Pattern Recognition 132 (2022) 108944

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

In the eye of the beholder: A survey of gaze tracking techniques

Jiahui Liu[a,*], Jiannan Chi[a,b], Huijie Yang[a], Xucheng Yin[c]

[a] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[b] Engineering Research Center of Intelligence Perception and Autonomous Control, Ministry of Education, Beijing 100124, China
[c] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

ARTICLE INFO

Article history:
Received 10 November 2021
Revised 12 June 2022
Accepted 25 July 2022
Available online 26 July 2022

Keywords:
Gaze estimation
eye features
appearance-based
personal calibration
head motion

ABSTRACT

Gaze tracking estimates and tracks the user's gaze by analyzing facial or eye features, it is an important way to realize automated vision-based interaction. This paper introduces the visual information used in gaze tracking, and discusses the commonly used gaze estimation methods and their research dynamics, including: 2D mapping-based methods, 3D model-based methods, and appearance-based methods. In this way, some key issues that need to be solved in these methods are considered, and their research trends are discussed. Their characteristics in system configuration, personal calibration, head motion, gaze accuracy and robustness are also compared. Finally, the applications of gaze tracking techniques are analyzed from various application factors and fields. This paper reviews the latest development of gaze tracking, focuses more on various gaze tracking algorithms and their existing challenges. The development trends of gaze tracking are prospected, which provides ideas for future theoretical research and practical applications.

1. Introduction

Gaze tracking captures the visual information from the user's face and eyes, and measures the user's attention on any object, so as to understand the user's desires and needs [1]. Due to its simple operation and good performance, it is widely used in human-computer interaction (HCI), medical diagnosis, intelligent transportation, virtual reality, and human factors analysis. Gaze tracking systems can be divided into intrusive and non-intrusive systems. Although the early intrusive systems have high accuracy, they require direct contact between the eye and the device, which causes great interference to the user. With the development of computer vision and artificial intelligence, the system not only needs to meet the accuracy requirements, but also pays attention to the user experience. Therefore, non-intrusive gaze tracking systems are the mainstream. According to the different ways of use, current gaze tracking systems can be divided into head-mounted and remote systems. Head-mounted gaze tracking systems obtain eye images through near-eye cameras, and are mainly used in smart glasses and helmets. Since the device moves with the user's head, the impact of head motion is small, which only includes the relative sliding between the user's head and the device. Remote gaze tracking systems use the external camera to obtain face images, and estimate the gaze direction or the point-of-regard (POR) by analyzing face or eye images. They have less interference and give users a better sense of experience, so they have promising development prospects in mobile devices.

The research on gaze estimation methods is mainly divided into three categories: 2D mapping-based methods, 3D model-based methods, and appearance-based methods. 2D mapping-based methods determine the mapping model of eye features and POR through personal calibration. Thus, the corresponding POR can be estimated according to the new features. 3D model-based methods estimate the 3D line-of-sight (LOS) based on the eyeball structure and geometric imaging model. Both 2D and 3D methods specifically study the relationship between the eye features and the gaze in the eye movement state. Although they often have certain requirements on the selection of system cameras and light sources, the effects of individual differences and head motion are easier to deal with, and the accuracy can reach 0.5° [2,3]. Appearance-based methods take face or eye images as input to train the mapping model between the appearance and the gaze, so as to determine the corresponding gaze based on the new appearance. They are robust due to the use of a large number of statistical samples, and they have lower system requirements since they do not need to extract the specific eye features. But the appearance is heavily influenced by individual differences, head motion, etc., making unconstrained gaze estimation a challenging task [4].

* Corresponding author.
E-mail addresses: ustbljh@ustb.edu.cn (J. Liu), ustbjnc@ustb.edu.cn (J. Chi), 15652760808@163.com (H. Yang), xuchengyin@ustb.edu.cn (X. Yin).

https://doi.org/10.1016/j.patcog.2022.108944
0031-3203/

At present, there have been some review papers on gaze tracking, which provide guidance for its technical research and development applications [1,5–7]. Hansen and Ji [1] presented a detailed review on video-based eye detection and gaze tracking. Cristina and Camilleri [5] focused on emerging passive and unobtrusive video-based eye-gaze tracking methods. They identified five challenges for pervasive eye-gaze tracking, and provided the avenues to address them. Cheng et al. [6] reviewed the appearance-based methods with deep learning from feature extraction, neural network architecture design, personal calibration as well as device and platform, and discussed the data pre-processing and post-processing methods for gaze estimation. Different from these review papers, this paper focuses more on a comprehensive overview of various gaze tracking methods from algorithmic principles, discusses the key issues faced by different methods and the avenues to break through the technical bottlenecks, and analyzes the application factors and possible application fields of different gaze tracking methods, which is convenient for researchers with specific application requirements.

This paper is organized as follows: Section 2 discusses the visual information required by gaze tracking, and analyzes its internal connection with the gaze. Section 3 gives a detailed overview of different gaze estimation methods, including: 2D mapping-based methods, 3D model-based methods, and appearance-based methods, and lists their key issues and research trends through a comparative analysis. We compare their characteristics from several aspects in Section 4, and give a discussion of the practical applications of gaze tracking techniques in Section 5. Finally, this paper is concluded and the development trends of gaze tracking are prospected in the conclusion.

2. Visual information in gaze tracking

2.1. Eyeball structure

The human eye is the sensory organ that obtains the most information and can intuitively reflect the gaze direction. As Fig. 1 shows, the eyeball is approximately a spherical shape with a radius of about 12mm, which consists of the eyeball wall and transparent contents. The outer membrane of the eyeball wall is composed of the cornea and the sclera. The iris is on the tunica vasculosa, its center has a circular aperture, called pupil, which regulates the amount of light coming into the eye. The inner membrane is the retina containing a large number of photoreceptor cells. The contents are composed of aqueous humor, lens and vitreous. The light entering the eye needs to traverse a series of eye optical media such as the cornea, aqueous humor, lens, and vitreous body, and be reflected and refracted at each media boundary before reaching the retina. The special region with the most acute vision on the retina is the fovea, which is used to perceive the details of the scene. The fovea is not exactly located on the symmetry axis of the eyeball, but forms the visual axis (VA) with the cornea center. The symmetry axis of the eyeball is the optical axis (OA), and there is a deviation angle between the OA and the VA, called the kappa angle, which is usually around 5° [8].

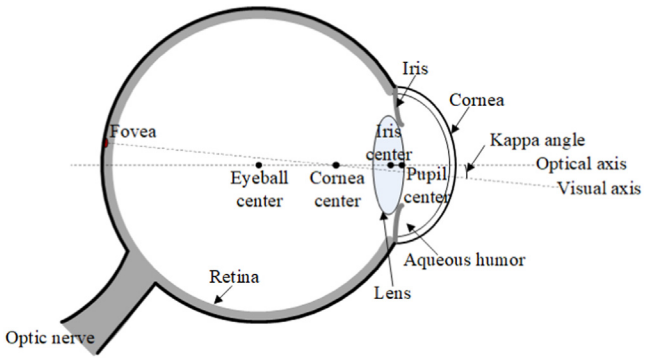


Fig. 1. Basic structure of the eyeball.

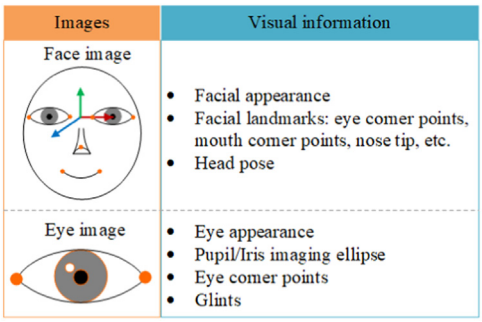


Fig. 2. Visual information in face or eye images.

2.2. Visual information

Gaze tracking refers to estimating the VA and determining the POR based on the observed scene. Some researchers also use the OA instead of the VA to simplify the gaze estimation [9]. The visual information that is necessary to be extracted from face or eye images to estimate the VA or OA of the eye, which can be summarized as shown in Fig. 2.

(1)*Facial landmarks/appearance*: The LOS is not only related to eye movement, but also related to head pose. The detected facial landmarks in the face image can be used to determine the head pose for gaze estimation. In addition, the face image contains some appearance information that is beneficial for gaze prediction [10].

(2)*Eye appearance*: During image acquisition, the features visible in some images may only be the eye region due to body pose, head pose, etc.. From the statistical perspective, when the head poses of users are relatively consistent, the appearance of their eyes looking at the same POR has a certain similarity. Therefore, the gaze direction can be predicted based on the similar eye appearance with known gaze direction.

(3)*Pupil/iris imaging ellipse*: When the user looks in different directions, the OA direction of the eye will change accordingly. The OA is the symmetry axis of the eyeball, which is always perpendicular to the pupil or iris plane. The kappa angle remains unchanged, therefore, the visible pupil/iris imaging ellipse can reflect the gaze information.

(4)*Eye corner points*: Except as one of the features for determining the head pose, eye corner points can be chosen as a reference point to form a vector with the feature that changes with the eye movement since they do not follow the eye movement. Eye corner points are also a key feature used to crop the eye image [11,12].

(5)*Glints*: The glint is formed by the corneal reflection of the light source, it can be used as a reference point for analyzing eye movement with the head fixed. When multiple light sources are used, multiple glints can not only be used to calculate some gaze-related parameters, but also improve the impact of head motion.

3. Gaze estimation methods

Gaze estimation methods are generally divided into 2D mapping-based methods, 3D model-based methods, and appearance-based methods. As Fig. 3 shows, they have different gaze estimation models that establish the relationship between visual information and 2D/3D gaze, which are discussed in detail as follows.

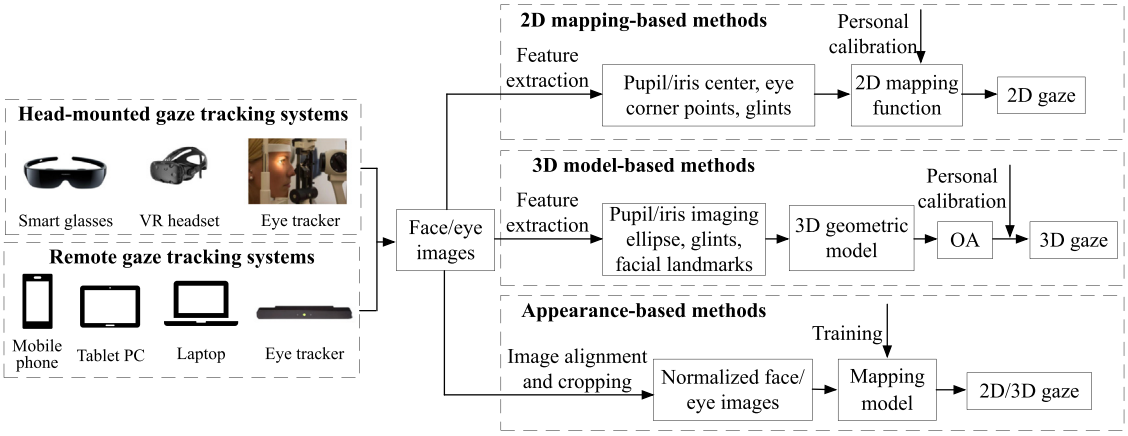


Fig. 3. Gaze estimation models in different methods.

3.1. 2D mapping-based gaze estimation methods

3.1.1. 2D gaze estimation algorithms

2D mapping-based methods are based on the invariant features of eye movement, and estimating the gaze by simulating the 2D mapping function between the variation features of eye movement and the POR. Common methods are: pupil/iris-corner technique (PCT/ICT)-based methods, pupil/iris-corneal reflection technique (PCRT/ICRT)-based methods, cross-ratio (CR)-based methods, and homography normalization (HN)-based methods. Besides, some other methods have emerged to improve 2D gaze estimation performance.

(1) PCT/ICT-based methods

The PCT/ICT-based method uses the mapping function between the vector formed by pupil/iris center and eye corner point and the POR. When the head is fixed, the eye corner position is fixed, and the pupil/iris feature changes with the gaze. Thus, the vector pointing from the pupil/iris center to the eye corner point (PCECV/ICECV) can be used as a 2D variation feature that contains the gaze information to construct a mapping function with the POR [13–15]. The user needs to stare at multiple calibration points preset on the screen in sequence during personal calibration, as shown in Fig. 4. In this way, multiple sets of corresponding vectors and PORs are obtained to regress the unknown mapping function coefficients a_i, b_i. When the user uses the system next time, the corresponding POR can be directly estimated using the vector extracted from the newly captured image according to the determined mapping function.

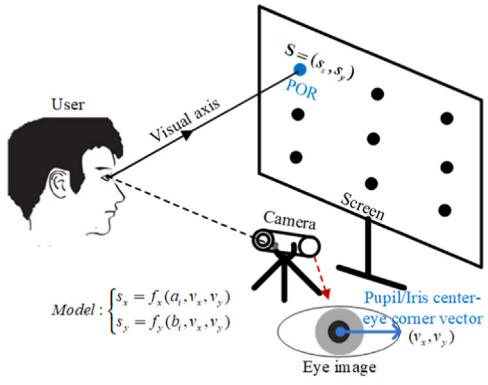


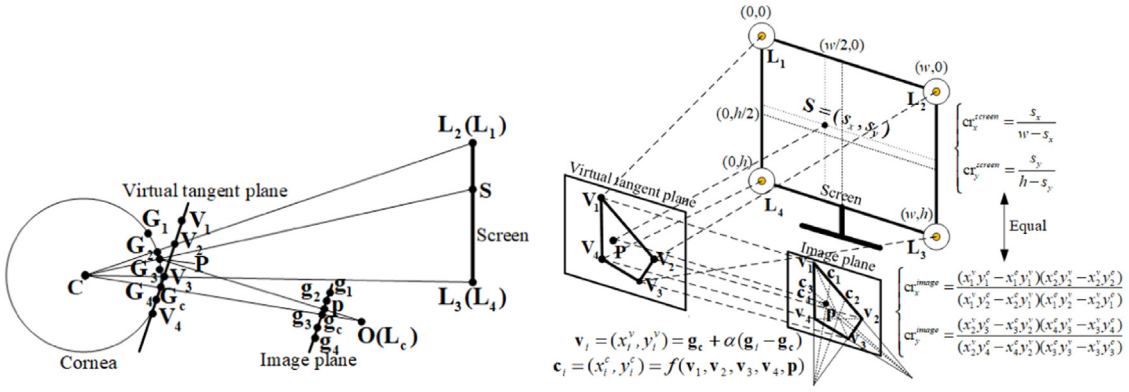
Fig. 4. The technical principle of PCT/ICT-based methods.

