaze Zone Estimation Using RGB-D Camera. *Sensors* **2019**, *19*, 1287. [[CrossRef](#)] [[PubMed](#)]
33. Kaminski, J.Y.; Knaan, D.; Shavit, A. Single image face orientation and gaze detection. *Mach. Vis. Appl.* **2008**, *21*, 85. [[CrossRef](#)]
34. Smith, P.; Shah, M.; Lobo, N. Determining driver visual attention with one camera. *IEEE Trans. Intell. Transp. Syst.* **2003**, *4*, 205–218. [[CrossRef](#)]

35. Valenti, R.; Sebe, N.; Gevers, T. Combining Head Pose and Eye Location Information for Gaze Estimation. *IEEE Trans. Image Process.* **2012**, *21*, 802–815. [CrossRef]
36. Zhu, Z.; Ji, Q. Eye and gaze tracking for interactive graphic display. *Mach. Vis. Appl.* **2004**, *15*, 139–148. [CrossRef]
37. Lu, F.; Sugano, Y.; Okabe, T.; Sato, Y. Adaptive Linear Regression for Appearance-Based Gaze Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2033–2046. [CrossRef]
38. Zhang, X.; Sugano, Y.; Fritz, M.; Bulling, A. MPIIGaze: Real-World Dataset and Deep Appearance-Based Gaze Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 162–175. [CrossRef]
39. Shehu, I.S.; Wang, Y.; Athuman, A.M.; Fu, X. Paradigm Shift in Remote Eye Gaze Tracking Research: Highlights on Past and Recent Progress. In Proceedings of the Future Technologies Conference (FTC 2020), Vancouver, BC, Canada, 5–6 November 2021; Volume 1. [CrossRef]
40. Zhang, X.; Huang, M.X.; Sugano, Y.; Bulling, A. Training Person-Specific Gaze Estimators from User Interactions with Multiple Devices. In Proceedings of the Conference on Human Factors in Computing Systems (CHI 2018), Montreal, QC, Canada, 21–26 April 2018. [CrossRef]
41. Zhu, Z.; Ji, Q. Novel Eye Gaze Tracking Techniques under Natural Head Movement. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 2246–2260. [CrossRef] [PubMed]
42. Sticky by Tobii Pro. Available online: <https://www.tobii.com/product-listing/sticky-by-tobii-pro/> (accessed on 10 October 2019).
43. Wood, E.; Bulling, A. EyeTab: Model-based gaze estimation on unmodified tablet computers. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA 2014), Safety Harbor, FL, USA, 26–28 March 2014; pp. 207–210. [CrossRef]
44. Zhang, Y.; Chong, M.K.; Müller, J.; Bulling, A.; Gellersen, H. Eye tracking for public displays in the wild. *Pers. Ubiquitous Comput.* **2015**, *19*, 967–981. [CrossRef]
45. The iMotions Screen-Based Eye Tracking Module. Available online: <https://imotions.com/blog/screen-based-eye-tracking-module/> (accessed on 5 February 2020).
46. Matsuno, S.; Sorao, S.; Susumu, C.; Akehi, K.; Itakura, N.; Mizuno, T.; Mito, K. Eye-movement measurement for operating a smart device: A small-screen line-of-sight input system. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON 2016), Singapore, 22–25 November 2016; pp. 3798–3800. [CrossRef]
47. How to Get a Good Calibration. Available online: <https://www.tobiidynavox.com/supporttraining/eye-tracker-calibration/how-to-get-a-good-calibration/> (accessed on 16 September 2019).
48. Drewes, H.; De Luca, A.; Schmidt, A. Eye-Gaze Interaction for Mobile Phones. In Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology, Singapore, 10–12 September 2007; pp. 364–371. [CrossRef]
49. Cheng, H.; Liu, Y.; Fu, W.; Ji, Y.; Yang, L.; Zhao, Y.; Yang, J. Gazing Point Dependent Eye Gaze Estimation. *Pattern Recognit.* **2017**, *71*, 36–44. [CrossRef]
50. Gaze Tracking Technology: The Possibilities and Future. Available online: <http://journal.jp.fujitsu.com/en/2014/09/09/01/> (accessed on 17 September 2019).
51. Cho, D.; Kim, W. Long-Range Gaze Tracking System for Large Movements. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3432–3440. [CrossRef]
52. Zhang, X.; Sugano, Y.; Bulling, A. Evaluation of Appearance-Based Methods and Implications for Gaze-Based Applications. In Proceedings of the Conference on Human Factors in Computing Systems (CHI 2019), Glasgow, UK, 4–9 May 2019. [CrossRef]
53. Zhang, X.; Sugano, Y.; Fritz, M.; Bulling, A. Appearance-based gaze estimation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–12 June 2015; pp. 4511–4520. [CrossRef]
54. Ramanauskas, N. Calibration of Video-Oculographical Eye Tracking System. *Electron. Electr. Eng.* **2006**, *8*, 65–68.
55. Kotus, J.; Kunka, B.; Czyzewski, A.; Szczuko, P.; Dalka, P.; Rybacki, R. Gaze-tracking and Acoustic Vector Sensors Technologies for PTZ Camera Steering and Acoustic Event Detection. In Proceedings of the 2010 Workshops on Database and Expert Systems Applications, Bilbao, Spain, 30 August–3 September 2010; pp. 276–280. [CrossRef]
56. Ohno, T.; Mukawa, N.; Yoshikawa, A. FreeGaze: A gaze tracking system for everyday gaze interaction. In Proceedings of the Eye Tracking Research & Application Symposium (ETRA 2002), New Orleans, LA, USA, 25–27 March 2002; pp. 125–132. [CrossRef]
57. Ebisawa, Y.; Satoh, S. Effectiveness of pupil area detection technique using two light sources and image difference method. In Proceedings of the 15th IEEE Engineering Conference in Medicine and Biology Society, San Diego, CA, USA, 31 October 1993; pp. 1268–1269. [CrossRef]
58. Morimoto, C.H.; Amir, A.; Flickner, M. Detecting eye position and gaze from a single camera and 2 light sources. In Proceedings of the International Conference on Pattern Recognition, Quebec City, QC, Canada, 11–15 August 2002; pp. 314–317. [CrossRef]
59. Tomono, A.; Lida, M.; Kobayashi, Y. A TV Camera System Which Extracts Feature Points for Non-Contact Eye Movement Detection. In *Optics, Illumination, and Image Sensing for Machine*; Svetkoff, D.J., Ed.; SPIE: Bellingham, WA, USA, 1990; Volume 1194. [CrossRef]
60. Coutinho, F.L.; Morimoto, C.H. Free head motion eye gaze tracking using a single camera and multiple light sources. In Proceedings of the 19th Brazilian Symposium on Computer Graphics and Image Processing, Amazonas, Brazil, 8–11 October 2006; pp. 171–178. [CrossRef]
61. Cheung, Y.; Peng, Q. Eye Gaze Tracking with a Web Camera in a Desktop Environment. *IEEE Trans. Hum. Mach. Syst.* **2015**, *45*, 419–430. [CrossRef]

62. Accuracy and Precision Test Method for Remote Eye Trackers: Test Specification (Version: 2.1.1). Available online: <https://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/use/what-affects-the-performance-of-an-eye-tracker/tobii-test-specifications-accuracy-and-precision-test-method.pdf/?v=2.1.1> (accessed on 10 February 2011).
63. Lupu, R.G.; Ungureanu, F. A survey of eye tracking methods and applications. *Bul. Inst. Politeh. Iasi* **2013**, *3*, 72–86.