The simplest system required for the PCT/ICT-based method is a single camera, but personal calibration is time-consuming due to the calculation of mapping function coefficients by staring at multiple calibration points. And the calibration process often requires keeping the head fixed to ensure the accuracy. Although Cheung and Peng [16] proposed an adaptive-weighted facial features embedded in the POSIT algorithm to improve the head pose estimation and compensated the POR using the head motion information, the gaze accuracy under slight head motion decreases by nearly 1° compared with no head motion. Hu et al. [17] adjusted the feature vector based on projective mapping correction to reduce the impact of head motion, and created a support vector regression (SVR) model between head motion and POR deviation to compensate the POR. But the average gaze accuracy is only 2.67°.

(2) PCRT/ICRT-based methods

The PCRT/ICRT-based method is the most typical 2D gaze estimation method, which uses the glint formed by the corneal reflection of light source. When the head is fixed, the glint does not move with the eye movement since the cornea is approximately spherical, then taking the glint as the reference point, the vector pointing from the pupil/iris center to the glint (PCGV/ICGV) will change with the gaze. Therefore, PCGV/ICGV can be used as a 2D variation feature to establish a mapping function with the POR [18]. Its mapping function is similar to the PCT/ICT-based method and usually described by a set of polynomials [19,20]. After the coefficients a_i, b_i are calibrated, the POR can be estimated by substituting the vector extracted from the newly captured image into the mapping function.

The PCRT/ICRT-based method uses the glint, which determines single-camera-single-light-source (SCSLS) is the simplest system using this method. Koshikawa et al. [21] proved that when the glint is far from the pupil center, the POR cannot be estimated correctly. Blignaut et al. [19] compared different 2D mapping functions, and concluded that the accuracy can reach 0.5° when the number of calibration points is more than 14. Cerrolaza et al. [20] also analyzed more than 400,000 mapping functions, indicating that higher-order polynomials cannot significantly improve the performance. They demonstrated the PCRT-based method in a SCSLS system is highly restricted to the head motion. When the user's head deviates from the calibration position, the accuracy would decrease significantly [22]. To deal with this issue, Zhang et al. [23] derived a compensation model in the head motion state, and corrected the gaze parameters caused by head motion back to the corresponding parameters at the calibration position. But the geometric approximation also introduces errors. Hu et al. [24] trained an error function by the BP neural network based on



(a) Corneal reflection and virtual projection (b) CR invariance of screen and image plane

Fig. 5. Basic principle of CR-based method.

particle swarm optimization to compensate the error caused by the head motion, but the analyzed head motion is quite limited. It can be seen that the PCRT/ICRT-based method can improve the effect of head motion through compensation or feedback, but the user's head is still required to be fixed at the calibration position to ensure the accuracy. Moreover, this method still requires complicated calibration for coefficient estimation.

Cerrolaza et al. [20] pointed out that it is necessary to use two light sources and normalize the PCGV to ensure the robustness under head motion. Some researchers have used a single-camera-multi-light-source (SCMLS) system to carry out the PCRT/ICRT-based method [25–29]. Wang et al. [26] detected the pupil and glints using the circular ring rays location method, and established the mapping function between the PCGV and the POR using an artificial neural network. When utilizing 16 calibration points on the wearable gaze tracking system, the gaze accuracy was within 0.5°. Mestre et al. [27] analyzed the potential benefits of 2/4/6/8/12 corneal reflections for gaze tracking, and confirmed that the PCGV normalization is beneficial to improve the accuracy. Sesma-Sanchez et al. [28] and Zhang et al. [29] dealt with the head motion by adding the relationship between two glints into the model. It can be seen that compared with SCSLS system, the PCRT/ICRT-based method using a SCMLS system can simplify the mapping model and reduce the required calibration points while ensuring the gaze accuracy. It can also increase the robustness under a certain range of head motion.

A few researchers also used multi-camera systems to study the PCRT/ICRT-based method. Some compensated for the impact of head motion[30], and some are pursuing higher accuracy. The use of multiple cameras can extract the eye features with high precision, ensuring the accuracy of mapping model calibration and subsequent gaze estimation.

(3) CR-based methods

The CR-based method exploits the invariance property of cross-ratios in projective transformations. As Fig. 5(a) shows, the system generally contains five light sources, $\mathbf{L_c}$ is placed at the OA of the camera to determine a virtual plane tangent to its corneal reflection point, and $\mathbf{L_i}(i = 1, 2, 3, 4)$ are respectively attached to the four corners of the screen. They have virtual projection points $\mathbf{V_i}$ on the virtual tangent plane, and their imaging points $\mathbf{v_i}$ can be calculated by the glints, where the scale factor α is calibrated, as shown in Fig. 5(b). According to the CR projective invariance, the CR on a certain side of the polygon formed by $\mathbf{L_i}$ is equal to the CR of the points on the corresponding side of the polygon formed by $\mathbf{v_i}$ on the basis of taking the virtual tangent plane as the medium. The pupil center **P** is imaged as **p**, assuming that **p** is the projection of the gaze point **S** on the image plane, thus **S** can be determined.

The CR-based method exploits an uncalibrated camera to estimate the POR under head motion. It only needs to know the positions of light sources and the screen size. However, the CR-based method has two deficiencies: (1) The CR mapping point of pupil center on the screen is regarded as the POR, which is essentially the intersection point of the OA and the screen, rather than the actual POR; (2) The premise of CR calculation is that the corneal reflection points of four on-screen light sources are coplanar with the pupil center, but they are only assumed to be coplanar. To address these deficiencies, Coutinho et al. [31] used an eye model to correct for the deviation of the VA from the OA, and calibrated the scale factor of virtual projection to adjust the glint position to the virtual plane. Cheng et al. [32] described the relationship between the reflections of LEDs and their virtual point using a dynamic matrix. But these methods involve complicated calibration. Arar et al. [33] introduced a personal calibration method based on regularized least squares regression and used an adaptive fusion scheme to determine the POR of both eyes, achieving higher accuracy when there are fewer calibration points. They also proposed a weighted regression based calibration method for a more convenient and user-friendly calibration process [34]. In addition, to improve the restriction of head motion using multiple light sources, Sasaki et al. [35] used a developed polarization camera system to detect the screen area on a cornea without near-infrared illumination. They also investigated the gaze accuracy of their proposed method under various illumination and display conditions [36]. But these methods have certain system requirements.

(4) HN-based methods

The HN-based method employs two homography transformations of the two projections from image to cornea and from cornea to screen,to estimate the POR based on the assumption that the pupil center and the corneal reflection plane are coplanar [37,38]. The system usually contains a camera and four light sources attached to the four corners of the screen, as shown in Fig. 6. Since the system is uncalibrated, a normalized plane Π_N is defined instead of the unknown Π_C. The homography matrix from the image to the normalized plane, represented by $\mathbf{H_I^N}$, is calculated by $\mathbf{g_i}$ and the corners of Π_N. And the homography matrix from the normalized plane to the screen, represented by $\mathbf{H_N^S}$, is determined through personal calibration. Therefore, the projection matrix from the image to the screen can be calculated. The corresponding POR on the screen can then be determined using the pupil center **p**.

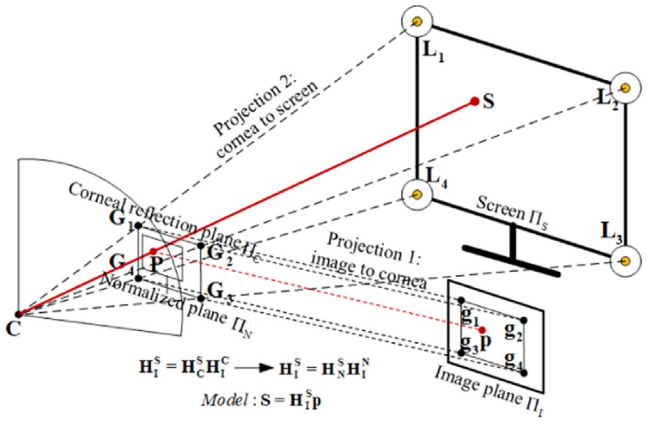


Fig. 6. Basic principle of HN-based method.

Compared with the CR-based method, the HN-based method does not require an extra light source except for the light sources attached to four screen corners, nor does it need the known screen size owing to the introduced normalized space [39]. The homography matrix normalizes the head motion before POR estimation and compensates for the kappa angle, so the HN-based method is more robust and less sensitive to head pose changes than the PCRT/ICRT-based method [40]. Kim et al. [41] proposed an elliptical error trajectory model to compensate the device rotation error caused by the kappa angle and the camera location change. However, at least four calibration points are needed to calculate $\mathbf{H}_N^S$. Choi et al. [39] constructed a calibration database through single-point calibration, and automatically found a mapping function best suited for the subject from the database without calculating the subject-specific parameters. But the head motions of the subjects are restrained. To reduce the requirements for the corneal reflection of four light sources, Ma et al. [42] proposed three alternative geometric transforms that are adaptive to the number of corneal reflections, so as to estimate the POR using two or three corneal reflections. Shin et al. [43] calculated the virtual glint to replace the actual glint by utilizing the mathematical and geometrical principles between the light source and the glint. Although these methods reduce the number of light sources, the accuracy is very sensitive to the imaging of actual light sources. Luo et al. [2] established a HN-based mapping model using pin-hole imaging and similarity conditions of triangles under a SCSLS system, but it requires the user to keep the head as still as possible.

(5) Other 2D gaze estimation methods

In addition to the above-mentioned methods, there are also many other methods that: 1) *Establish 2D mapping models between the pupil parameters and the POR* [44–46]; 2) *Simplify the model to linear* [44,47,48]; 3) *Use simple system configuration or few samples* [47,49]; 4) *Consider different application scenarios such as HCI* [50,51]*, Intelligent transportation* [52]*, Virtual reality* [48,53]. For example, Han et al. [46] used the UNet architecture to extract the parameters of pupil imaging ellipse, and then calibrated a gaze difference network for gaze estimation, which took the difference of pupil ellipse parameters of two consecutive image frames as input and the difference of their PORs as the output. Yoon et al. [52] proposed a driver's gaze estimation method suitable for actual vehicle environment. When the corneal reflection cannot be detected due to a driver's excessive head rotation, the model from the pupil center to the gaze was compensated by the inner corner points.

3.1.2. Comparison of 2D gaze estimation algorithms

Table 1 summarizes the eye features, head motion requirements, personal calibration parameters and main error factors in common 2D gaze estimation methods. In a single-camera system, when no or only one light source is used, the 2D mapping model cannot guarantee the invariance of head motion, so the user's head is required to keep fixed. The PCT/ICT-based method and PCRT/ICRT-based method do not need to calibrate the camera, but they require multiple calibration points to determine the user-dependent mapping function coefficients, and the mapping model shows poor performance in the state of head motion. By adding additional cameras or light sources, not only can the mapping model be simplified to reduce the number of calibration points, but also the eye features that change with head motion can be added to the model to improve the performance. However, the complexity of system calibration increases accordingly. The system configurations required for the CR-based method and the HN-based method are similar. The research object of these methods is the glints formed by the corneal reflection of light sources, so they have higher requirements on glint detection. The CR-based method needs to calibrate the scale factor of virtual projection and compensate the kappa angle through multi-point calibration. The HN-based method needs to preset a normalized plane instead of the unknown corneal reflection plane and calibrate the homography matrix from the normalized plane to the screen. It has certain advantages in accuracy and robustness owing to introducing the normalized space.

3.1.3. Key issues of 2D gaze estimation

2D mapping-based methods are to first establish a mapping model between the variation features of eye movement and the POR, and then substitute the new variation features into the mapping model to estimate the corresponding POR. In this process, the key issues mainly include the following two points:

(1) **Head motion**: The mapping model is usually established while keeping the head at a certain calibration position. When the head deviates from the calibration position, the established mapping model cannot accurately reflect the mapping between

Table 1
Characteristics of common 2D gaze estimation methods.

Methods	Cameras	Lights	Eye features	Head motion	Calibration	Main error factors	References
PCT/ICT-based	1	0	Pupil/iris center,corner point	Fixed	a_i, b_i	Corner detection, head motion	[13],[14],[15],[16],[17]
	1	1	Pupil/iris center,glint	Fixed	a_i, b_i	Head motion	[18],[19],[20],[22],[23],[24]
PCRT/ICRT-based	1	≥2	Pupil/iris center,glints	Certain range	a_i, b_i	Head motion	[21],[25],[26],[27],[28],[29]
	≥2	≥1	Pupil/iris center,glints	Free	a_i, b_i	Camera calibration	[30]
CR-based	≥1	≥5	Pupil center,glints	Certain range	α	Coplanar approximation	[31],[32],[33],[34],[35],[36],[40],[54]
HN-based	≥1	≥4	Pupil center,glints	Certain range	$\mathbf{H}_N^S$	Glint detection	[37],[38],[39],[41],[42],[43],[55]

the variation features of eye movement and the POR. To solve this problem, there are two main methods: one is to consider the head motion in the mapping model, such as feature normalization [20,27,28], using eye features associated with head motion [29]. The second is to compensate the eye features after head motion to the calibrated position [17,23,24,30]. However, the allowable range of head motion for these two methods is still limited, especially the latter, which often requires additional system configuration or some known eye features.

(2) **Personal calibration**: The construction of a 2D mapping model requires personal calibration to determine some user-dependent parameters. However, most calibration methods require users to stare at multiple calibration points to obtain sufficient calibration data. The process is relatively complicated. Although Choi et al. [39] proposed to build a calibration database to estimate the POR based on one-point calibration, this method also requires a cumbersome process when constructing the database.

3.1.4. Research trends of 2D gaze estimation

Based on the research status of 2D mapping-based methods, the further research on 2D gaze estimation can be carried out from the following aspects:

(1) **2D gaze estimation based on simple mapping model**: Cerrolaza et al. [20] demonstrated that higher order polynomials do not noticeably improve system behavior. Some researchers proved that a simple mapping model can still obtain the ideal gaze accuracy [29,44,56]. Therefore, studying a simple mapping model while ensuring the accuracy can not only reduce the system requirements such as suitable for low-resolution cameras or reducing the number of light sources, but also simplify personal calibration such as reducing calibration points or calibration time.

(2) **2D gaze estimation based on implicit calibration**: The current method needs to obtain some user-dependent parameters through an explicit calibration, but performing complicated calibration before use is not an ideal way of HCI. Therefore, studying the 2D gaze estimation method based on implicit calibration is not only a theoretical breakthrough, but also has important practical value. For example, the calibration data is obtained by watching a video, or clicking the mouse while using the computer. It can effectively enhance the user's experience.

(3) **Accurate gaze estimation under natural head motion**: When the 2D mapping-based method is applied to a head-mounted system, it is less disturbed by the head motion, and higher accuracy can be achieved using the multi-camera-multi-light-source (MCMLS) system. However, even if multiple cameras and multiple light sources are used to compensate for the effect of head motion in a remote system, the performance is difficult to reach the level of a head-mounted system. To deal with this issue, combining 2D mapping and 3D model is an effective way to achieve accurate gaze estimation under natural head motion.

3.2. 3D model-based gaze estimation methods

3.2.1. 3D gaze estimation algorithms

3D model-based methods calibrate some eye invariant features and estimate some spatial eye parameters according to the eyeball structure and the spatial geometric imaging model, so as to reconstruct the OA of the eye and estimate the 3D LOS. The most common method is based on corneal reflection and pupil refraction (CRPR-based method), which can achieve high accuracy with the support of high-resolution cameras, infrared (IR) illumination, and high-precision feature extraction. To achieve 3D gaze estimation under natural light conditions, some researchers have also proposed passive methods based on the extraction of facial features (FF-based methods), which use a single camera without IR illumination. In recent years, the 3D gaze estimation method based on depth sensor (DS-based method) is rapidly developed since the 3D gaze estimation model can be simplified with the depth information.

(1) CRPR-based method

The CRPR-based method estimates the 3D cornea center and the 3D pupil center using the geometric relationship in corneal reflection and pupil refraction, and then construct the OA to estimate the 3D LOS. The POR can be obtained by intersecting the 3D LOS with the object of gaze. This method requires at least a single camera and a single light source. Table 2 lists the specific calculation methods of each procedure when different systems are used.

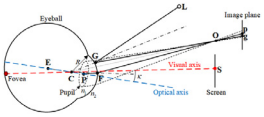
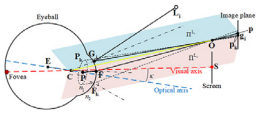
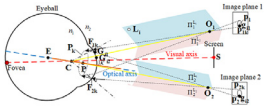
1) *3D cornea center estimation*: 3D cornea center is estimated by the corneal reflection of light sources. According to the law of reflection, both incident and reflected lights are contained in a plane together with the light source $\mathbf{L_i}(i = 1, 2, ..., N)$, 3D cornea center $\mathbf{C}$, camera optical center $\mathbf{O_j}(j = 1, 2, ..., M)$, and glint $\mathbf{g_{ij}}$. Moreover, the angles of the incident and reflected light relative to the normal line are equal. Since the reflection point $\mathbf{G_{ij}}$ is located on the corneal surface, the distance between $\mathbf{G_{ij}}$ and $\mathbf{C}$ is equal to the cornea radius. Therefore, the equations about 3D cornea center can be obtained. However, in a SCSLS system, if the cornea radius are unknown, 3D cornea center cannot be solved [57]. This method under a SCSLS system usually relies on setting some eye invariant parameters according to population averages or the classic eyeball model [57,58]. In a SCMLS system, as long as there are two non-collinear light sources, the line $\mathbf{OC}$ can be obtained by intersecting the reflection planes of two light sources, which can simplify solving three coordinates of $\mathbf{C}$ into solving a scale factor t. If multiple non-coplanar cameras are used on this basis, the solution of $\mathbf{C}$ can be further simplified to calculate the intersection of the line $\mathbf{O_jC}$ [59].

2) *3D pupil center estimation*: 3D pupil center is estimated by the pupil refraction when passing through the cornea. When considering the refraction of pupil center, according to the law of refraction, both incident and refracted lights are contained in a plane together with 3D pupil center $\mathbf{P}$, refraction point $\mathbf{F}$, 3D cornea center $\mathbf{C}$, and camera optical center. It also satisfies Snell's law. In addition, the distance between $\mathbf{F}$ and $\mathbf{C}$ is equal to the cornea radius. Assuming that the distance of pupil center to cornea center (DPC) is d, an equation can also be established. Then the equations about 3D pupil center can be obtained, where n_1 is the effective refraction index of the aqueous humor and cornea combined, and n_2 is the index of refraction of air. When the DPC is unknown, the condition for solving $\mathbf{P}$ is insufficient. Moreover, due to image distortion, the image point of 3D pupil center is not the center of pupil imaging ellipse [60], so 3D pupil center estimated from the center of pupil imaging ellipse is not accurate. From the analysis of the refraction of pupil edge points $\mathbf{P_k}(k = 1, 2, ..., W)$, the unit direction of refracted light $\mathbf{f_k}$ can be expressed by the unit direction of incident light $\mathbf{i_k}$ and the normal unit direction vector $\mathbf{n_k}$, and $\mathbf{P_k}$ can also be expressed. According to the equal distance from $\mathbf{P_k}$ to $\mathbf{P}$, the distance from $\mathbf{F_k}$ to $\mathbf{C}$ is equal to the cornea radius, and the line $\mathbf{CP}$ is perpendicular to the pupil plane, a set of equations can be established. When W is not less than 5, $\mathbf{P}$ can be estimated [57]. It is worth noting that when a MCMLS system is adopted, the 3D pupil center estimation is not necessary. The reason will be discussed in OA reconstruction.

3) *OA reconstruction*: After estimating $\mathbf{C}$ and $\mathbf{P}$, the OA of the eye is reconstructed by these two points [61]. Particularly, since $\mathbf{C}$ and $\mathbf{P}$ are both located in the refraction plane of pupil center, the OA can be directly obtained by intersecting the refraction planes when a MCMLS system is used, in which the refraction planes are determined by the 3D cornea center $\mathbf{C}$, the center of the pupil imaging ellipse $\mathbf{p_j}$, and the camera optical center $\mathbf{O_j}$ [57,62,63].

4) *Kappa angle calibration*: When the subject looks at some known on-screen calibration points in turn, the corresponding eye

Table 2
CRPR-based gaze estimation model.

Configuration	SCSLS system	SCMLS system	MCMLS system
Figure			
3D cornea center estimation	$(\mathbf{L}-\mathbf{O})\times(\mathbf{G}-\mathbf{O})\bullet(\mathbf{C}-\mathbf{O})=0$ $\frac{(\mathbf{L}-\mathbf{G})\bullet(\mathbf{G}-\mathbf{C})}{\|\mathbf{L}-\mathbf{G}\|}=\frac{(\mathbf{O}-\mathbf{G})\bullet(\mathbf{G}-\mathbf{C})}{\|\mathbf{O}-\mathbf{G}\|}$ $\|\mathbf{G}-\mathbf{C}\|=R$ where $\mathbf{G}=\mathbf{O}+u(\mathbf{O}-\mathbf{g})$**3 equations and 5 unknowns:** $\mathbf{C}, R, u$(Insufficient conditions)	$\frac{(\mathbf{L_i}-\mathbf{G_i})\bullet(\mathbf{G_i}-\mathbf{C})}{\|\mathbf{L_i}-\mathbf{G_i}\|}=\frac{(\mathbf{O}-\mathbf{G_i})\bullet(\mathbf{G_i}-\mathbf{C})}{\|\mathbf{O}-\mathbf{G_i}\|}$ $\|\mathbf{G_i}-\mathbf{C}\|=R$ where $\mathbf{G_i}=\mathbf{O}+u_i(\mathbf{O}-\mathbf{g_i})$, $\mathbf{C}=\mathbf{O}+t\frac{[(\mathbf{L_1}-\mathbf{O})\times(\mathbf{g_1}-\mathbf{O})]\times[(\mathbf{L_2}-\mathbf{O})\times(\mathbf{g_2}-\mathbf{O})]}{\|[(\mathbf{L_1}-\mathbf{O})\times(\mathbf{g_1}-\mathbf{O})]\times[(\mathbf{L_2}-\mathbf{O})\times(\mathbf{g_2}-\mathbf{O})]\|}$ $2N$ **equations and** $(N+2)$ **unknowns:** u_i, R, t	$(\mathbf{L_i}-\mathbf{O_j})\times(\mathbf{g_{ij}}-\mathbf{O_j})\bullet(\mathbf{C}-\mathbf{O_j})=0$ NM **equations and 3 unknowns: C**
3D pupil center estimation	**Using pupil center:** $(\mathbf{F}-\mathbf{O})\times(\mathbf{C}-\mathbf{O})\bullet(\mathbf{P}-\mathbf{O})=0$ $n_1\frac{\|(\mathbf{F}-\mathbf{C})\times(\mathbf{P}-\mathbf{F})\|}{\|\mathbf{P}-\mathbf{F}\|}=$ $n_2\frac{\|(\mathbf{F}-\mathbf{C})\times(\mathbf{O}-\mathbf{F})\|}{\|\mathbf{O}-\mathbf{F}\|}\|\mathbf{F}-\mathbf{C}\|=R\ \|\mathbf{P}-\mathbf{C}\|=d$ where $\mathbf{F}=\mathbf{O}+m(\mathbf{O}-\mathbf{p})$ **4 equations and 5 unknowns:** $\mathbf{P}, d, m$ (Insufficient conditions)	**Using pupil edge points:** $\|\mathbf{P_1}-\mathbf{P}\|=\|\mathbf{P_2}-\mathbf{P}\|=\ldots=\|\mathbf{P_W}-\mathbf{P}\|$ $\|\mathbf{F_k}-\mathbf{C}\|=R\ \frac{(\mathbf{P}-\mathbf{C})\bullet(\mathbf{P_j}-\mathbf{C})}{d}-d=0$ where $\mathbf{i_k}=\frac{\mathbf{p_k}-\mathbf{O}}{\|\mathbf{p_k}-\mathbf{O}\|}$, $\mathbf{n_k}=\frac{\mathbf{F_k}-\mathbf{C}}{\|\mathbf{F_k}-\mathbf{C}\|}$, $\mathbf{f_k}=\frac{n_2}{n_1}(\mathbf{i_k}-((\mathbf{i_k}\bullet\mathbf{n_k})+$ $\sqrt{(\frac{n_2}{n_1})^2-1+(\mathbf{i_k}\bullet\mathbf{n_k})^2})\mathbf{n_k})$, $\mathbf{F_k}=\mathbf{O}+m_k(\mathbf{O}-\mathbf{p_k})$, $\mathbf{P_k}=\mathbf{F_k}+\delta_k\mathbf{f_k}$ $(3W-1)$ **equations and** $(2W+4)$ **unknowns:** $\mathbf{P}, d, m_k, \delta_k$	-
OA reconstruction	$\mathbf{V^o}=\frac{\mathbf{P}-\mathbf{C}}{\|\mathbf{P}-\mathbf{C}\|}$	$\mathbf{V^o}=\frac{\mathbf{P}-\mathbf{C}}{\|\mathbf{P}-\mathbf{C}\|}$	$\mathbf{V^o}=\frac{[(\mathbf{O_1}-\mathbf{p_1})\times(\mathbf{C}-\mathbf{O_1})]\times[(\mathbf{O_2}-\mathbf{p_2})\times(\mathbf{C}-\mathbf{O_2})]}{\|[(\mathbf{O_1}-\mathbf{p_1})\times(\mathbf{C}-\mathbf{O_1})]\times[(\mathbf{O_2}-\mathbf{p_2})\times(\mathbf{C}-\mathbf{O_2})]\|}$
Kappa angle calibration	$\mathbf{V^p}=\mathbf{MV^o}$	$\mathbf{V^p}=\mathbf{MV^o}$	$\mathbf{V^p}=\mathbf{MV^o}$
3D gaze estimation	$\mathbf{S}=\mathbf{C}+\mu\mathbf{V^p}$ where $\mu=\frac{\mathbf{n_s}\bullet\mathbf{C}}{\mathbf{n_s}\bullet\mathbf{V^p}}$	$\mathbf{S}=\mathbf{C}+\mu\mathbf{V^p}$ where $\mu=\frac{\mathbf{n_s}\bullet\mathbf{C}}{\mathbf{n_s}\bullet\mathbf{V^p}}$	$\mathbf{S}=\mathbf{C}+\mu\mathbf{V^p}$ where $\mu=\frac{\mathbf{n_s}\bullet\mathbf{C}}{\mathbf{n_s}\bullet\mathbf{V^p}}$

features are extracted to estimate the 3D cornea center and the 3D pupil center. The OA direction vector $\mathbf{V^o}$ is represented by the 3D cornea center and the 3D pupil center. The VA direction vector $\mathbf{V^p}$ is represented by the 3D cornea center and the known calibration point. Therefore, multiple sets of OA and VA direction vectors during calibration can be obtained. The transformation relationship between the OA and the VA, represented by the transformation matrix $\mathbf{M}$, can be solved by the least square method [30].