64. Kim, S.M.; Sked, M.; Ji, Q. Non-intrusive eye gaze tracking under natural head movements. In Proceedings of the 26th IEEE Engineering Conference in Medicine and Biology Society, San Francisco, CA, USA, 1–4 September 2004; Volume 1, pp. 2271–2274. [[CrossRef](#)]
65. Hennessey, C.; Fiset, J. Long range eye tracking: Bringing eye tracking into the living room. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA 2012), Santa Barbara, CA, USA, 28–30 March 2012; pp. 249–252. [[CrossRef](#)]
66. Jafari, R.; Ziou, D. Gaze estimation using Kinect/PTZ camera. In Proceedings of the IEEE International Symposium on Robotic and Sensors Environments, Magdeburg, Germany, 16–18 November 2012; pp. 13–18. [[CrossRef](#)]
67. Lee, H.C.; Lee, W.O.; Cho, C.W.; Gwon, S.Y.; Park, K.R.; Lee, H.; Cha, J. Remote gaze tracking system on a large display. *Sensors* **2013**, *13*, 13439–13463. [[CrossRef](#)]
68. Kar, A.; Corcoran, P. A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms. *IEEE Access* **2017**, *5*, 16495–16519. [[CrossRef](#)]
69. Mansouryar, M.; Steil, J.; Sugano, Y.; Bulling, A. 3D Gaze Estimation from 2D Pupil Positions on Monocular Head-Mounted Eye Trackers. In Proceedings of the 9th ACM International Symposium on Eye Tracking Research & Applications (ETRA 2016), Charleston, SC, USA, 14–17 March 2016; pp. 197–200. [[CrossRef](#)]
70. Venkateswarlu, R. Eye gaze estimation from a single image of one eye. In Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; Volume 1, pp. 136–143. [[CrossRef](#)]
71. Ferhat, O.; Vilariño, F. Low Cost Eye Tracking: The Current Panorama. *Comput. Intell. Neurosci.* **2016**, *2016*, 8680541. [[CrossRef](#)] [[PubMed](#)]
72. Wang, X.; Liu, K.; Qian, X. A Survey on Gaze Estimation. In Proceedings of the 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE 2015), Taipei, Taiwan, 24–27 November 2015; pp. 260–267. [[CrossRef](#)]
73. Ki, J.; Kwon, Y.M. 3D Gaze Estimation and Interaction. In Proceedings of the IEEE 3DTV Conference: The True Vision—Capture, Transmission and Display of 3D Video, Istanbul, Turkey, 28–30 May 2008; pp. 373–376. [[CrossRef](#)]
74. Model, D.; Eizenman, M. User-calibration-free remote eye-gaze tracking system with extended tracking range. In Proceedings of the 24th Canadian Conference on Electrical and Computer Engineering (CCECE 2011.), Niagara Falls, ON, Canada, 8–11 May 2011; pp. 001268–001271. [[CrossRef](#)]
75. Pichitwong, W.; Chamnongthai, K. 3-D gaze estimation by stereo gaze direction. In Proceedings of the 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON 2016), Chiang Mai, Thailand, 28 June–1 July 2016; pp. 1–4. [[CrossRef](#)]
76. Zhu, Z.; Ji, Q. Eye gaze tracking under natural head movements. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 918–923. [[CrossRef](#)]
77. Wen, Q.; Bradley, D.; Beeler, T.; Park, S.; Hilliges, O.; Yong, J.; Xu, F. Accurate real-time 3D Gaze Tracking Using a Lightweight Eyeball Calibration. *Comput. Graph. Forum* **2020**, *39*, 475–485. [[CrossRef](#)]
78. Wang, K.; Ji, Q. Real Time Eye Gaze Tracking with 3D Deformable Eye-Face Model. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1003–1011. [[CrossRef](#)]
79. Funes Mora, K.A.; Odobez, J.M. Geometric Generative Gaze Estimation (G3E) for Remote RGB-D Cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1773–1780. [[CrossRef](#)]
80. Li, Y.; Monaghan, D.S.; O'Connor, N.E. Real-Time Gaze Estimation Using a Kinect and a HD Webcam. In Proceedings of the International Conference on Multimedia Modeling, Dublin, Ireland, 6–10 January 2014; pp. 506–517. [[CrossRef](#)]
81. Chen, J.; Ji, Q. 3D gaze estimation with a single camera without IR illumination. In Proceedings of the 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4. [[CrossRef](#)]
82. Sun, L.; Liu, Z.; Sun, M. Real time gaze estimation with a consumer depth camera. *Inf. Sci.* **2015**, *320*, 346–360. [[CrossRef](#)]
83. Xiong, X.; Cai, Q.; Liu, Z.; Zhang, Z. Eye Gaze Tracking Using an RGBD Camera: A Comparison with an RGB Solution. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014), Seattle, WA, USA, 13–17 September 2014; pp. 1113–1121. [[CrossRef](#)]
84. Pieszala, J.; Diaz, G.; Pelz, J.; Speir, J.; Bailey, R. 3D Gaze Point Localization and Visualization Using LiDAR-based 3D reconstructions. In Proceedings of the ACM Symposium on Eye Tracking Research & Applications (ETRA 2016), Charleston, SC, USA, 14–17 March 2016; pp. 201–204. [[CrossRef](#)]
85. Wang, H.; Pi, J.; Qin, T.; Shen, S.; Shi, B.E. SLAM-based localization of 3D gaze using a mobile eye tracker. In Proceedings of the ACM Symposium on Eye Tracking Research & Applications (ETRA 2018), Warsaw, Poland, 14–17 June 2018; pp. 1–5. [[CrossRef](#)]
86. How to Position Participants and the Eye Tracker. Available online: <https://www.tobiipro.com/learnand-support/learn/steps-in-an-eye-tracking-study/run/how-to-position-the-participant-and-the-eye-tracker/> (accessed on 26 December 2019).
87. Hansen, D.W.; Ji, Q. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 478–500. [[CrossRef](#)]

88. Sireesha, M.V.; Vijaya, P.A.; Chellamma, K. A Survey on Gaze Estimation Techniques. In Proceedings of the International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking (VCASAN-2013), Bangalore, India, 17–19 June 2013; pp. 353–361. [[CrossRef](#)]
89. Jiang, J.; Zhou, X.; Chan, S.; Chen, S. Appearance-Based Gaze Tracking: A Brief Review. In Proceedings of the International Conference on Intelligent Robotics and Applications, Shenyang, China, 8–11 August 2019; pp. 629–640. [[CrossRef](#)]
90. Lindén, E.; Sjöstrand, J.; Proutiere, A. Learning to Personalize in Appearance-Based Gaze Tracking. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW 2019), Seoul, Korea, 27–28 October 2019; pp. 1140–1148. [[CrossRef](#)]
91. Al-Rahayfeh, A.; Faezipour, M. Eye Tracking and Head Movement Detection: A State-of-Art Survey. *IEEE J. Transl. Eng. Health Med.* **2013**, *1*, 2100212. [[CrossRef](#)]
92. Tonsen, M.; Steil, J.; Sugano, Y.; Bulling, A. InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–21. [[CrossRef](#)]
93. Wood, E.; Baltrušaitis, T.; Morency, L.P.; Robinson, P.; Bulling, A. Learning an Appearance Based Gaze Estimator from One Million Synthesised Images. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA 2016), Charleston, SC, USA, 14–17 March 2016; pp. 131–138. [[CrossRef](#)]