5) *3D gaze estimation*: After constructing the OA direction vector using the real-time estimated 3D cornea center $\mathbf{C}$ and 3D pupil center $\mathbf{P}$, the transformation matrix $\mathbf{M}$ is used to convert from the OA direction vector to the VA direction vector. Combined with $\mathbf{C}$, the 3D LOS can be determined. If the screen has been calibrated and the normal vector of the screen plane is known to be $\mathbf{n_s}$, then the POR can be obtained.

In a SCSLS system, the conditions of the CRPR-based method are insufficient, and some eye invariant parameters need to be set according to population averages or the classic eyeball model. To solve this problem, Liu et al. [8] used the 3D reconstruction of a spatial circular target to represent the iris center and its normal vector, thereby the iris radius was calibrated with the constraint that the kappa angles of the left and right eyes are approximately equal. They also calibrated the cornea radius using the calibrated iris radius [64], but this method relies on accurate iris detection. In a SCMLS system, it is necessary to obtain the cornea radius and kappa angle through multi-point calibration. O'Reilly et al. [65] proposed a gaze tracking system that includes a single camera and 9 IR LEDs. A mathematical model with two solutions for estimating the cornea center were developed using the lines of LED lights to handle missing glints and allows head motion. However, multi-point calibration is still required. Compared with a single-camera system, a MCMLS system can simplify the estimation of 3D cornea center and OA. It can also simplify personal calibration to one-point calibration, but it would be more robust using more calibration points. The MCMLS system can allow natural head motion by utilizing multiple cameras to track the eye and head pose [66]. However, the use of a MCMLS system is costly and relatively complicated in system calibration.

(2) FF-based method

The FF-based method generally exploits the 2D image information to obtain the 3D face features, from which the head pose can be computed, and estimates the 3D LOS using an eyeball imaging model [67–70], as shown in Fig. 7. By tracking the facial feature points such as eye contour, nose tip, and mouth corners, the head pose and the midpoint of 3D eye corners can be estimated using a 3D universal face model. According to the eyeball structure and pinhole imaging model, 3D eyeball center, 3D pupil center and 3D cornea center can then be represented. Through the process of staring at N calibration points, $7N$ equations can be established using the distance relationship among 3D eye features and the consistency of estimated PORs and calibration points. In the case where K and K_0 are set by population averages, the number of unknowns is $6N+5$, including: kappa angle (α, β), the vector pointing from $\mathbf{E}$ to $\mathbf{M}$, etc. [71]. These parameters can be used for subsequent 3D gaze estimation.

The system required by the FF-based method is simple and does not require additional illumination, but the required user calibration is complicated. To alleviate the calibration, Jeni and Cohn [72] reconstructed face shape by fitting a part-based 3D model to estimate a canonical view of eyes, and trained a linear SVR from the binary features involving six eye contour points and pupil cen-

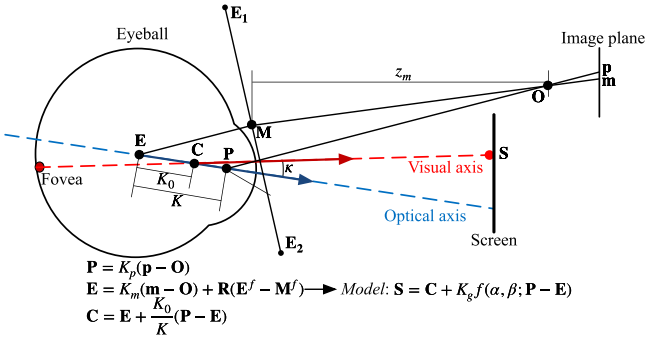


Fig. 7. FF-based gaze estimation model.

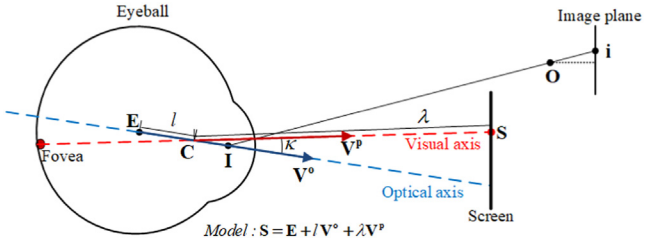


Fig. 8. DS-based gaze estimation model.

ter to the 3D gaze direction. Cristina and Camilleri [73] estimated the 3D gaze from a single camera based on a cylindrical head and spherical eyeball model. They eliminated prior camera calibration and simplified the user calibration to frontal eye and head pose detection, which is suitable for disabled people interacting with the computer. The FF-based method considers the head pose and is robust to head motion, but it uses population averages to set some eyeball parameters, and the facial feature detection in low-quality images is also challenging. Its accuracy is often lower than active methods with IR illumination.

(3) DS-based method

The DS-based method usually uses Kinect to obtain depth information and modeling the eye by combining image and depth information to estimate the 3D POR [74–78], as shown in Fig. 8. The origin of head coordinate system in the camera coordinate system and the transformation matrix between head and camera coordinate systems are obtained first through head pose tracking. According to the property that the eyeball center remains fixed relative to the head position, the eyeball center in the head coordinate system can be calibrated, so the 3D eyeball center $\mathbf{E}$ in the camera coordinate system can be transformed. The 3D iris center $\mathbf{I}$ is estimated by pin-hole imaging, thereby the OA direction vector $\mathbf{V^o}$ is constructed by $\mathbf{E}$ and $\mathbf{I}$. After calibrating the kappa angle (transformation matrix $\mathbf{M}$ or horizontal and vertical rotation angles α, β), the VA direction vector can be obtained by $\mathbf{V^p} = \mathbf{MV^o}$(or $f(\mathbf{V^o}, \alpha, \beta)$). Then the POR can be estimated, where l is generally 5.3mm [57] or ignored; λ is expressed as $\lambda = -\frac{(\mathbf{E}+l\mathbf{V^o})^T \bullet \mathbf{V^s} + n}{(\mathbf{V^p})^T \mathbf{V^s}}$. For any point $\mathbf{S}$, $\mathbf{V^s}$ and n can be obtained by system calibration, satisfying $\mathbf{S} \bullet \mathbf{V^s} = -n$ [74].

The DS-based method directly converts the 3D POR relying on the eyeball center and the iris center by vector superposition. Due to its simple system configuration and allowing head motion, the use of depth sensors is consistent with the wider application requirements of gaze tracking system. But this method is highly dependent on accurate feature detection and estimation of intermediate quantities. In addition, many parameters need to be calibrated. Sun et al. [74] calibrated the eyeball radius and the vector between eyeball center and eye inner corner in the head coordinate system to estimate the eyeball center and the iris center in the camera coordinate system. Finally, the line connecting the eyeball center and the iris center was used to indicate the LOS. Later, they calibrated the kappa angle to consider the individual differences [75]. Although Wang and Ji [76] used the pupil features instead of iris features, and achieved comparable accuracy with higher speeds using low-resolution images, the calibrated parameters including the eyeball radius, the distance between the eyeball center and the cornea center, and the kappa angle are needed. Zhou et al. [78] presented a one-point calibration method, but its gaze point accuracy is relatively low as the LOS is approximately presented by the line connecting the iris center and the POR.

(4) Other 3D gaze estimation methods

3D gaze estimation methods are often used in remote gaze tracking systems. A few researchers have proposed the methods based on a head-mounted system [79,80]. There are also some methods for specific issues, such as to *improve the impact of head motion* [81,82], *simplify personal calibration* [83,84], and *simplify system configuration* [85,86]. For example, Wang and Ji [83] studied 3D gaze estimation without explicit personal calibration, using the following constraints: 1) 2D and 3D gazes are complementary to each other, 2) most of the fixations are concentrated on the center region of the screen, 3) the POR is within the screen region, 4) kappa angle theoretical range constraints. Wen et al. [86] proposed to calibrate the eyeball radius, eyeball center and kappa angle without knowing the specific calibration point using a single RGB camera. They used the vergence constraint and photometric consistency constraint to estimate the 3D gaze online, and used the linear regression to reduce the systematic error.

3.2.2. Comparison of 3D gaze estimation algorithms

Table 3 summarizes the features, head motion requirements, personal calibration parameters and main error factors in common 3D gaze estimation methods. When the CRPR-based method is applied to a SCSLC system, the estimates of 3D cornea center and 3D pupil center are lack of conditions unless the eye invariant parameters such as cornea radius are unknown as a prior. These parameters have individual differences, so the CRPR-based method has limited applicability when using a SCSLS system. In a SCMLS system, the eye invariant parameters can be calibrated according to the spatial geometric model in addition to being set to fixed values. But they are optimally solved in the nonlinear equations, and the solutions are numerical. In a MCMLS system, 3D cornea center and the OA can be directly estimated to obtain the LOS. The wide-view camera tracks the head pose to allow natural head motion, but the system calibration will be relatively complicated. The FF-based method combines the facial and eye features to estimate the 3D LOS with a single camera and ambient illumination. It is difficult to accurately detect facial features and estimate head pose in a simple system, which determines that the accuracy is lower than the CRPR-based method. The DS-based method uses a depth sensor to estimate the spatial points such as eyeball center and iris center. The gaze estimation model is a superposition of the spatial points and vectors, which requires high accuracy for these spatial points, and is complicated to calibrate the eyeball center and the kappa angle.

3.2.3. Key issues of 3D gaze estimation

3D model-based gaze estimation methods are to first estimate the spatial eyeball parameters, and then reconstruct the OA of the eye, so as to estimate the 3D LOS. The key issues in 3D gaze estimation are as follows:

(1) **Head motion**: When the head moves in translation, pitch and yaw, the pupil/iris will change accordingly and can be used to estimate the 3D LOS. However, when the head rolls, the rotation of the OA around itself cannot be characterized by the visual features or the estimated cornea center or pupil center as they are all located on the OA. Even with a light source, the glint still cannot characterize this rotation component since the cornea is approximately spherical. The VA cannot be accurately estimated from the OA if the spatial eye state cannot be characterized. Therefore, 3D gaze estimation allows the translation, pitch and yaw of the head, but head rolling is often restricted.

(2) **Personal calibration**: The required eye invariant parameters can be preset according to population averages or the classic eyeball model, but it does not consider the individual differences. In most cases, personal calibration is an important part of 3D gaze estimation. The CRPR-based method needs to calibrate the cornea radius, kappa angle, etc.; even if a MCMLS system is used, kappa angle calibration is unavoidable. After setting the distance between 3D eyeball center and 3D cornea center, and the distance between 3D eyeball center and 3D pupil center, the FF-based method still

Table 3
Characteristics of common 3D gaze estimation methods.

Methods	Cameras	Lights	Features	Head motion	Calibration	Main error factors	References
CRPR-based	1	1	Pupil, glint	Certain range	R,kappa	Presetting eye invariant parameters	[57], [58]
	1	≥2	Pupil, glints	Certain range	R,kappa	Solving or presetting eye invariant parameters	[57], [59], [65], [87]
	≥2	≥1	Pupil, glints	Free(Without rolling)	kappa	System calibration	[57], [59], [61], [62], [63], [66]
FF-based	1	0	Eye contour, nose, mouth, Pupil/iris	Certain range	K_p, K_m, K_g, $\overrightarrow{\mathbf{EM}}$,kappa	Facial feature detection, head pose estimation	[67], [68], [69], [70], [71]
DS-based	1	0	Pupil/iris, corner point	Certain range	**E**,kappa	Spatial point estimation	[74], [75], [76], [77], [78]

needs to calibrate some scale factors between eyeball, camera and image plane, and the kappa angle. The DS-based method needs to calibrate the eyeball center, kappa angle, etc. 3D gaze estimation can simplify personal calibration to one-point calibration, but the results are often less robust than using multiple calibration points.

(3) **System configuration**: Most 3D gaze estimation methods use a SCMLS system or a MCMLS system. In addition, to improve the pupil detection, the light source usually adopts the IR illumination. The 3D gaze estimation method has better performance under a higher system configuration, but it does not meet the requirements of widespread application. Although the system configuration required by the FF-based method and the DS-based method is relatively simple, the accuracy is very dependent on the feature detection and parameter estimation.

3.2.4. Research trends of 3D gaze estimation

Based on the research status of 3D model-based method, the further research on 3D gaze estimation can be carried out from the following aspects:

(1) **3D gaze estimation based on simple system**: At present, allowing head motion and simplifying personal calibration are mainly achieved by using a SCMLS or MCMLS system, but high system configuration will hinder its implementation and application. With the development of machine learning and pattern recognition, the ideal results can be obtained based on a simple system by combining the advantages of neural networks in image processing in the solution of eyeball parameters.

(2) **3D gaze estimation based on implicit calibration or calibration-free**: Although the existing methods can simplify personal calibration to one-point calibration, its robustness is not as strong as multi-point calibration. Explicit calibration is also not an ideal way of HCI. Therefore, studying implicit calibration or calibration-free method is the trend of 3D gaze estimation. For example, the eye invariant parameters are implicitly calibrated while using the computer; or the 3D LOS is adaptively iterated by taking some habits as constraints.

(3) **High-precision gaze estimation under natural head motion**: High-precision gaze estimation relies in part on restricting the head motion within a certain range, so as to collect higher-quality images. However, it is necessary to meet that the 3D LOS can still be estimated accurately under natural head motion in practical applications. Brousseau et al. [87] pointed out that the gaze estimation accuracy in handheld mobile devices depends on the eye-system relative rolling angle and the kappa angle. This is an urgent and challenging bottleneck of gaze tracking. Track the head pose or combining left and right eyes can predict the degree of head rolling.

3.3. Appearance-based gaze estimation methods

3.3.1. Gaze tracking datasets

With the rapid development of machine learning, the appearance-based method has been paid increasing attention and many datasets have been proposed for gaze tracking. Table 4 lists some datasets available online in recent years in terms of acquisition environment, number of subjects, data types, head pose requirements, gaze direction range and gaze annotations. The download links are also provided. Some typical datasets are respectively discussed below.

(1) *MPIIGaze*: It is the first in-the-wild dataset, and is collected during natural everyday laptop use. It not only provides binocular images, but also annotates the eye landmarks, 2D and 3D gazes, 3D head pose, and 3D eyeball centers. A standard evaluation set including 1,500 images for each eye taken from each subject randomly is also provided. In the annotation subset, six facial landmarks (four eye corners, two mouth corners) and pupil centers in 10,848 samples are manually annotated. Later, Zhang et al. [10] provided the additional facial landmark annotation and face regions in MPIIFaceGaze. Compared with the laboratory environment, it presents a more significant variability in eye appearance and illumination. It is currently the most widely used dataset in the study of appearance-based methods. However, the head pose covered by MPIIGaze is mainly frontal, and the camera-subject distances are relatively small, which is not suitable for remote gaze tracking.

(2) *GazeCapture*: It is collected by crowdsourcing on mobile devices. Krafka et al. [88] built an iOS application to collect face images from 1,474 subjects, in which 1,249 subjects used iPhones while 225 used iPads. It provides face images under various lighting conditions and head poses and their 2D gaze annotations, which is convenient for learning an unconstrained 2D gaze estimation model. But the head pose does not include some extreme head poses, while the camera-subject distances are also small.

(3) *RT-GENE*: It provides accurate ground truth gaze annotations and head pose labels without specifying an explicit gaze target. Fischer et al. [89] used a commercial motion capture system for head pose detection, and the mobile eye tracking glasses to automatically annotate the subject's ground truth. A total of 122,531 labeled training images and 154,755 unlabeled images of 15 subjects where the mobile eye tracking glasses are not worn are provided. It covers large camera-subject distances, and wide ranges of head poses and gaze directions. However, the face appearance is altered due to the use of a head-mounted eye tracker to obtain the ground truth gaze annotations. Although semantic inpainting is used to reduce the difference with real appearance, the error cannot be eliminated.

Table 4
Comparison of gaze tracking datasets.

Datasets	Environments	Subjects	Data types	Head poses(H;V)	Gaze directions(H;V)	Annotations	Year
Columbia [91][a]	Laboratory	56	5880 images	5(0°;±30°)	21(±15°;±10°)	3D gaze	2013
EyeDiap [92][b]	Laboratory	16	94 videos	CONT(±15°;30°)	CONT(±25°;20°)	2D/3D gaze	2014
UT Multiview [93][c]	Laboratory	50	64000 images	8+synthesised(±36°)	160(±50°;±36°)	3D gaze	2014
MPIIGaze [94][d]	Daily life	15	213659 images	CONT(±15°;30°)	CONT(±20°;−30°)	2D/3D gaze	2015
OMEG [95][e]	Laboratory	50	44827 images	3(±30°;0°)+CONT	10(−38° − 36°;−10° − 29°)	2D/3D gaze	2015
GazeCapture [88][f]	Daily life	1474	2445504 images	CONT(±30°;40°)	CONT+13(±20°)	2D gaze	2016
MPIIFaceGaze [10][g]	Daily life	15	37667 images	CONT(±15°;30°)	20(±20°;−30°)	2D/3D gaze	2017
InvisibleEye [96][h]	Laboratory	17	280000 images	4	25(±35°;±22°)	2D gaze	2017
TabletGaze [97][i]	Laboratory	51	816 videos	CONT	35(3.42*cm*;3.41*cm*)	2D gaze	2017
RT-GENE [89][j]	Laboratory	15	122531 images	CONT(±40°)	CONT(±40°;−40°)	3D gaze	2018
NVGaze [98][k]	Laboratory	30	4.5M images	CONT	CONT	2D gaze	2019
Gaze360 [90][l]	Indoor,outdoor	238	169748 images	CONT(±90°;/)	CONT(±140°;−50°)	3D gaze	2019
ShanghaiTechGaze [99][m]	Laboratory	137	233796 images	CONT	CONT	2D gaze	2019
ETH-XGaze [100][n]	Laboratory	110	1083492 images	CONT(±80°)	CONT(±120°;±70°)	2D/3D gaze	2020
EVE [101][o]	Laboratory	54	4.2K videos	CONT(±40°)	CONT(±30°)	2D/3D gaze	2020
OpenEDS2020 [102][p]	Laboratory	80	550400 images	CONT	CONT+17(±20°)	3D gaze	2020

[a] https://cs.columbia.edu/CAVE/databases/columbia_gaze
[b] https://idiap.ch/dataset/eyediap
[c] https://ut-vision.org/datasets
[d] https://www.mpi-inf.mpg.de/mpiigaze
[e] http://www.cse.oulu.fi/CMV/Downloads
[f] https://gazecapture.csail.mit.edu
[g] https://mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/gaze-based-human-computer-interaction/its-written-all-over-your-face-full-face-appearance-based-gaze-estimation
[h] https://mpi-inf.mpg.de/invisibleeye
[i] https://sh.rice.edu/cognitive-engagement/tabletgaze
[j] https://github.com/Tobias-Fischer/rt_gene
[k] https://sites.google.com/nvidia.com/nvgaze
[l] https://gaze360.csail.mit.edu
[m] https://github.com/dongzelian/multi-view-gaze
[n] https://ait.ethz.ch/projects/2020/ETH-XGaze
[o] https://ait.ethz.ch/projects/2020/EVE
[p] http://research.fb.com/programs/openeds-2020-challenge

(4) *Gaze360*: It covers a wide range of head poses and gaze directions, variety of indoor and outdoor capture environments and diversity of subjects, especially the entire horizontal range of 360°. Kellnhofer et al. [90] collected the images in five indoor and two outdoor locations with a Ladybug5 360° panoramic camera and a large moving rigid target board marked with an AprilTag and a cross. A total of 129K training, 17K validation and 26K test images with 3D gaze annotation are efficiently collected, which is suitable for 3D gaze estimation based on deep learning.

The head poses, individual differences, gaze range, and environmental differences included in the datasets are all important factors that affect the gaze estimation accuracy. Although they are taken into consideration when collecting the data over time, it is time-consuming to collect samples and manually annotate them in actual scenes, and the data coverage is also limited. Therefore, some researchers have proposed to synthesize the images [93,103–106]. Wood et al. [104] proposed UnityEyes, which contained large amounts of variable eye region images that rapidly synthesized by combining a statistically-derived generative 3D model of the eye region with a real-time rendering framework. However, the distribution of synthetic images compared to real images is different. Although Shrivastava et al. [105] proposed to refine the gap between synthetic images and real images using generative adversarial networks, the appearance and gaze diversity of refined images are still limited to the variations found in the real images. This makes it difficult to apply networks trained with these synthetic datasets to real images.

3.3.2. Appearance-based gaze estimation algorithms

Appearance-based methods aim to learn a mapping model of the appearance in eye or face images and the gaze through a large number of training samples, so as to find the corresponding gaze of new appearance. Many conventional machine learning methods can be used to obtain the appearance-based mapping model, mainly: linear regression (LR), K-nearest neighbor (KNN), random forest (RF), gaussian process (GP), support vector machine (SVM), etc. The method based on convolutional neural network (CNN) also embodies greater advantages in recent years.

(1) **Gaze estimation method based on conventional machine learning**

1) *LR*: LR approximates the relationship between the appearance and the gaze with a linear mapping function, but the number of samples is usually more than the feature dimension. Therefore, adaptive linear regression(ALR) is to determine a linear model that satisfies the mapping relationship of a certain subset of all samples by reducing the sample dimension. Lu et al. [107] used an ALR method based on the l-optimization framework to predict the gaze position from sparsest eye images according to the local linearity in the eye appearance manifold, reducing the number of training samples while ensuring accuracy. The ALR algorithm is convenient for decision analysis, but it is not suitable for processing nonlinear data or data with complex feature correlation.

2) *KNN*: To estimate the gaze direction of an eye image, the k-nearest neighbors corresponding to it in the training set should be found, and their weights are determined by the distance between the image and each neighbor. Then the gaze direction is represented by calculating the weighted average of the gaze directions of k-nearest neighbors. Wang et al. [108] selected the closest samples with more relevant features using the information of head pose, pupil center and eye appearance, and then predicted the gaze direction by the neighbor regression between the appearance and the gaze angle. The KNN algorithm is effective and weighted average shows great advantages especially in small or unbalanced samples, but its computational complexity is large, and the results are sensitive to the selection of k.

3) *RF*: RF is to estimate the output of a new input by the average of the output of each tree that is generated for multiple sample datasets with random replacement from the original data [97].

Table 5
Conventional appearance-based gaze estimation methods in recent years.

Algorithms	Refs	Experimental setups/datasets	Calibration	Accuracy	Year
ALR	[107]	7 subjects;9/18/23/33 training samples and nearly 100 test samples	√	$2.37 \pm 1.42^{\circ}$(23 training samples)	2014
KNN	[104]	UnityEyes	×	9.95°(Cross-dataset)	2016
	[108]	UT Multiview	×	$3.26 \pm 1.12^{\circ}$(Cross-subject)	2018
RF	[93]	UT Multiview	×	$6.5 \pm 1.5^{\circ}$(Cross-subject)	2014
	[114]	6 subjects; 22 groups of videos; 107681 images	√	$4-8^{\circ}$(Cross-subject)	2016
	[109]	200k synthetic training samples and 17k test samples of 42 people	×	H:3.65°,V:3.88°(Cross-dataset)	2017
	[115]	20 subjects; 90791 photos	√	94.12%(Within dataset)	2020
GP	[110]	12 subjects; 4 different experiment cases; videos	√	H:1.42°,V:1.40°(Within dataset)	2014
	[116]	10 subjects; 250 training data and two parts of test phase	√	1.23°(Within dataset)	2019
	[111]	16 driving sequences	√	82.5%(95% confidence interval)	2020
	[117]	5 subjects; 82 calibration points	√	2.00°(Within dataset)	2020
SVR	[112]	15 subjects; 800 training images and 200 testing images	×	95.67%(Without SVM scale)	2014
	[118]	4 experimental scenarios; 8 gaze directions; videos	×	97.4%(Sport-utility-vehicle)	2014

Sugano et al. [93] learned a gaze angle estimator based on random regression forests with some redundancy of head poses using a large number of synthetic eye images. Kacete et al. [109] used RF regression to estimate the gaze vector, in which the input vector was built by integrating the depth information around the face. The training speed of RF algorithm is fast, which is suitable for processing high-dimensional data, but overfitting is prone to occur in high-noise regression problems.

4) *GP*: GP regression is based on the prior probability distribution of the predicted value of test set. It uses the joint probability distribution between the samples in the training set to calculate the posterior probability distribution of the predicted value of the test set through the Bayesian formula. Ferhat et al. [110] considered to use GP interpolator to map the average eye images to the screen coordinates. Shirpour et al. [111] applied a GP regression model from the head pose to estimate the gaze direction and describe the drivers visual attention. The GP regression model is a probabilistic model. It is not suitable for large datasets due to its high time and space complexity.

5) *SVM*: SVM utilizes the well-known "kernel trick" to map data into a higher-dimensional space, and transforms the original low-dimensional nonlinear regression problem into a high-dimensional linear regression problem. Wu et al. [112] located the eye region by modifying the characteristics of active appearance model, and used SVM to classify the five gaze directions. Lu et al. [113] fed the texture features from local pattern model and the spatial coordinates together into SVM to match a gaze mapping function, and subsequently tracked the gaze direction under allowable head motion. SVR has good robustness even in the presence of certain deviations, but it is not effective when dealing with large-scale training samples and multi-classification problems.

Table 5 lists some methods based on conventional machine learning in recent years. They are susceptible to individual differences and head motion. For this reason, some researchers have proposed person-specific models to consider the differences [119,120]. To improve the impact of head motion, there are mainly two common methods. One is to compensate the unknown head poses and correct the gaze [121,122]. Mora and Odobez [121] obtained the accurate head pose using a multimodal Kinect sensor and generated a person-specific 3D mesh model based on a 3D morphable model, so as to create a frontal pose face image and estimate the head pose. They corrected the final gaze vector by the estimated head pose. The second is to use head pose information in the gaze estimation model [123,124]. Yuan et al. [124] combined multi-scale eye image sparse features and head pose to predict the gaze region, which still has high accuracy even under large head motion. The gaze estimation method based on conventional machine learning requires a small number of samples and short training time, but its generalization ability is poor. When the eye or facial appearance is affected by individuals, head poses, lighting conditions, etc., the obtained accuracy is often low, which limits its development and application to a large extent.

(2) **Gaze estimation method based on CNN**

The CNN-based gaze estimation method uses a multi-layer neural network to learn the mapping model between appearance and gaze. It uses the gradient descent method to minimize the loss function to layer-by-layer reversely adjust the weight parameters in the network, and improves the accuracy of the network through frequent iterative training. According to different learning methods, the models of the CNN-based method are divided into: supervised learning model, weakly/semi/self-supervised learning model and unsupervised learning model.

1) *Supervised learning model*: Most appearance-based methods use supervised learning models. In addition to the above-mentioned conventional methods, supervised CNNs are the most widely used. They need to train a mapping model using a large number of samples with known gaze labels. The input of the model is eye images, face images, or eye images combined with face images, and the output is the corresponding gaze direction or gaze point. Zhang et al. [94] first proposed to use a multimodal CNN model to predict the gaze angle. They used the LeNet architecture, and added the head pose to the output of the fully connected layer for training a linear regression layer. Krafka et al. [88] first proposed using CNN to determine a POR estimation model. They used the face image, left and right eye images, and face grid information as inputs.