94. Blignaut, P. Mapping the Pupil-Glint Vector to Gaze Coordinates in a Simple Video-Based Eye Tracker. *J. Eye Mov. Res.* **2014**, *7*, 1–11. [[CrossRef](#)]
95. Cerrolaza, J.; Villanueva, A.; Cabeza, R. Taxonomic Study of Polynomial Regressions Applied to the Calibration of Video-Oculographic Systems. In Proceedings of the Eye Tracking Research and Applications Symposium (ETRA 2008), Savannah, GA, USA, 26–28 March 2008; pp. 259–266. [[CrossRef](#)]
96. Cherif, Z.R.; Nait-Ali, A.; Motsch, J.F.; Krebs, M.O. An adaptive calibration of an infrared light device used for gaze tracking. In Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No.00CH37276), Anchorage, AK, USA, 21–23 May 2002; Volume 2, pp. 1029–1033. [[CrossRef](#)]
97. Jian-nan, C.; Chuang, Z.; Yan-tao, Y.; Yang, L.; Han, Z. Eye Gaze Calculation Based on Nonlinear Polynomial and Generalized Regression Neural Network. In Proceedings of the Fifth International Conference on Natural Computation, Tianjian, China, 14–16 August 2009; Volume 3, pp. 617–623. [[CrossRef](#)]
98. Hennessey, C.; Noureddin, B.; Lawrence, P. A Single Camera Eye-Gaze Tracking System with Free Head Motion. In Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA 2006), San Diego, CA, USA, 27–29 March 2006; pp. 87–94. [[CrossRef](#)]
99. Meyer, A.; Böhme, M.; Martinetz, T.; Barth, E. A Single-Camera Remote Eye Tracker. In Proceedings of the International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems, Kloster Irsee, Germany, 19–21 June 2006; pp. 208–211. [[CrossRef](#)]
100. Jian-nan, C.; Peng-yi, Z.; Si-yi, Z.; Chuang, Z.; Ying, H. Key Techniques of Eye Gaze Tracking Based on Pupil Corneal Reflection. In Proceedings of the WRI Global Congress on Intelligent Systems, Xiamen, China, 19–21 May 2009; Volume 2, pp. 133–138. [[CrossRef](#)]
101. Cai, H.; Yu, H.; Zhou, X.; Liu, H. Robust Gaze Estimation via Normalized Iris Center-Eye Corner Vector. In Proceedings of the International Conference on Intelligent Robotics and Applications, Tokyo, Japan, 22–24 August 2016; Volume 9834, pp. 300–309. [[CrossRef](#)]
102. Wu, H.; Chen, Q.; Wada, T. Conic-based algorithm for visual line estimation from one image. In Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 19 May 2004; pp. 260–265. [[CrossRef](#)]
103. Hansen, D.W.; Pece, A. Eye typing off the shelf. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 2, p. II. [[CrossRef](#)]
104. Yamazoe, H.; Utsumi, A.; Yonezawa, T.; Abe, S. Remote and head-motion-free gaze tracking for real environments with automated head-eye model calibrations. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, AK, USA, 23–28 June 2008; pp. 1–6. [[CrossRef](#)]
105. Huang, S.; Wu, Y.; Hung, W.; Tang, C. Point-of-Regard Measurement via Iris Contour with One Eye from Single Image. In Proceedings of the IEEE International Symposium on Multimedia, Taichung, Taiwan, 13–15 December 2010; pp. 336–341. [[CrossRef](#)]
106. Ohno, T.; Mukawa, N.; Kawato, S. Just Blink Your Eyes: A Head-Free Gaze Tracking System. In Proceedings of the CHI '03 Extended Abstracts on Human Factors in Computing Systems, Ft. Lauderdale, FL, USA, 5–10 April 2003; pp. 950–957. [[CrossRef](#)]
107. Wu, H.; Kitagawa, Y.; Wada, T.; Kato, T.; Chen, Q. Tracking Iris Contour with a 3D Eye-Model for Gaze Estimation. In Proceedings of the Asian Conference on Computer Vision, Tokyo, Japan, 18–22 November 2007; pp. 688–697. [[CrossRef](#)]
108. Wang, Y.; Zhao, T.; Ding, X.; Peng, J.; Bian, J.; Fu, X. Learning a gaze estimator with neighbor selection from large-scale synthetic eye images. *Knowl.-Based Syst.* **2017**, *139*, 41–49. [[CrossRef](#)]
109. Baluja, S.; Pomerleau, D. Non-Intrusive Gaze Tracking Using Artificial Neural Networks. *Tech. Rep.* **1994**, 1–16. Available online: <https://www.aaai.org/Papers/Symposia/Fall/1993/FS-93-04/FS93-04-032.pdf> (accessed on 23 August 2021).
110. Sewell, W.; Komogortsev, O. Real-Time Eye Gaze Tracking with an Unmodified Commodity Webcam Employing a Neural Network. In Proceedings of the CHI '10 Extended Abstracts on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; pp. 3739–3744. [[CrossRef](#)]

111. Cheng, Y.; Lu, F.; Zhang, X. Appearance-Based Gaze Estimation via Evaluation-Guided Asymmetric Regression. In *Computer Vision—ECCV*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 105–121. [[CrossRef](#)]
112. Fischer, T.; Chang, H.J.; Demiris, Y. RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments. In *Computer Vision—ECCV*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 339–357. [[CrossRef](#)]
113. Yu, Y.; Liu, G.; Odobez, J.M. *Deep Multitask Gaze Estimation with a Constrained Landmark-Gaze Model*. *Computer Vision—ECCV*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 456–474. [[CrossRef](#)]
114. Zhang, X.; Sugano, Y.; Fritz, M.; Bulling, A. It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 2299–2308. [[CrossRef](#)]
115. Huang, Y.; Dong, X.; Hao, M. Eye gaze calibration based on support vector regression machine. In Proceedings of the 9th World Congress on Intelligent Control and Automation, Taipei, Taiwan, 21–25 June 2011; pp. 454–456. [[CrossRef](#)]
116. Zhu, Z.; Ji, Q.; Bennett, K.P. Nonlinear Eye Gaze Mapping Function Estimation via Support Vector Regression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; Volume 1, pp. 1132–1135. [[CrossRef](#)]
117. Sugano, Y.; Matsushita, Y.; Sato, Y. Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1821–1828. [[CrossRef](#)]
118. Wang, Y.; Shen, T.; Yuan, G.; Bian, J.; Fu, X. Appearance-based Gaze Estimation using Deep Features and Random Forest Regression. *Knowl.-Based Syst.* **2016**, *110*, 293–301. [[CrossRef](#)]
119. Alnajar, F.; Gevers, T.; Valenti, R.; Ghebreab, S. Calibration-Free Gaze Estimation Using Human Gaze Patterns. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 137–144.
120. Lu, F.; Okabe, T.; Sugano, Y.; Sato, Y. Learning gaze biases with head motion for head pose-free gaze estimation. *Image Vis. Comput.* **2014**, *32*, 169–179. [[CrossRef](#)]
121. Lu, F.; Sugano, Y.; Okabe, T.; Sato, Y. Head pose-free appearance-based gaze sensing via eye image synthesis. In Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba, Japan, 11–15 November 2012; pp. 1008–1011.