CNN models are widely used to deal with head motion, individual differences, gaze range differences, and environmental differences. To consider the head motion, the head pose is usually learned separately [125–127]. Wang et al. [127] learned the head pose from human face for gaze angle estimation when both face images and head pose labels are available, and learned the head pose from the eye deformations when they are not available. In case of the latter, the labeled data in UnitEyes was used to learn the mappings from landmark hidden features to gaze and head-pose through the movements of pupil and iris. Many researchers have studied person-specific gaze estimation models [4,94,128,129]. Wang et al. [128] leveraged on eye movement dynamics to improve generalization capabilities. It is also an effective way to calibrate the model using a few calibration samples to deal with individual differences [130–132]. Liu et al. [132] trained a differential CNN to predict the gaze difference between two eye images of the same subject. The gaze direction of a new eye image was predicted by inferring the gaze difference of a few subject-specific calibration samples. Limited gaze range will degrade the perfor-

mance. In addition to expanding the gaze range covered by the datasets [90,99,127], Palmero et al. [129] combined the face, eyes region, and face landmarks as individual streams in a CNN to estimate the gaze in still images, and then used the dynamic nature of gaze by feeding the learned features of all frames in a sequence to a many-to-one recurrent module, which predicts the 3D gaze vector of the last frame. Most datasets are collected in a controlled environment like a laboratory. In addition to some subject/scenario-specific methods, some methods are specific for different scenarios [127,133]. Salvalaio and Ramos [133] introduced an eye tracking algorithm based on online deep appearance, which continuously fine-tunes the eye tracking model by online transfer learning to make the model quickly adapt to different environments. In particular, Wu et al. [134] proposed an EyeNet architecture based on appearance and geometric tasks, which can robustly locate pupil and glint from off-axis eye images, and then estimate the VA based on the 3D model. There are various methods for locating eye features such as pupil based on CNN [135–137]. Accurate segmentation of these eye features can effectively improve the gaze estimation performance of appearance-based methods.

Supervised CNNs mainly rely on the design of network architecture and the adjustment of module parameters. With reasonable module and structure design, good performance can be achieved, but it requires a large number of samples and long training time. Although many methods of synthesizing images have been proposed to reduce the complexity of sample collection and labeling, there are essential differences between the synthetic and real images.

2)*Weakly/semi/self-supervised learning model*: The weakly/semi/self-supervised learning model mainly relies on analyzing some features related to labeled samples from a large number of unlabeled samples to improve gaze estimation performance [138–146]. Yu et al. [140] used a few annotated eye images from a user to generate some gaze-redirected images, and then used these images to fine-tune a generic gaze estimator. Lindén et al. [141] modeled personal variations as a low-dimensional latent parameter space, and captured the range of personal variations by calibrating a spatial adaptive gaze estimator for a new person. Wang et al. [143] proposed a Bayesian adversarial learning method to deal with the appearance and head pose changes and the overfitting issue of point estimation, so as to improve the generalization ability of gaze estimation. Kothari et al. [146] used the activity that “people looking at each other” in videos as a constraint for weakly-supervised gaze estimation.

Compared with the supervised learning model, the weakly/semi/self-supervised learning model can significantly reduce the demand for labeled samples and increase the learning rate. With the help of a large number of unlabeled samples, the gaze estimation performance can be effectively improved, and the generalization ability is strong. The weakly/semi/self-supervised learning model has a broad development space and application prospects.

3) *Unsupervised learning model*: The unsupervised learning model uses only unlabeled samples for training, and generally learns some hidden gaze representations from the image domains to estimate the gaze [142,147–149]. Guo et al. [147] designed domain adaptation gaze estimation network, which learns embedding representation with prediction consistency to ensure that linear relationships between gaze directions in different domains remain consistent on gaze space and embedding space. Yu and Odobez [148] synthesized gaze redirection images based on two unlabeled images with the same or at least close head poses, and learned low dimensional gaze representations without gaze annotations based on the redirection loss in image domain. He et al. [142] proposed an unsupervised few-shot personalization method based on heterogeneous teacher-student network. They used the user embedding and query features to predict the gaze location.

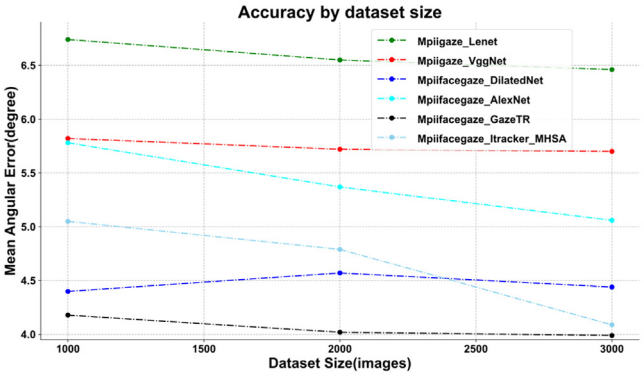


Fig. 9. Dependence of the training set size on the algorithm accuracy.

The unsupervised learning model estimates the gaze information based on the eye or face structural features without gaze annotation. However, it is difficult to improve the performance without using labeled samples. The model accuracy also cannot be directly evaluated.

Table 6 summarizes some CNN-based gaze estimation methods in recent years in terms of learning methods, input features and addressed issues. In general, supervised CNNs based on a large number of labeled samples are the most reliable and widely used, but the collection and annotation of images is cumbersome. The quality of labeled samples is also a key factor affecting the performance. Weakly/semi/self-supervised CNNs can estimate accurate gaze from a small number of labeled samples. They have advantages in performance and efficiency, and have gained more and more attention. Unsupervised CNNs take unlabeled samples as input to obtain gaze information from hidden gaze representations, but it is prone to deviate from the ground-truth due to the issues such as individual differences. The performance can be effectively improved using a small number of labeled samples to fine-tune the model.

Since the CNN-based gaze estimation method trains the model with a large number of samples, we evaluated the effect of the training set size on the algorithm accuracy. Both MPIIGaze and MPIIFaceGaze were selected in our evaluation. For MPIIGaze, LeNet [94] and VGGNet [4] were tested; for MPIIFaceGaze, four models including DilatedNet [161], AlexNet [10], GazeTR [164] and Itracker_MHSA [165] were tested. The results are shown in Fig. 9. The horizontal axis represents the number of images collected from each subject, and the vertical axis represents the mean angular error of the leave-one-out test for all subjects. The polylines represent the results when combining different datasets and models. It can be seen that when using LeNet and VGGNet on MPIIGaze, as the training images collected from each subject increase, the overall angle error shows a downward trend; when using different network architectures on MPIIFaceGaze, the algorithm accuracy generally improves with an increasing dataset size, but the dependence of different networks on the dataset size is different. The DilatedNet even shows the opposite result, which may be caused by the network overfitting in feature extraction. In summary, increasing the size of a small dataset is beneficial to improve the algorithm accuracy, but as the dataset gradually increases, the improvement rate will decrease, and there will even be unwanted results such as overfitting.

3.3.3. Comparison of appearance-based gaze estimation algorithms

Table 7 compares the required sample size, training time, cross-subject performance and main error factors of the appearance-

Table 6
Appearance-based gaze estimation methods based on CNN in recent years.

Methods / Addressed issues / Image categories		Head motion	Individual differences	Gaze range differences	Environmental differences	Calibration-free	Low-resolution images
Supervised CNNs	Eye	[94], [4], [150], [128], [151]	[94], [4], [128], [130], [131], [132], [152]	[99], [4]	[94], [4], [128], [152]	[94], [99], [4], [153], [150], [128], [154], [151]	[154]
	Face	[10], [90], [125], [129], [155]	[10], [129], [156]	[10], [90], [129]	[10], [90], [133], [155]	[10], [90], [157], [125], [129], [155], [158]	[158]
	Eye+Face	[88], [89], [126], [127], [159], [160], [161], [162]	[88], [89], [126], [127], [163], [160], [161]	[89], [127]	[88], [89], [126], [127], [160]	[88], [89], [126], [127], [159], [161], [162]	[89], [127], [161]
Weakly/semi/self-supervised CNNs	Eye	[139], [143], [145]	[140], [143], [145]	-	[139], [143], [145]	[139], [140], [143], [145]	-
	Face	[142,146]	[144], [146]	[146]	[146]	[146]	-
	Eye+Face	[138], [141]	[141]	-	[138]	[138]	-
Unsupervised CNNs	Eye	[148]	[147], [148]	-	[148]	-	-
	Face	[149]	[142], [149]	-	-	-	-

Table 7
Characteristics of common appearance-based gaze estimation methods.

Methods	Sample size	Training time	Calibration	Cross-subject performance	Main error factors
Conventional machine learning-based	Small	Short	×	Poor	Individual, head motion
			√	General	Head motion, environment
CNN-based	Large	Long	×	General	Individual, environment
			√	Good	Environment, gaze range

based method. The conventional machine learning-based method has the advantages of small sample size and short training time, but it cannot cope well with the impact of individual differences, head motion, environmental differences, etc. Even though the performance can be improved to a certain extent through personal calibration, high-precision and robustness cannot be guaranteed under natural conditions. The CNN-based method can effectively filter irrelevant information through continuous convolution and pooling operations, and improve the recognition efficiency without loss of effect. The performance is better than most standard machine learning algorithms. Most of CNN-based methods do not need calibration process, and the cross-subject performance is general. If a small number of calibration samples are used, the performance can be further improved. However, the CNN-based method tends to require more samples and longer training time than the conventional machine learning-based method. Although some corresponding methods to the issues such as head motion, individual differences, gaze range differences and environmental differences have been proposed, they are not completely compatible. With the increasing number of datasets available online, it is currently the mainstream of research on appearance-based gaze estimation methods.

3.3.4. Key issues of appearance-based gaze estimation

The appearance-based gaze estimation method learns a mapping model from the eye or facial appearance to the gaze to predict the gaze in a new image. The key issues involved are:

(1) **Model input selection**: Appearance-based methods mostly take eye or facial appearance as the model input, and a few methods take both eye and facial appearance as the input. As far as we know, there is no specific research comparing which input is more effective. This is the fundamental issue that should be paid attention to in the study of appearance-based methods. The eye appearance seems to lack facial landmarks and head pose when compared to the facial appearance, but it is not easy to accurately obtain the head pose from the facial appearance. If the quality of features in the facial appearance is not high, the performance may be worse than just inputting the eye appearance.

(2) **Personal calibration**: Although most appearance-based gaze estimation models are calibration-free, they are more susceptible to large individual differences, extreme head poses, etc. Some researchers have proposed to fine-tune the model with a few calibration samples [132,133,141,142,163], but it is time-consuming to collect labeled samples through explicit calibration in real scenes, which will affect the degree of automation and real-time performance of the system. More natural and convenient calibration methods should be studied.

(3) **Complex appearance**: Unconstrained gaze estimation is mentioned in many articles [90,97,109,127], but the performance is limited by the visual information covered in the training set. Zhang et al. [4] studied the key challenges of unconstrained gaze estimation including the gaze range, illumination conditions, and facial appearance variation. These differences complicate the eye appearance, which affects the gaze estimation performance. To meet the application requirements of natural HCI, the realization of accurate gaze estimation of complex appearance is still a key issue to be solved in appearance-based methods.

3.3.5. Research trends of appearance-based gaze estimation

Appearance-based gaze estimation has many challenges, and its research trends can be summarized as follows:

(1) **Gaze estimation based on small labeled datasets**: To improve the collection and labeling of samples, gaze estimation based on small labeled datasets is a key research direction. Head motion, individual differences, and gaze range are all important factors that hinder the development of appearance-based methods, which will be particularly prominent in gaze estimation based on small labeled datasets. The gaze estimation method needs to have a strong ability of derivative learning to obtain more representative derivative features using small labeled datasets, and expand the feature ranges covered by training samples through transformation between features. Or the gaze estimation model can be learned us-

Table 8
Comparison of gaze estimation methods.

Categories	2D mapping-based methods	3D model-based methods	Appearance-based methods
System configuration	Mainly with SCSLS, SCMLS	Mainly with SCMLS, MCMLS	Single/multi-camera or mobile devices
Research object	Pupil, iris, glints, eye corner	Pupil, iris, glints, facial landmarks	Eye images, face images
Personal calibration	Mapping model coefficients	Eye invariant parameters	Fine-tune the network
Head motion	Keep fixed or in a certain range	Limit the head rolling	Keep in a certain range
Gaze accuracy	Usually between 0.5°-2°	Usually between 0.5°-2°	Usually above 3°
Main error factors	Feature detection, head motion	Feature detection	Individual, head motion, environment
Robustness	General	Poor	Strong

ing small labeled datasets with the help of the features in existing resources.

(2) **Gaze estimation based on automatic or implicit calibration**: Although researchers have proposed some CNN architectures to achieve calibration-free gaze estimation, personal calibration is effective for improving the accuracy. The current implicit calibration method is mainly to obtain information from the subjects' daily habits without their attention. For example, assuming that the position where the subject clicks the mouse is the subject's gaze point at that moment [133,166]; the eye image is captured from the subject watching a video, and the corresponding relationship is established with the saliency map [11]. Obtaining some calibration samples through a simple implicit calibration, and learning the mapping model adaptively, can not only ensure the accuracy and increase the robustness, but also can improve the system automation extent.

(3) **Unconstrained gaze estimation**: Unconstrained gaze estimation requires that gaze estimation is completely independent of person, scene or device, and has significant performance for cross-dataset evaluation. Before the key issues such as head motion, individual differences, gaze range differences, and environmental differences are effectively handled, it is still difficult to achieve unconstrained gaze estimation. To meet the application requirements of natural HCI, unconstrained gaze estimation will be the goal gaze tracking has been pursuing.

4. Comparison of gaze estimation methods

2D mapping-based methods and 3D model-based methods study the pupil, iris, eye corners, and glints, while appearance-based methods are based on eye or face images. Their characteristics are summarized in Table 8.

The most simplified system of 2D mapping-based methods is a single-camera-non-light-source (SCNLS) system, which uses the PCT/ICT-based method for 2D POR estimation. The most commonly used system is a SCSLS or SCMLS system, where the PCRT/ICRT-based method, CR-based method or HN-based method can be used. When establishing a 2D mapping model, some parameters with individual differences need to be calibrated. 2D mapping-based methods can achieve high accuracy when the user's head is kept fixed or moves in a certain range. It is widely used in systems containing fixed components such as headrests and chin rests, as well as helmets and smart glasses. However, the natural head motion in remote gaze tracking systems will significantly reduce the accuracy.

3D model-based methods usually need to use a SCMLS or MCMLS system to estimate the required eyeball parameters for estimating the 3D LOS. Using head pose information or depth information, the system can be simplified to a SCNLS system or a depth sensor. The method needs to calibrate user-dependent eye invariant parameters before estimating the parameters required for OA reconstruction and 3D gaze estimation. The head rolling cannot be characterized based on eye features when it is analyzed solely from the eye image, therefore, 3D model-based methods usually restrict the head rolling. When the system consists of high-resolution cameras and infrared light sources and the features are extracted with high-precision, 3D model-based methods can reach an accuracy within 1°. They are often used in remote gaze tracking systems.

The system configuration of appearance-based methods is diversified. The images can be collected on mobile devices such as smartphones and tablets. To improve the cross-subject performance of the trained gaze estimation model, a few calibration samples can be collected to fine-tune the network. However, due to the limitations of the dataset, the results are greatly affected by head motion, individual differences, gaze range differences, environmental differences, etc. Even if a deeper network is used, the gaze accuracy of appearance-based methods cannot be comparable to feature-based methods, but the robustness is stronger than feature-based methods since the model is learned through a large number of training samples.

5. Applications of gaze tracking techniques

5.1. Applications factors

5.1.1. System type and gaze tracking accuracy

Gaze tracking systems are generally divided into head-mounted and remote systems. For practical applications, most of them use MCMLS systems to ensure the accuracy. Head-mounted gaze tracking systems follow the user's motion. The pupil and the glints on the corneal surface are obtained by the high-resolution near-eye cameras, and the gaze can be accurately estimated. For example, aSee Glasses[1] uses one camera and eight infrared light sources to track each eye, the accuracy is less than 0.5° based on corneal reflection. Pupil Invisible[2] is the world's first deep learning powered eye tracking glasses. In addition to a scene camera, two IR near-eye cameras are one on each side, and an IR LED is used for sufficient illumination of the respective eye region. It delivers unbiased gaze estimates with less than about 4° of random spread over most of the field of view, without the need of any calibration. Remote gaze tracking systems usually have their own operating distances. For example, the operating distance of Tobii Pro Fusion[3] is 50-80cm, and the head motion range is 40cm×25cm@65cm and 45cm×30cm@80cm. Based on pupil corneal reflection combined with bright and dark pupil tracking, the accuracy can reach 0.3° in optimal conditions. The number of cameras installed in Smart Eye Pro[4] is flexible, which determines the operating distance and tracking range. Its accuracy can reach 0.5°.

Current gaze tracking devices on the market, whether head-mounted or remote, use a tracking technique that is based on pupil/corneal reflection. They use high-quality system to accurately obtain the pupil and glints, and estimate the accurate gaze based

[1] https://pupil-labs.com/products/invisible/
[2] https://www.7invensun.com/yjsxqy
[3] https://www.tobiipro.cn/product-listing/tobii-pro-fusion/
[4] https://smarteye.se

Table 9
Comparison of computational resources and execution speed of different methods.

Categories	Refs	System type	Accuracy	Computational resources	Execution speed(fps)
2D mapping-based method	[13]	Remote	1.33°	2.5 GHz Core2 duo, 2 GB RAM	>100
	[27]	Head-mounted	0.6°	Intel i5-4200M CPU, 8GB RAM	25
	[34]	Remote	~ 1°	Intel i7 CPU(3.2GHz)	~ 30
	[39]	Remote	1.03°	2.67 GHz quad-core CPU, 4GB RAM	37
	[40]	Remote	0.49° ± 0.21°	Xeon2.8GHz CPU	83
	[41]	Remote	0.76°	3.4 GHz quad-core CPU, 8GB RAM	40
	[42]	Remote	1.36° ± 0.61° (STN method)	2.8 GHz quad-core CPU, 4GB RAM	43
	[43]	Remote	1.03°	3.5 GHz quad-core CPU, 8GB RAM	40
3D model-based method	[59]	Remote	X:0.69°,Y:0.71°	Dual processor Pentium with 1.7 GHz and 256MB memory	60
	[75]	Remote	1.38°-2.71°	Intel Core2 quad-core CPU Q9550 (2.83 GHz), GT310 Graphic adapter, 4GB RAM	12
	[68]	Remote	H:5.9°,V:4.4°	Intel Core i5-2400S CPU, 8GB RAM	3.03
	[86]	Remote	3.45°	Intel Core i7-4790 CPU(3.6GHz), 32GB RAM	28.6
	[70]	Remote	3.5°	Intel Core i7-4770 CPU(3.4GHz), 16GB RAM	30
Appearance-based method	[88]	Remote	2.58°	Mobile device	10-15
	[89]	Head-mounted	7.7°	Inter i7-6900K, Nvidia 1070, 64GB RAM	25.3
	[106]	Remote	7.8°	Single GTX Titan X GPU	20
	[107]	Remote	2.37° ± 1.42° (Slight head motion)	2.8GHz CPU, 8GB RAM	28
	[118]	Remote	H:1.3mm,V:1.5mm	Intel Core i5 CPU(2.5GHz), 4GB RAM	15-20
	[166]	Remote	2.9°	2.67GHz dual-core CPU, 3GB RAM	20
	[134]	Head-mounted	3.11°	Single Nvidia1080 Ti GPU	83

on 2D mapping and 3D model. The accuracy can reach about 0.5°. However, these devices often require a calibration process before they can be used. In different application scenarios, the selection of gaze tracking system should be determined by the specific requirements. Head-mounted systems allow natural head motion, which is more advantageous in the scenarios that require high stability and accuracy. Remote systems are contactless with the user, and are the best choice for mobile devices such as smartphones and tablets. In applications that do not require high precision, the system configuration can also be appropriately simplified to reduce the costs.

5.1.2. Computational resources and execution speed

When the gaze tracking system is based on the 2D mapping or 3D model, most of the time required for gaze tracking is used for feature detection, while the execution time of gaze estimation algorithms is very short. For example, Zhang et al. [29] mentioned that the execution times of their method for corneal reflection localization and iris center localization were less than 100ms and 150ms respectively, while the time for POR estimation was only less than 2ms. When the gaze tracking system is based on the appearance, especially based on CNN, more time needs to be used to train the model. After the model has been trained, the time required for gaze estimation can reach the millisecond [11]. Table 9 summarizes the computational resources and average execution speeds mentioned in some methods. The current eye trackers of the well-known brands such as Tobii and SmartEye can reach 60fps [29]. The 2D mapping-based method and the 3D model-based method can generally achieve an execution speed of more than 30fps in common personal computers. Among them, the DD-based method is limited by the resolution of depth sensor, its execution speed is relatively slow [75]. In comparison, the appearance-based method performs slower under the same computational resources, generally not exceeding 20fps. This indicates that the appearance-based method, in addition to its relatively low accuracy, still has great challenges to achieve unconstrained gaze estimation in mobile devices with less computational resources, and cannot satisfy comfortable and natural real-time interaction.

5.2. Application fields

Gaze tracking has been applied in HCI, medical diagnosis and treatment, intelligent transportation, virtual reality, and human factors analysis. According to its application type, some of the state-of-the-art algorithms are categorized in Table 10, indicating the fields in which they may be applied. 2D mapping-based methods are mainly applied in HCI, medical diagnosis, especially in virtual reality. Qian et al. [53] proposed an PCT-based virtual reality system to improve the anxiety of patients, especially vulnerable groups and children when they receive MRI examinations, and hence improve the effectiveness of the examination. The system accuracy reached 0.76° when the head motion range was within 30mm. Drakopoulos et al. [48] proposed a linear gaze mapping between the iris centroid and the POR for eye tracking interaction on mobile virtual reality headsets. The average accuracy was 1.17° and 1.86° in x- and y-axis respectively. 3D model-based methods are usually applied in HCI, medical diagnosis and treatment, human factors analysis, etc. Iqbal et al. [167] proposed user interface for hand-held mobile devices using CRPR-based method. The accuracy meets the requirements of many simple menu-driven user interface applications, even both glint and pupil center have pixel error. Wyder et al. [168] applied CRPR-based method to a navigation scheme for proton beam radiation of intraocular tumors, and reached an accuracy of 0.59mm on the depth of the retina. They integrated the gaze tracker into the proton beam radiation facility of the Paul Scherrer Institute in Villigen, Switzerland, for noninvasive proton therapy of the eye. It promotes the noninvasive of the proton therapy of the eye. Appearance-based methods are used in HCI, intelligent transportation, human factors analysis, etc. Steil et al. [169] extracted rich semantic information about the users vi-

Table 10
Classification of possible applications of different gaze tracking algorithms.

Methods / Applications / Categories	HCI	Medical diagnosis and treatment	Intelligent transportation	Virtual reality	Human factor analysis
2D mapping-based methods	[13], [15], [17], [25], [21], [24], [34], [54], [32], [35], [36], [39], [55], [41], [50], [51]	[27], [2], [172], [173]	[52]	[56], [53], [48]	-
3D model-based methods	[8], [64], [87], [65], [80], [83], [167], [174]	[168]	-	[84], [86]	[77], [78], [68], [69], [79], [175]
Appearance-based methods	[97], [88], [96], [89], [153], [154], [138], [130], [121], [124], [119], [131], [133], [169], [134]	-	[12], [115], [111], [126], [125], [127], [170], [176], [177]	[116], [117], [150], [128]	[10], [90], [108], [114], [109], [126], [157], [171]

sual scene for object and face detection, semantic scene segmentation and depth reconstruction, and used these features to predict attentive behavior during everyday mobile interactions from real phone-integrated and body-worn sensors. Hu et al. [170] extracted the low-level features, static visual saliency map and dynamic optical flow information as input feature maps, which combined the high-level semantic descriptions and a gaze probability map transformed from the gaze direction, and proposed a multi-resolution neural network for driver attention estimation. O'Dwyer et al. [171] presented a continuous affect prediction system using the combination of eye gaze and speech. They extracted 31 eye gaze features to fuse with and 2,268 speech features. It shows a 19.5% improvement in valence prediction and a 3.5% improvement in arousal prediction compared with using speech alone in a bi-modal system.

6. Conclusion

In this paper, we discuss 2D mapping-based methods, 3D model-based methods, and appearance-based methods in gaze tracking, involving the adopted visual information, commonly used algorithms, key issues, research trends, and practical applications, etc. The purpose of this paper is to comprehensively clarify the research status of gaze tracking, put forward some targeted research goals, and promote the development of gaze tracking techniques.

Although gaze tracking has been applied in many fields, it still belongs to a dedicated device. In order to promote gaze tracking to be an important channel for natural interaction in daily life, the following directions can be studied:

(1) *Independent of person, head pose or environment*: The eyeball parameters and gaze states of different people are different, and the head pose and environment of people in daily life are also full of diversity. At present, these issues are mainly addressed by using a high-quality system and perform personal calibration. But this is not applicable to all fields, such as for visual interaction in mobile devices. Investigating gaze estimation methods independent of person, head pose, or environment is the key to ensuring universal and stable application of gaze tracking.

(2) *Simplify or eliminate personal calibration*: To improve gaze tracking performance, personal calibration is required in many cases, but this is contrary to the pursuit of natural HCI. Research on a more natural personal calibration method can improve the operation and real-time performance of gaze tracking, and realize the instant use.

(3) *Improve execution rate*: Appearance-based methods seem to be best suited for natural interaction since they have low system requirements and strong robustness. However, they need to process a large number of samples to train the gaze estimation model, which requires more computational resources. It is of great significance to achieve accurate, real-time visual interaction based on a small number of samples. Combining features and appearance organically can be considered to balance the gaze tracking performance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key Research and Development Program of China [grant numbers 2018YFC2001700]; the Beijing Municipal Natural Science Foundation [grant numbers 4212023]; the Scientific and Technological Innovation Foundation of Shunde Graduate School, USTB; the Foundation of Engineering Research Center of Intelligence Perception and Autonomous Control, Ministry of Education, P. R. China; and the Fundamental Research Funds for the Central Universities [grant numbers FRF-GF-20-04A].

References

[1] D.W. Hansen, Q. Ji, In the eye of the beholder: A survey of models for eyes and gaze, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (3) (2010) 478–500.
[2] K. Luo, X. Jia, H. Xiao, D. Liu, P. Han, A new gaze estimation method based on homography transformation derived from geometric relationship, Applied Sciences 10 (24) (2020) 9079.
[3] N.M. Bakker, B.A.J. Lenseigne, S. Schutte, E.B.M. Geukers, P.P. Jonker, F.C.T. van der Helm, H.J. Simonsz, Accurate gaze direction measurements with free head movement for strabismus angle estimation, IEEE Transactions on Biomedical Engineering 60 (11) (2013) 3028–3035.
[4] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Mpiigaze: Real-world dataset and deep appearance-based gaze estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (1) (2019) 162–175.
[5] S. Cristina, K.P. Camilleri, Unobtrusive and pervasive video-based eye-gaze tracking, Image and Vision Computing 74 (2018) 21–40.
[6] Y. Cheng, H. Wang, Y. Bao, F. Lu, Appearance-based gaze estimation with deep learning: A review and benchmark, CoRR abs/2104.12668 (2021).
[7] A.F. Klaib, N.O. Alsrehin, W.Y. Melhem, H.O. Bashtawi, A.A. Magableh, Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and internet of things technologies, Expert Systems with Applications 166 (1) (2021) 114037.
[8] J. Liu, J. Chi, N. Lu, Z. Yang, Z. Wang, Iris feature-based 3-d gaze estimation method using a one-camera-one-light-source system, IEEE Transactions on Instrumentation and Measurement 69 (7) (2020) 4940–4954.
[9] F. Lu, Y. Gao, X. Chen, Estimating 3d gaze directions using unlabeled eye images via synthetic iris appearance fitting, IEEE Transactions on Multimedia 18 (9) (2016) 1772–1782.
[10] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, 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 (CVPR) Workshops, IEEE, 2017, pp. 2299–2308.