122. Lu, F.; Sugano, Y.; Okabe, T.; Sato, T. Gaze Estimation from Eye Appearance: A Head Pose-Free Method via Eye Image Synthesis. *IEEE Trans. Image Process.* **2015**, *24*, 3680–3693. [[CrossRef](#)] [[PubMed](#)]
123. Sugano, Y.; Matsushita, Y.; Sato, Y. Appearance-Based Gaze Estimation Using Visual Saliency. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 329–341. [[CrossRef](#)]
124. Ferhat, O.; Vilariño, F.; Sánchez, F.J. A cheap portable eye-tracker solution for common setups. *J. Eye Mov. Res.* **2014**, *7*, 1–10. [[CrossRef](#)]
125. Williams, O.; Blake, A.; Cipolla, R. Sparse and semi-supervised visual mapping with the S³GP. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 1, pp. 230–237. [[CrossRef](#)]
126. Sesma-Sanchez, L.; Villanueva, A.; Cabeza, R. Gaze Estimation Interpolation Methods Based on Binocular Data. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2235–2243. [[CrossRef](#)]
127. Shih, S.W.; Wu, Y.T.; Liu, J. A calibration-free gaze tracking technique. In Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, 3–7 September 2000; Volume 4, pp. 201–204. [[CrossRef](#)]
128. Sesma, L.; Villanueva, A.; Cabeza, R. Evaluation of Pupil Center-Eye Corner Vector for Gaze Estimation Using a Web Cam. In Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara, CA, USA, 28–30 March 2012; pp. 217–220. [[CrossRef](#)]
129. Guo, Z.; Zhou, Q.; Liu, Z. Appearance-based gaze estimation under slight head motion. *Multimed. Tools Appl.* **2016**, *76*, 2203–2222. [[CrossRef](#)]
130. Tan, K.H.; Kriegman, D.J.; Ahuja, N. Appearance-based eye gaze estimation. In Proceedings of the 6th IEEE Workshop on Applications of Computer Vision, Orlando, FL, USA, 4 December 2002; pp. 191–195. [[CrossRef](#)]
131. Lukander, K. Measuring Gaze Point on Handheld Mobile Devices. In Proceedings of the CHI '04 Extended Abstracts on Human Factors in Computing. Association for Computing Machinery, Vienna, Austria, 24–29 April 2004; p. 1556. [[CrossRef](#)]
132. Martinez, F.; Carbone, A.; Pissaloux, E. Gaze estimation using local features and non-linear regression. In Proceedings of the IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 1961–1964. [[CrossRef](#)]
133. Majaranta, P.; Räihä, K.J. Twenty years of eye typing: Systems and design issues. In Proceedings of the Eye Tracking Research and Applications Symposium, New Orleans, LA, USA, 25–27 March 2002; pp. 15–22. [[CrossRef](#)]
134. Kawato, S.; Tetsutani, N. Detection and tracking of eyes for gaze-camera control. *Image Vis. Comput.* **2004**, *22*, 1031–1038. [[CrossRef](#)]
135. Long, X.; Tonguz, O.K.; Kiderman, A. A High Speed Eye Tracking System with Robust Pupil Center Estimation Algorithm. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007; pp. 3331–3334. [[CrossRef](#)]

136. Alioua, N.; Amine, A.; Rziza, M.; Aboutajdine, D. Eye state analysis using iris detection based on Circular Hough Transform. In Proceedings of the International Conference on Multimedia Computing and Systems, Ouarzazate, Morocco, 7–9 April 2011; pp. 1–5. [\[CrossRef\]](#)
137. Juhong, A.; Treebupachatsakul, T.; Pintavirooj, C. Smart eye-tracking system. In Proceedings of the International Workshop on Advanced Image Technology, Chiang Mai, Thailand, 7–9 January 2018; pp. 1–4. [\[CrossRef\]](#)
138. Söylemez, Ö.F.; Ergen, B. Circular hough transform based eye state detection in human face images. In Proceedings of the Signal Processing and Communications Applications Conference, Haspolat, Turkey, 24–26 April 2013; pp. 1–4. [\[CrossRef\]](#)
139. Kocejko, T.; Bujnowski, A.; Wtorek, J. Eye mouse for disabled. In Proceedings of the Conference on Human System Interactions, Krakow, Poland, 25–27 May 2009; pp. 199–202. [\[CrossRef\]](#)
140. Zhu, J.; Yang, J. Subpixel Eye Gaze Tracking. In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 21 May 2002; pp. 131–136. [\[CrossRef\]](#)
141. Shubhangi, T.; Meshram, P.M.; Rahangdale, C.; Shivhare, P.; Jindal, L. 2015. Eye Gaze Detection Technique to Interact with Computer. *Int. J. Eng. Res. Comput. Sci. Eng.* **2015**, *2*, 92–96.
142. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; Volume 1, p. 1. [\[CrossRef\]](#)
143. Villanueva, A.; Cerrolaza, J.J.; Cabeza, R. Geometry Issues of Gaze Estimation. In *Advances in Human Computer Interaction*; Pinder, S., Ed.; InTechOpen: London, UK, 2008.
144. Świrski, L.; Bulling, A.; Dodgson, N. Robust real-time pupil tracking in highly off-axis images. In Proceedings of the Eye Tracking Research and Applications Symposium (ETRA), Santa Barbara, CA, USA, 28–30 March 2012; pp. 173–176. [\[CrossRef\]](#)
145. Li, D.; Winfield, D.; Parkhurst, D.J. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)—Workshops, San Diego, CA, USA, 21–23 September 2005; p. 79. [\[CrossRef\]](#)
146. Santini, T.; Fuhl, W.; Kasneci, E. PuRe: Robust pupil detection for real-time pervasive eye tracking. *Comput. Vis. Image Underst.* **2018**, *170*, 40–50. [\[CrossRef\]](#)
147. Fuhl, W.; Santini, T.C.; Kübler, T.; Kasneci, E. ElSe: Ellipse selection for robust pupil detection in real-world environments. In Proceedings of the ETRA '16: 2016 Symposium on Eye Tracking Research and Applications, Charleston, SC, USA, 14–17 March 2016; pp. 123–130. [\[CrossRef\]](#)
148. Kassner, M.; Patera, W.; Bulling, A. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In Proceedings of the UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 1151–1160. [\[CrossRef\]](#)
149. Fuhl, W.; Kübler, T.; Sippel, K.; Rosenstiel, W.; Kasneci, E. Excuse: Robust pupil detection in real-world scenarios. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Valletta, Malta, 2–4 September 2015; pp. 39–51. [\[CrossRef\]](#)
150. Fitzgibbon, A.; Pilu, M.; Fisher, R.B. Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Intell.* **1999**, *21*, 476–480. [\[CrossRef\]](#)
151. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [\[CrossRef\]](#)
152. Ramanauskas, N.; Daunys, G.; Dervinis, D. Investigation of Calibration Techniques in Video Based Eye Tracking System. In Proceedings of the 11th international conference on Computers Helping People with Special Needs, Linz, Austria, 9–11 July 2008; pp. 1208–1215. [\[CrossRef\]](#)
153. Hansen, J.P.; Mardanbegi, D.; Biermann, F.; Bækgaard, P. A gaze interactive assembly instruction with pupillometric recording. *Behav. Res. Methods* **2018**, *50*, 1723–1733. [\[CrossRef\]](#)
154. Hansen, D.W.; Hammoud, R.I. An improved likelihood model for eye tracking. *Comput. Vis. Image Underst.* **2007**, *106*, 220–230. [\[CrossRef\]](#)
155. Lemley, J.; Kar, A.; Drimbarean, A.; Corcoran, P. Convolutional Neural Network Implementation for Eye-Gaze Estimation on Low-Quality Consumer Imaging Systems. *IEEE Trans. Consum. Electron.* **2019**, *65*, 179–187. [\[CrossRef\]](#)
156. Arar, N.M.; Gao, H.; Thiran, J.P. A Regression-Based User Calibration Framework for Real-Time Gaze Estimation. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *27*, 2623–2638. [\[CrossRef\]](#)
157. Dubey, N.; Ghosh, S.; Dhall, A. Unsupervised learning of eye gaze representation from the web. In Proceedings of the 2019 International Joint Conference on Neural Networks, Budapest, Hungary, 14–19 July 2019; pp. 1–7. [\[CrossRef\]](#)
158. Yu, Y.; Odobez, J.M. Unsupervised representation learning for gaze estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2020; pp. 7314–7324.
159. Chen, Z.; Deng, D.; Pi, J.; Shi, B.E. Unsupervised Outlier Detection in Appearance-Based Gaze Estimation. In Proceedings of the International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019; pp. 1088–1097. [\[CrossRef\]](#)
160. Akashi, T.; Wakasa, Y.; Tanaka, K.; Karungaru, S.; Fukumi, M. Using Genetic Algorithm for Eye Detection and Tracking in Video Sequence. *J. Syst. Cybern. Inform.* **2007**, *5*, 72–78.
161. Amarnag, S.; Kumaran, R.S.; Gowdy, J.N. Real time eye tracking for human computer interfaces. In Proceedings of the International Conference on Multimedia and Expo. ICME '03, Baltimore, MD, USA, 6–9 July 2003; Volume 3, p. III-557. [\[CrossRef\]](#)

162. Haro, A.; Flickner, M.; Essa, I. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, SC, USA, 15 June 2000; Volume 1, pp. 163–168. [[CrossRef](#)]
163. Coetzer, R.C.; Hancke, G.P. Eye detection for a real-time vehicle driver fatigue monitoring system. In Proceedings of the IEEE Intelligent Vehicles Symposium, Baden-Baden, Germany, 5–9 June 2011; pp. 66–71. [[CrossRef](#)]
164. Sung Ho Park, S.H.; Yoon, H.S.; Park, K.R. Faster R-CNN and Geometric Transformation-Based Detection of Driver's Eyes Using Multiple Near-Infrared Camera Sensors. *Sensors* **2019**, *19*, 197. [[CrossRef](#)] [[PubMed](#)]
165. Gudi, A.; Li, X.; Gemert, J. Efficiency in Real-time Webcam Gaze Tracking. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 529–543.
166. Schneider, T.; Schauerte, B.; Stiefelhagen, R. Manifold Alignment for Person Independent Appearance-Based Gaze Estimation. In Proceedings of the International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 1167–1172. [[CrossRef](#)]
167. Bäck, D. Neural Network Gaze Tracking Using Web Camera. Master's Thesis, Linköping University, Linköping, Sweden, 2005.
168. Wang, J.; Zhang, G.; Shi, J. 2D Gaze Estimation Based on Pupil-Glint Vector Using an Artificial Neural Network. *Appl. Sci.* **2016**, *6*, 174. [[CrossRef](#)]
169. Cho, S.W.; Baek, N.R.; Kim, M.C.; Koo, J.H.; Kim, J.H.; Park, K.R. Face Detection in Nighttime Images Using Visible-Light Camera Sensors with Two-Step Faster Region-Based Convolutional Neural Network. *Sensors* **2018**, *18*, 2995. [[CrossRef](#)] [[PubMed](#)]
170. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
171. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [[CrossRef](#)]
172. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
173. Cortacero, K.; Fischer, T.; Demiris, Y. RT-BENE: A Dataset and Baselines for Real-Time Blink Estimation in Natural Environments. In Proceedings of the International Conference on Computer Vision Workshop, Seoul, Korea, 27–28 October 2019; pp. 1159–1168. [[CrossRef](#)]
174. Xia, Y.; Liang, B. Gaze Estimation Based on Deep Learning Method. In Proceedings of the 4th International Conference on Computer Science and Application Engineering, Sanya, China, 20–22 October 2020; pp. 1–6. [[CrossRef](#)]
175. Ansari, M.F.; Kasprowski, P.; Obetkal, M. Gaze Tracking Using an Unmodified Web Camera and Convolutional Neural Network. *Appl. Sci.* **2021**, *11*, 9068. [[CrossRef](#)]
176. Zhou, X.; Lin, J.; Jiang, J.; Chen, S. Learning a 3d gaze estimator with improved Itracker combined with bidirectional LSTM. In Proceedings of the IEEE International Conference on Multimedia and Expo, Shanghai, China, 8–12 July 2019; pp. 850–855. [[CrossRef](#)]
177. Palmero, C.; Selva, J.; Bagheri, M.A.; Escalera, S. Recurrent CNN for 3d gaze estimation using appearance and shape cues. In Proceedings of the The British Machine Vision Conference, Safety Harbor, FL, USA, 26–28 March 2018.
178. Kim, J.H.; Jeong, J.W. Gaze Estimation in the Dark with Generative Adversarial Networks. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA '20 Adjunct). Association for Computing Machinery, Stuttgart, Germany, 2–5 June 2020; Volume 33, pp. 1–3. [[CrossRef](#)]
179. Kim, J.-H.; Jeong, J.W. Gaze Estimation in a Low-Light Environment with Generative Adversarial Networks. *Sensors* **2020**, *20*, 4935. [[CrossRef](#)] [[PubMed](#)]
180. Wang, K.; Zhao, R.; Ji, Q. A Hierarchical Generative Model for Eye Image Synthesis and Eye Gaze Estimation. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 440–448. [[CrossRef](#)]
181. Wang, K.; Zhao, R.; Su, H.; Ji, Q. Generalizing Eye Tracking with Bayesian Adversarial Learning. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11899–11908. [[CrossRef](#)]
182. He, Z.; Spurr, A.; Zhang, X.; Hilliges, O. Photo-Realistic Monocular Gaze Redirection Using Generative Adversarial Networks. In Proceedings of the International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 6931–6940. [[CrossRef](#)]
183. Wang, K.; Ji, Q. 3D gaze estimation without explicit personal calibration. *Pattern Recognit.* **2018**, *79*, 216–227. [[CrossRef](#)]
184. Khan, S.; Rahmani, H.; Shah, S.A.; Bennamoun, M. A Guide to Convolutional Neural Networks for Computer Vision. *Synth. Lect. Comput. Vis.* **2018**, *8*, 1–207. [[CrossRef](#)]
185. Park, S.; Mello, S.D.; Molchanov, P.; Iqbal, U.; Hilliges, O.; Kautz, J. Few-Shot Adaptive Gaze Estimation. In Proceedings of the International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9367–9376. [[CrossRef](#)]
186. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2019**, *53*, 5455–5516. [[CrossRef](#)]
187. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. In Proceedings of the 22nd ACM international conference on Multimedia. Association for Computing Machinery, Orlando, FL, USA, 3–7 November 2014; pp. 675–678. [[CrossRef](#)]
188. Zhu, W.; Deng, H. Monocular Free-Head 3D Gaze Tracking with Deep Learning and Geometry Constraints. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3162–3171. [[CrossRef](#)]

189. Zhang, Z.; Lian, D.; Gao, S. RGB-D-based gaze point estimation via multi-column CNNs and facial landmarks global optimization. *Vis. Comput.* **2020**, *37*, 1731–1741. [[CrossRef](#)]
190. Zhang, X.; Park, S.; Beeler, T.; Bradley, D.; Tang, S.; Hilliges, O. ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head Pose and Gaze Variation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 365–381. [[CrossRef](#)]
191. George, A.; Routray, A. Real-time eye gaze direction classification using convolutional neural network. In Proceedings of the International Conference on Signal Processing and Communications, Bangalore, India, 12–15 June 2016; pp. 1–5. [[CrossRef](#)]
192. Park, S.; Aksan, E.; Zhang, X.; Hilliges, O. Towards End-to-end Video-based Eye-Tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 747–763. [[CrossRef](#)]
193. Zheng, Y.; Park, S.; Zhang, X.; De Mello, S.; Hilliges, O. Selflearning transformations for improving gaze and head redirection. *arXiv* **2020**, arXiv:2010.12307.
194. Chen, J.; Zhang, J.; Sangineto, E.; Chen, T.; Fan, J.; Sebe, N. Coarseto-fine gaze redirection with numerical and pictorial guidance. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 3665–3674. [[CrossRef](#)]
195. Shrivastava, A.; Pfister, T.; Tuzel, O.; Susskind, J.; Wang, W.; Webb, R. Learning from simulated and unsupervised images through adversarial training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2107–2116.
196. Ahmed, M.; Laskar, R.H. Evaluation of accurate iris center and eye corner localization method in a facial image for gaze estimation. *Multimed. Syst.* **2021**, *27*, 429–448. [[CrossRef](#)]
197. Min-Allah, N.; Jan, F.; Alrashed, S. Pupil detection schemes in human eye: A review. *Multimed. Syst.* **2021**, *27*, 753–777. [[CrossRef](#)]
198. Wang, X.; Zhang, J.; Zhang, H.; Zhao, S.; Liu, H. Vision-based Gaze Estimation: A Review. *IEEE Trans. Cogn. Dev. Syst.* **2021**, *99*, 1–19. [[CrossRef](#)]
199. Park, S.; Zhang, X.; Bulling, A.; Hilliges, O. Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings. In Proceedings of the ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland, 14–17 June 2018; Volume 21, pp. 1–10. [[CrossRef](#)]
200. Bayouhd, K.; Knani, R.; Hamdaoui, F.; Mtibaa, A. A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets. *Vis. Comput.* **2021**, 1–32. [[CrossRef](#)]
201. Feit, A.M.; Williams, S.; Toledo, A.; Paradiso, A.; Kulkarni, H.; Kane, S.; Morris, M.R. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In Proceedings of the CHI '17: CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 1118–1130. [[CrossRef](#)]
202. Eye Tracker Accuracy and Precision. Available online: <https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/what-affects-the-accuracy-and-precision-of-an-eye-tracker/> (accessed on 25 July 2021).
203. Shih, S.W.; Liu, J. A novel approach to 3-D gaze tracking using stereo cameras. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2004**, *34*, 234–245. [[CrossRef](#)]
204. Pérez, A.; Córdoba, M.L.; García, A.; Méndez, R.; Muñoz, M.L.; Pedraza, J.L.; Sánchez, F. A Precise Eye-Gaze Detection and Tracking System. In Proceedings of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Pilsen, Czech Republic, 3–7 February 2003; pp. 105–108.
205. Kang, J.J.; Guestrin, E.D.; Maclean, W.J.; Eizenman, M. Simplifying the cross-ratios method of point-of-gaze estimation. *CMBES Proc.* **2007**, *30*, 1–4.
206. Villanueva, A.; Cabeza, R. A Novel Gaze Estimation System with One Calibration Point. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2008**, *38*, 1123–1138. [[CrossRef](#)] [[PubMed](#)]
207. Ohno, T.; Mukawa, N. A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. In Proceedings of the Symposium on Eye Tracking Research & Applications, San Antonio, TX, USA, 22–24 March 2004; Volume 22, pp. 115–122. [[CrossRef](#)]
208. Hansen, D.W.; Nielsen, M.; Hansen, J.P.; Johansen, A.S.; Stegmann, M.B. Tracking Eyes Using Shape and Appearance. In Proceedings of the IAPR Workshop on Machine Vision Applications, Nara, Japan, 11–13 December 2002; pp. 201–204.
209. Hansen, D.W.; Hansen, J.P.; Nielsen, M.; Johansen, A.S.; Stegmann, M.B. Eye typing using Markov and active appearance models. In Proceedings of the 6th IEEE Workshop on Applications of Computer Vision, Orlando, FL, USA, 4 December 2002; pp. 132–136. [[CrossRef](#)]
210. Nguyen, P.; Fleureau, J.; Chamaret, C.; Guillotel, P. Calibration-free gaze tracking using particle filter. In Proceedings of the IEEE International Conference on Multimedia and Expo, San Jose, CA, USA, 15–19 July 2013; pp. 1–6. [[CrossRef](#)]
211. Zhang, C.; Yao, R.; Cai, J. Efficient eye typing with 9-direction gaze estimation. *Multimed. Tools Appl.* **2018**, *77*, 19679–19696. [[CrossRef](#)]
212. Kar, A.; Corcoran, P. Performance evaluation strategies for eye gaze estimation systems with quantitative metrics and visualizations. *Sensors* **2018**, *18*, 3151. [[CrossRef](#)] [[PubMed](#)]
213. Asteriadis, S.; Soufleros, D.; Karpouzis, K.; Kollias, S. A natural head pose and eye gaze dataset. In Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots, Boston, MA, USA, 6 November 2009; pp. 1–4. [[CrossRef](#)]
214. McMurrough, C.D.; Metsis, V.; Kosmopoulos, D.; Maglogiannis, I.; Makedon, F. A dataset for point of gaze detection using head poses and eye images. *J. Multimodal User Interfaces* **2013**, *7*, 207–215. [[CrossRef](#)]

215. Ponz, V.; Villanueva, A.; Cabeza, R. Dataset for the evaluation of eye detector for gaze estimation. In Proceedings of the ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012; pp. 681–684. [[CrossRef](#)]
216. Smith, B.A.; Yin, Q.; Feiner, S.K.; Nayar, S.K. Gaze locking: Passive eye contact detection for human-object interaction. In Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, St. Andrews, UK, 8–11 October 2013; pp. 271–280. [[CrossRef](#)]
217. Villanueva, A.; Ponz, V.; Sesma-Sanchez, L.; Ariz, M.; Porta, S.; Cabeza, R. Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2013**, *9*, 1–20. [[CrossRef](#)]
218. Weidenbacher, U.; Layher, G.; Strauss, P.M.; Neumann, H. A comprehensive head pose and gaze database. In Proceedings of the 3rd IET International Conference on Intelligent Environments, Ulm, Germany, 24–25 September 2007; pp. 455–458. [[CrossRef](#)]
219. He, Q.; Hong, X.; Chai, X.; Holappa, J.; Zhao, G.; Chen, X.; Pietikäinen, M. OMEG: Oulu multi-pose eye gaze dataset. In Proceedings of the Scandinavian Conference on Image Analysis, Copenhagen, Denmark, 15–17 June 2015; pp. 418–427. [[CrossRef](#)]
220. Schöning, J.; Faion, P.; Heidemann, G.; Krumnack, U. Providing video annotations in multimedia containers for visualization and research. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Santa Rosa, CA, USA, 24–31 March 2017; pp. 650–659. [[CrossRef](#)]
221. Wood, E.; Baltrusaitis, T.; Zhang, X.; Sugano, Y.; Robinson, P.; Bulling, A. Rendering of eyes for eye-shape registration and gaze estimation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3756–3764. [[CrossRef](#)]
222. Cheng, Y.; Zhang, X.; Lu, F.; Sato, Y. Gaze estimation by exploring two-eye asymmetry. *IEEE Trans. Image Process.* **2020**, *29*, 5259–5272. [[CrossRef](#)]
223. Funes Mora, K.A.; Monay, F.; Odobez, J.M. Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras. In Proceedings of the Symposium on Eye Tracking Research and Applications, Safety Harbor, FL, USA, 26–28 March 2014; Volume 26, pp. 255–258. [[CrossRef](#)]
224. Cheng, Y.; Huang, S.; Wang, F.; Qian, C.; Lu, F. A coarse-to-fine adaptive network for appearance-based gaze estimation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 10623–10630. [[CrossRef](#)]
225. Zhao, T.; Yan, Y.; Shehu, I.S.; Fu, X. Image purification networks: Real-time style transfer with semantics through feed-forward synthesis. In Proceedings of the International Joint Conference on Neural Networks, Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7. [[CrossRef](#)]
226. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576. [[CrossRef](#)]
227. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711. [[CrossRef](#)]
228. Selim, A.; Elgharib, M.; Doyle, L. Painting style transfer for head portraits using convolutional neural networks. *ACM Trans. Graph.* **2016**, *35*, 1–18. [[CrossRef](#)]
229. Zhao, T.; Yan, Y.; Shehu, I.S.; Fu, X.; Wang, H. Purifying naturalistic images through a real-time style transfer semantics network. *Eng. Appl. Artif. Intell.* **2019**, *81*, 428–436. [[CrossRef](#)]
230. Zhao, T.; Yan, Y.; Shehu, I.S.; Wei, H.; Fu, X. Image purification through controllable neural style transfer. In Proceedings of the International Conference on Information and Communication Technology Convergence, Jeju, Korea, 17–19 October 2018; pp. 466–471. [[CrossRef](#)]
231. Xiong, Y.; Kim, H.J.; Singh, V. Mixed effects neural networks (menets) with applications to gaze estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7743–7752. [[CrossRef](#)]
232. Yu, Y.; Liu, G.; Odobez, J.M. Improving few-shot user-specific gaze adaptation via gaze redirection synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11937–11946. [[CrossRef](#)]
233. Duchowski, A. A breadth-first survey of eye-tracking applications. *Behav. Res. Methods Instrum. Comput.* **2002**, *34*, 455–470. [[CrossRef](#)] [[PubMed](#)]
234. Armstrong, T.; Olatunji, B.O. Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clin. Psychol. Rev.* **2012**, *32*, 704–723. [[CrossRef](#)] [[PubMed](#)]
235. Kanowski, M.; Rieger, J.W.; Noesselt, T.; Tempelmann, C.; Hinrichs, H. Endoscopic eye tracking system for fMRI. *J. Neurosci. Methods* **2007**, *160*, 10–15. [[CrossRef](#)] [[PubMed](#)]
236. Papageorgiou, E.; Hardiess, G.; Mallot, H.A.; Schiefer, U. Gaze patterns predicting successful collision avoidance in patients with homonymous visual field defects. *Vis. Res.* **2012**, *65*, 25–37. [[CrossRef](#)] [[PubMed](#)]
237. Fu, B.; Yang, R. Display control based on eye gaze estimation. In Proceedings of the 4th International Congress on Image and Signal Processing, Shanghai, China, 15–17 October 2011; Volume 1, pp. 399–403. [[CrossRef](#)]
238. Heidenburg, B.; Lenisa, M.; Wentzel, D.; Malinowski, A. Data mining for gaze tracking system. In Proceedings of the Conference on Human System Interactions, Krakow, Poland, 25–27 May 2008; pp. 680–683. [[CrossRef](#)]
239. Top 8 Eye Tracking Applications in Research. Available online: <https://imotions.com/blog/top-8-applications-eye-tracking-research/> (accessed on 16 February 2020).

240. Chen, M.; Chen, Y.; Yao, Z.; Chen, W.; Lu, Y. Research on eye-gaze tracking network generated by augmented reality application. In Proceedings of the Second International Workshop on Knowledge Discovery and Data Mining, Moscow, Russia, 23–25 January 2009; pp. 594–597. [\[CrossRef\]](#)
241. Danforth, R.; Duchowski, A.; Geist, R.; McAliley, E. A platform for gaze-contingent virtual environments. In *Smart Graphics (Papers from the 2000 AAAI Spring Symposium, Technical Report SS-00-04)*; American Association for Artificial Intelligence: Palo Alto, CA, USA, 2000; pp. 66–70.
242. Nilsson, S. Interaction without gesture or speech—A gaze controlled AR system. In Proceedings of the 17th International Conference on Artificial Reality and Telexistence, Esbjerg, Denmark, 28–30 November 2007; pp. 280–281.
243. Roy, D.; Ghitza, Y.; Bartelma, J.; Kehoe, C. Visual memory augmentation: Using eye gaze as an attention filter. In Proceedings of the 8th International Symposium on Wearable Computers, Arlington, VA, USA, 31 October–3 November 2004; Volume 1, pp. 128–131. [\[CrossRef\]](#)
244. Tateno, K.; Takemura, M.; Ohta, Y. Enhanced eyes for better gaze-awareness in collaborative mixed reality. In Proceedings of the Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality, Vienna, Austria, 5–8 October 2005; pp. 100–103. [\[CrossRef\]](#)
245. Calvi, C.; Porta, M.; Sacchi, D. e5Learning, an e-learning environment based on eye tracking. In Proceedings of the Eighth IEEE International Conference on Advanced Learning Technologies, Santander, Spain, 1–5 July 2008; pp. 376–380. [\[CrossRef\]](#)
246. Georgiou, T.; Demiris, Y. Adaptive user modelling in car racing games using behavioural and physiological data. *User Model. User-Adapt. Interact.* **2017**, *27*, 267–311. [\[CrossRef\]](#) [\[PubMed\]](#)
247. Porta, M.; Ricotti, S.; Perez, C.J. Emotional e-learning through eye tracking. In Proceedings of the IEEE Global Engineering Education Conference, Marrakech, Morocco, 17–20 April 2012; pp. 1–6. [\[CrossRef\]](#)
248. Rajashekar, U.; Van Der Linde, I.; Bovik, A.C.; Cormack, L.K. GAFFE: A gaze-attentive fixation finding engine. *IEEE Trans. Image Process.* **2008**, *17*, 564–573. [\[CrossRef\]](#)
249. Rasouli, A.; Kotseruba, I.; Tsotsos, J.K. Agreeing to cross: How drivers and pedestrians communicate. In Proceedings of the IEEE Intelligent Vehicles Symposium, Los Angeles, CA, USA, 11–14 June 2017; pp. 264–269. [\[CrossRef\]](#)
250. Chen, J.; Luo, N.; Liu, Y.; Liu, L.; Zhang, K.; Kolodziej, J. A hybrid intelligence-aided approach to affect-sensitive e-learning. *Computing* **2016**, *98*, 215–233. [\[CrossRef\]](#)
251. De Luca, A.; Denzel, M.; Hussmann, H. Look into my Eyes! Can you guess my Password? In Proceedings of the 5th Symposium on Usable Privacy and Security, Mountain View, CA, USA, 15–17 July 2009; pp. 1–12. [\[CrossRef\]](#)
252. De Luca, A.; Weiss, R.; Drewes, H. Evaluation of eye-gaze interaction methods for security enhanced PIN-entry. In Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, Adelaide, Australia, 28–30 November 2007; pp. 199–202. [\[CrossRef\]](#)
253. Fookes, C.; Maeder, A.; Sridharan, S.; Mamic, G. Gaze based personal identification. In *Behavioral Biometrics for Human Identification: Intelligent Applications*; Wang, L., Geng, X., Eds.; IGI Global: Hershey, PA, USA, 2010; pp. 237–263. [\[CrossRef\]](#)
254. Kumar, M.; Garfinkel, T.; Boneh, D.; Winograd, T. Reducing shoulder-surfing by using gaze-based password entry. In Proceedings of the 3rd Symposium on Usable Privacy and Security, Pittsburgh, PA, USA, 18–20 July 2007; pp. 13–19. [\[CrossRef\]](#)
255. Weaver, J.; Mock, K.; Hoanca, B. Gaze-based password authentication through automatic clustering of gaze points. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK, USA, 9–12 October 2011; pp. 2749–2754. [\[CrossRef\]](#)
256. Klaib, A.F.; Alsrehin, N.O.; Melhem, W.Y.; Bashtawi, H.O. IoT Smart Home Using Eye Tracking and Voice Interfaces for Elderly and Special Needs People. *J. Commun.* **2019**, *14*, 614–621. [\[CrossRef\]](#)
257. Wu, M.; Louw, T.; Lahijanian, M.; Ruan, W.; Huang, X.; Merat, N.; Kwiatkowska, M. Gaze-based intention anticipation over driving manoeuvres in semi-autonomous vehicles. In Proceedings of the IEEE/RISJ International Conference on Intelligent Robots and Systems, Macau, China, 3–8 November 2019; pp. 6210–6216. [\[CrossRef\]](#)
258. Subramanian, M.; Songur, N.; Adjei, D.; Orlov, P.; Faisal, A.A. A.Eye Drive: Gaze-based semi-autonomous wheelchair interface. In Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Berlin, Germany, 23–27 July 2019; pp. 5967–5970. [\[CrossRef\]](#)
259. Kamp, J.; Sundstedt, V. Gaze and Voice controlled drawing. In Proceedings of the 1st Conference on Novel Gaze-Controlled Applications (NGCA '11), Karlskrona Sweden, 26–27 May 2011; Volume 9, pp. 1–8. [\[CrossRef\]](#)
260. Scalera, L.; Seriani, S.; Gallina, P.; Lentini, M.; Gasparetto, A. Human–Robot Interaction through Eye Tracking for Artistic Drawing. *Robotics* **2021**, *10*, 54. [\[CrossRef\]](#)
261. Santella, A.; DeCarlo, D. Abstracted painterly renderings using eye-tracking data. In Proceedings of the 2nd International Symposium on Non-Photorealistic Animation and Rendering (NPAR '02), Annecy, France, 3–5 June 2002; p. 75. [\[CrossRef\]](#)
262. Scalera, L.; Seriani, S.; Gasparetto, A.; Gallina, P. A Novel Robotic System for Painting with Eyes. In *Advances in Italian Mechanism Science. IFToMM ITALY 2020. Mechanisms and Machine Science*; Niola, V., Gasparetto, A., Eds.; Springer: Cham, Switzerland, 2020; Volume 91. [\[CrossRef\]](#)
263. Lallé, S.; Conati, C.; Carenini, G. Predicting Confusion in Information Visualization from Eye Tracking and Interaction Data. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16), New York, NY, USA, 9–15 July 2016; pp. 2529–2535.

264. Salminen, J.; Jansen, B.J.; An, J.; Jung, S.G.; Nielsen, L.; Kwak, H. Fixation and Confusion: Investigating Eye-tracking Participants' Exposure to Information in Personas. In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA, 11–15 March 2018; pp. 110–119. [\[CrossRef\]](#)
265. Sims, S.D.; Putnam, V.; Conati, C. Predicting confusion from eye-tracking data with recurrent neural networks. *arXiv* **2019**, arXiv:1906.11211.
266. Hayhoe, M.M.; Matthis, J.S. Control of gaze in natural environments: Effects of rewards and costs, uncertainty and memory in target selection. *Interface Focus* **2018**, *8*, 1–7. [\[CrossRef\]](#)
267. Jording, M.; Engemann, D.; Eckert, H.; Bente, G.; Vogeley, K. Distinguishing Social from Private Intentions Through the Passive Observation of Gaze Cues. *Front. Hum. Neurosci.* **2019**, *13*, 442. [\[CrossRef\]](#) [\[PubMed\]](#)
268. Uma, S.; Eswari, R. Accident prevention and safety assistance using IOT and machine learning. *J. Reliab. Intell. Environ.* **2021**, 1–25. [\[CrossRef\]](#)
269. Shimauchi, T.; Sakurai, K.; Tate, L.; Tamura, H. Gaze-Based Vehicle Driving Evaluation of System with an Actual Vehicle at an Intersection with a Traffic Light. *Electronics* **2020**, *9*, 1408. [\[CrossRef\]](#)
270. Ledezma, A.; Zamora, V.; Sipele, Ó.; Sesmero, M.P.; Sanchis, A. Implementing a Gaze Tracking Algorithm for Improving Advanced Driver Assistance Systems. *Electronics* **2021**, *10*, 1480. [\[CrossRef\]](#)
271. Berkovsky, S.; Taib, R.; Koprinska, I.; Wang, E.; Zeng, Y.; Li, J.; Kleitman, S. Detecting Personality Traits Using Eye-Tracking Data. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; Volume 221, p. 12. [\[CrossRef\]](#)
272. Brunyé, T.T.; Drew, T.; Weaver, D.L.; Elmore, J.G. A review of eye tracking for understanding and improving diagnostic interpretation. *Cogn. Res. Princ. Implic.* **2019**, *4*, 7. [\[CrossRef\]](#) [\[PubMed\]](#)
273. Maurage, P.; Masson, N.; Bollen, Z.; D'Hondt, F. Eye tracking correlates of acute alcohol consumption: A systematic and critical review. *Neurosci. Biobehav. Rev.* **2019**, *108*, 400–422. [\[CrossRef\]](#) [\[PubMed\]](#)
274. Iannizzotto, G.; Nucita, A.; Fabio, R.A.; Capri, T.; Lo Bello, L. Remote Eye-Tracking for Cognitive Telerehabilitation and Interactive School Tasks in Times of COVID-19. *Information* **2020**, *11*, 296. [\[CrossRef\]](#)
275. Jin, N.; Mavromatis, S.; Sequeira, J.; Curcio, S. A Robust Method of Eye Torsion Measurement for Medical Applications. *Information* **2020**, *11*, 408. [\[CrossRef\]](#)
276. Maimon-Mor, R.O.; Fernandez-Quesada, J.; Zito, G.A.; Konnaris, C.; Dziemian, S.; Faisal, A.A. Towards free 3D end-point control for robotic-assisted human reaching using binocular eye tracking. In Proceedings of the International Conference on Rehabilitation Robotics (ICORR), London, UK, 17–20 July 2017; pp. 1049–1054. [\[CrossRef\]](#)
277. Palinko, O.; Sciutti, A.; Wakita, Y.; Matsumoto, Y.; Sandini, G. If looks could kill: Humanoid robots play a gaze-based social game with humans. In Proceedings of the 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), Cancun, Mexico, 15–17 November 2016; pp. 905–910. [\[CrossRef\]](#)
278. Schwab, D.; Fejza, A.; Vial, L.; Robert, Y. The GazePlay Project: Open and Free Eye-Trackers Games and a Community for People with Multiple Disabilities. In Proceedings of the CCHP: International Conference on Computers Helping People with Special Needs, Linz, Austria, 11–13 July 2018; pp. 254–261. [\[CrossRef\]](#)
279. Wöhle, L.; Gebhard, M. Towards Robust Robot Control in Cartesian Space Using an Infrastructureless Head- and Eye-Gaze Interface. *Sensors* **2021**, *21*, 1798. [\[CrossRef\]](#)
280. Bozkir, E.; Günlü, O.; Fuhl, W.; Schaefer, R.F.; Kasneci, E. Differential Privacy for Eye Tracking with Temporal Correlations. *PLoS ONE* **2020**, *16*, e0255979. [\[CrossRef\]](#) [\[PubMed\]](#)
281. Bozkir, E.; Ünal, A.B.; Akgün, M.; Kasneci, E.; Pfeifer, N. Privacy Preserving Gaze Estimation using Synthetic Images via a Randomized Encoding Based Framework. In Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA 2020), Stuttgart, Germany, 2–5 June 2020; Volume 21, pp. 1–5. [\[CrossRef\]](#)
282. Liu, A.; Xia, L.; Duchowski, A.; Bailey, R.; Holmqvist, K.; Jain, E. Differential Privacy for Eye-Tracking Data. In Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA 2019), Denver, CO, USA, 25–28 June 2019; Volume 28, p. 10. [\[CrossRef\]](#)
283. Steil, J.; Hagedstedt, I.; Huang, M.X.; Bulling, A. Privacy-Aware Eye Tracking Using Differential Privacy. In Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications (ETRA 2019), Denver, CO, USA, 25–28 June 2019; Volume 27, pp. 1–9. [\[CrossRef\]](#)
284. Abdrabou, Y.; Khamis, M.; Eisa, R.M.; Ismail, S.; Elmougy, A. Just Gaze and Wave: Exploring the Use of Gaze and Gestures for Shoulder-Surfing Resilient Authentication. In Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, Denver, CO, USA, 25–28 June 2019; Volume 29, p. 10. [\[CrossRef\]](#)
285. Khamis, M.; Alt, F.; Hassib, M.; Zezschwitz, E.V.; Hasholzner, R.; Bulling, A. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 2156–2164. [\[CrossRef\]](#)
286. Khamis, M.; Hasholzner, R.; Bulling, A.; Alt, F. GTmoPass: Two-Factor Authentication on Public Displays Using Gaze-Touch Passwords and Personal Mobile Devices. In Proceedings of the 6th ACM International Symposium on Pervasive Displays, Lugano, Switzerland, 7–9 June 2017; Volume 8, p. 9. [\[CrossRef\]](#)

287. Khamis, M.; Hassib, M.; Zezschwitz, E.V.; Bulling, A.; Alt, F. GazeTouchPIN: Protecting Sensitive Data on Mobile Devices Using Secure Multimodal Authentication. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; pp. 446–450. [CrossRef]
288. Mathis, F.; Vaniea, K.; Williamson, J.; Khamis, M. RubikAuth: Fast and Secure Authentication in Virtual Reality. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI 2020), Honolulu, HI, USA, 25–30 April 2020; pp. 1–9. [CrossRef]
289. Top 12 Eye Tracking Hardware Companies. Available online: <https://imotions.com/blog/top-eyetracking-hardware-companies/> (accessed on 23 August 2021).
290. Tobii. Available online: <https://www.tobii.com/> (accessed on 23 August 2021).
291. SensoMotoric. Available online: <http://www.smivision.com/> (accessed on 23 August 2021).
292. EyeLink. Available online: <http://www.eyelinkinfo.com/> (accessed on 23 August 2021).
293. NNET. Available online: <https://userweb.cs.txstate.edu/~jok11/nnet.html> (accessed on 23 August 2021).
294. EyeTab. Available online: <https://github.com/errollw/EyeTab> (accessed on 23 August 2021).
295. Opengazer. Available online: <http://www.inference.phy.cam.ac.uk/opengazer/> (accessed on 23 August 2021).
296. TurkerGaze. Available online: <https://github.com/PrincetonVision/TurkerGaze> (accessed on 23 August 2021).
297. Camgaze. Available online: <https://github.com/wallarelvo/camgaze> (accessed on 23 August 2021).
298. ITU. Available online: <https://github.com/devinbarry/GazeTracker> (accessed on 23 August 2021).
299. CVC ET. Available online: <https://github.com/tiendan/> (accessed on 23 August 2021).
300. Xlabs. Available online: <https://xlabsgaze.com/> (accessed on 23 August 2021).
301. Gazepointer. Available online: <https://sourceforge.net/projects/gazepointer/> (accessed on 23 August 2021).
302. MyEye. Available online: <https://myeye.jimdofree.com/> (accessed on 23 August 2021).
303. NetGazer. Available online: <http://sourceforge.net/projects/netgazer/> (accessed on 23 August 2021).
304. OpenEyes. Available online: <http://thirtysixthspan.com/openEyes/software.html> (accessed on 23 August 2021).
305. Ogama. Available online: <http://www.ogama.net/> (accessed on 23 August 2021).
306. GazeParser. Available online: <http://gazeparser.sourceforge.net/> (accessed on 23 August 2021).
307. Pygaze. Available online: <http://www.pygaze.org/> (accessed on 23 August 2021).
308. Paperswithcodes. Available online: <https://www.paperswithcode.com/task/gaze-estimation?page=2> (accessed on 23 August 2021).