[11] Y. Sugano, Y. Matsushita, Y. Sato, Appearance-based gaze estimation using visual saliency, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2) (2013) 329–341, doi:10.1109/TPAMI.2012.101.
[12] F. Lu, X. Chen, Y. Sato, Appearance-based gaze estimation via uncalibrated gaze pattern recovery, IEEE Transactions on Image Processing 26 (4) (2017) 1543–1553.
[13] A. George, A. Routray, Fast and accurate algorithm for eye localisation for gaze tracking in low-resolution images, IET Computer Vision 10 (7) (2016) 660–669.
[14] M.X. Yu, Y.Z. Lin, X.Y. Tang, J. Xu, D. Schmidt, X.Z. Wang, Y. Guo, An easy iris center detection method for eye gaze tracking system, Journal of Eye Movement Research 8 (3) (2015) 1–20.
[15] H. Cai, H. Yu, X. Zhou, H. Liu, Robust gaze estimation via normalized iris center-eye corner vector, in: International Conference on Intelligent Robotics and Applications, volume 9834, Springer, Cham, 2016, pp. 300–309.
[16] Y.M. Cheung, Q.M. Peng, Eye gaze tracking with a web camera in a desktop environment, IEEE Transactions on Human-Machine Systems 45 (4) (2015) 419–430.
[17] D. Hu, H. Qin, H. Liu, S. Zhang, Gaze tracking algorithm based on projective mapping correction and gaze point compensation in natural light*, in: 2019 IEEE 15th International Conference on Control and Automation (ICCA), IEEE, Edinburgh, UK, 2019, pp. 1150–1155.
[18] S. Rattarom, N. Aunsri, S. Uttama, A framework for polynomial model with head pose in low cost gaze estimation, in: 2017 International Conference on Digital Arts, Media and Technology (ICDAMT), IEEE, Chiang Mai, Thailand, 2017, pp. 24–27.
[19] P. Blignaut, Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker, Journal of Eye Movement Research 7 (1) (2013) 1–11.
[20] J.J. Cerrolaza, A. Villanueva, R. Cabeza, Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems, in: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA, 2008, pp. 259–266.
[21] K. Koshikawa, M. Sasaki, T. Utsu, K. Takemura, Polarized near-infrared light emission for eye gaze estimation, in: ACM Symposium on Eye Tracking Research and Applications, Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–4.
[22] J. Sigut, S.A. Sidha, Iris center corneal reflection method for gaze tracking using visible light, IEEE Transactions on Biomedical Engineering 58 (2) (2011) 411–419.
[23] C. Zhang, J.N. Chi, Z.H. Zhang, X. Gao, T. Hu, Z.L. Wang, Gaze estimation in a gaze tracking system, Science China Information Sciences 54 (11) (2011) 2295–2306.
[24] Y. Hu, J. Wei, S. Mei, Gaze estimation method based on pupils and corneal reflection technique, Computer Engineering and Applications 54 (14) (2018) 7–10.
[25] J.H. Park, J.B. Park, A novel approach to the low cost real time eye mouse, Computer Standards and Interfaces 44 (2016) 169–176.
[26] J. Wang, G. Zhang, J. Shi, 2d gaze estimation based on pupil-glint vector using an artificial neural network, Applied Sciences 6 (174) (2016) 1–17.
[27] C. Mestre, J. Gautier, J. Pujol, Robust eye tracking based on multiple corneal reflections for clinical applications, Journal of Biomedical Optics 23 (3) (2018) 035001.
[28] L. Sesma-Sanchez, A. Villanueva, R. Cabeza, Gaze estimation interpolation methods based on binocular data, IEEE Transactions on Biomedical Engineering 59 (8) (2012) 2235–2243.
[29] T.N. Zhang, J.J. Bai, C.N. Meng, S.J. Chang, Eye-gaze tracking based on one camera and two light sources, Journal of Optoelectronics Laser 23 (10) (2012) 1990–1995.
[30] Z. Zhu, Q. Ji, Novel eye gaze tracking techniques under natural head movement, IEEE Transactions on Biomedical Engineering 54 (12) (2008) 2246–2260.
[31] F.L. Coutinho, C.H. Morimoto, Augmenting the robustness of cross-ratio gaze tracking methods to head movement, in: Proceedings of the Symposium on Eye Tracking Research and Applications, Association for Computing Machinery, New York, NY, USA, 2012, pp. 59–66.
[32] H. Cheng, Y. Liu, W. Fu, Y. Ji, Y. Lu, Z. Yang, Y. Jie, Gazing point dependent eye gaze estimation, Pattern Recognition 71 (2017) 36–44.
[33] N.M. Arar, H. Gao, J.-P. Thiran, Towards convenient calibration for cross-ratio based gaze estimation, in: 2015 IEEE Winter Conference on Applications of Computer Vision, IEEE, Waikoloa, HI, USA, 2015, pp. 642–648.
[34] N.M. Arar, H. Gao, J.P. Thiran, A regression-based user calibration framework for real-time gaze estimation, IEEE Transactions on Circuits and Systems for Video Technology 27 (12) (2017) 2623–2638.
[35] M. Sasaki, T. Nagamatsu, K. Takemura, Cross-ratio based gaze estimation for multiple displays using a polarization camera, in: The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–3.
[36] M. Sasaki, T. Nagamatsu, K. Takemura, Screen corner detection using polarization camera for cross-ratio based gaze estimation, in: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA, 2019.
[37] D.W. Hansen, J.S. Agustin, A. Villanueva, Homography normalization for robust gaze estimation in uncalibrated setups, in: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA, 2010, pp. 13–20.
[38] C.H. Morimoto, F.L. Coutinho, W.H. Dan, Screen-light decomposition framework for point-of-gaze estimation using a single uncalibrated camera and multiple light sources, Journal of Mathematical Imaging and Vision 62 (12) (2020) 585–605.
[39] K.A. Choi, C. Ma, S.J. Ko, Improving the usability of remote eye gaze tracking for human-device interaction, IEEE Transactions on Consumer Electronics 60 (3) (2014) 493–498.
[40] F.L. Coutinho, C.H. Morimoto, Improving head movement tolerance of cross-ratio based eye trackers, International Journal of Computer Vision 101 (3) (2013) 459–481.
[41] S.T. Kim, K.A. Choi, Y.G. Shin, M.C. Kang, S.J. Ko, Remote eye-gaze tracking method robust to the device rotation, Optical Engineering 55 (8) (2016) 083108.
[42] C. Ma, S.J. Baek, K.A. Choi, S.J. Ko, Improved remote gaze estimation using corneal reflection-adaptive geometric transforms, Optical Engineering 53 (5) (2014) 053112.
[43] Y.G. Shin, K.A. Choi, S.T. Kim, S.J. Ko, A novel single ir light based gaze estimation method using virtual glints, IEEE Transactions on Consumer Electronics 61 (2) (2015) 254–260.
[44] L. Xia, B. Sheng, W. Wu, L. Ma, P. Li, Accurate gaze tracking from single camera using gabor corner detector, Multimedia Tools and Applications 75 (1) (2016) 221–239.
[45] J.H. Kim, D.I. Han, C.O. Min, D.W. Lee, W.S. Eom, Ir vision-based los tracking using non-uniform illumination compensation, International Journal of Precision Engineering and Manufacturing 14 (8) (2013) 1355–1360.
[46] S.Y. Han, N.I. Cho, User-independent gaze estimation by extracting pupil parameter and its mapping to the gaze angle, in: 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 1993–2000.
[47] E. Skodras, V.G. Kanas, N. Fakotakis, On visual gaze tracking based on a single low cost camera, Signal Processing: Image Communication 36 (2015) 29–42.
[48] P. Drakopoulos, G.-A. Koulieris, K. Mania, Eye tracking interaction on unmodified mobile vr headsets using the selfie camera, ACM Transactions on Applied Perception 18 (3) (2021) Article11.
[49] R. Banaeeyan, A.A. Halin, M. Bahari, Nonintrusive eye gaze tracking using a single eye image, in: 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), IEEE, Kuala Lumpur, Malaysia, 2015, pp. 139–144.
[50] Y. Eom, S. Mu, S. Satoru, T. Liu, A method to estimate eye gaze direction when wearing glasses, in: 2019 International Conference on Technologies and Applications of Artificial Intelligence, IEEE, Kaohsiung, Taiwan, 2019, pp. 1–6.
[51] B. Yan, S. Yu, T. Pei, Y. Hu, Vision interaction method based on visual attention mechanism, in: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, Siem Reap, Cambodia, 2017, pp. 930–935.
[52] H.S. Yoon, H.G. Hong, D.E. Lee, K.R. Park, Driver's eye-based gaze tracking system by one-point calibration, Multimedia Tools and Applications 78 (6) (2019) 7155–7179.
[53] K. Qian, T. Arichi, A. Price, S. Dall'Orso, J. Eden, Y. Noh, K. Rhode, E. Burdet, M. Neil, A.D. Edwards, J. Hajnal, An eye tracking based virtual reality system for use inside magnetic resonance imaging systems, Scientific Reports 11 (16301) (2021).
[54] K. Huang, M. Khalil, E. Luciani, D. Melesse, T. Ning, A data-driven approach for gaze tracking, in: 2018 14th IEEE International Conference on Signal Processing (ICSP), IEEE, Beijing, China, 2018, pp. 494–499.
[55] K.-H. Uhm, M.-C. Kang, J.-Y. Kim, S.-J. Ko, Improving the robustness of gaze tracking under unconstrained illumination conditions, Multimedia Tools and Applications 79 (2) (2020) 20603–20616.
[56] C.-L. Jen, Y.-L. Chen, Y.-J. Lin, C.-H. Lee, A. Tsai, M.-T. Li, Vision based wearable eye-gaze tracking system, in: 2016 IEEE International Conference on Consumer Electronics (ICCE), 2016, pp. 202–203.
[57] E.D. Guestrin, M. Eizenman, General theory of remote gaze estimation using the pupil center and corneal reflections, IEEE Transactions on Biomedical Engineering 53 (6) (2006) 1124–1133.
[58] T. Ohno, N. Mukawa, A. Yoshikawa, Freegaze: A gaze tracking system for everyday gaze interaction, in: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA, 2002, pp. 125–132.
[59] A. Villanueva, R. Cabeza, A novel gaze estimation system with one calibration point, IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics 38 (4) (2008) 1123–1138.
[60] J. Chi, J. Liu, F. Wang, Y. Chi, Z.G. Hou, 3d gaze estimation method using a multi-camera-multi-light-source system, IEEE Transactions on Instrumentation and Measurement 69 (12) (2020) 9695–9708.
[61] K. Zhang, X. Zhao, Z. Ma, Y. Man, A simplified 3d gaze tracking technology with stereo vision, in: 2010 International Conference on Optoelectronics and Image Processing, IEEE, Haikou, China, 2011, pp. 131–134.
[62] C.C. Lai, S.W. Shih, Y.P. Hung, Hybrid method for 3-d gaze tracking using glint and contour features, IEEE Transactions on Circuits and Systems for Video Technology 25 (1) (2015) 24–37.
[63] C.-C. Lai, S.-W. Shih, H.-R. Tsai, Y.-P. Hung, 3-d gaze tracking using pupil contour features, in: 2014 22nd International Conference on Pattern Recognition, IEEE, Stockholm, Sweden, 2014, pp. 1162–1166.
[64] J. Liu, J. Chi, W. Hu, Z. Wang, 3d model-based gaze tracking via iris features with a single camera and a single light source, IEEE Transactions on Human–Machine Systems 51 (2) (2021) 75–86.
[65] J. O'Reilly, A.S. Khan, Z. Li, J. Cai, X. Hu, M. Chen, Y. Tong, A novel remote eye gaze tracking system using line illumination sources, in: 2019 IEEE Confer-

ence on Multimedia Information Processing and Retrieval (MIPR), IEEE, San Jose, CA, USA, 2019, pp. 449–454.
[66] M. Lidegaard, D.W. Hansen, N. Krüger, Head mounted device for point-of-gaze estimation in three dimensions, in: Proceedings of the Symposium on Eye Tracking Research and Applications, Association for Computing Machinery, New York, NY, USA, 2014, pp. 83–86.
[67] J. Li, S. Li, Eye-model-based gaze estimation by rgb-d camera, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, Columbus, USA, 2014, pp. 592–596.
[68] J. Li, S. Li, Gaze estimation from color image based on the eye model with known head pose, IEEE Transactions on Human-Machine Systems 46 (3) (2016) 414–423.
[69] A. Kacete, R. Séguier, M. Collobert, J. Royan, Head pose free 3d gaze estimation using rgb-d camera, in: Eighth International Conference on Graphic and Image Processing, volume 10225, SPIE, Tokyo, Japan, 2017, p. 102251S.
[70] K. Wang, Q. Ji, Real time eye gaze tracking with 3d deformable eye-face model, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, Italy, 2017, pp. 1003–1011.
[71] J. Chen, Q. Ji, 3d gaze estimation with a single camera without ir illumination, in: 2008 19th International Conference on Pattern Recognition, IEEE, Tampa, FL, USA, 2008, pp. 1–4.
[72] L.A. Jeni, J.F. Cohn, Person-independent 3d gaze estimation using face frontalization, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Las Vegas, NV, USA, 2016, pp. 792–800.
[73] S. Cristina, K.P. Camilleri, Model-based head pose-free gaze estimation for assistive communication, Computer Vision and Image Understanding 149 (2016) 157–170.
[74] L. Sun, M. Song, Z. Liu, M.T. Sun, Real-time gaze estimation with online calibration, IEEE Multimedia 21 (4) (2014) 28–37.
[75] L. Sun, Z. Liu, M.T. Sun, Real time gaze estimation with a consumer depth camera, Information Sciences 320 (1) (2015) 346–360.
[76] K. Wang, Q. Ji, Real time eye gaze tracking with kinect, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, Cancun, 2016, pp. 2752–2757.
[77] X. Zhou, H. Cai, Z. Shao, H. Yu, H. Liu, 3d eye model-based gaze estimation from a depth sensor, in: 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, Qingdao, China, 2016, pp. 369–374.
[78] X. Zhou, H. Cai, Y. Li, H. Liu, Two-eye model-based gaze estimation from a kinect sensor, in: 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, Singapore, 2017, pp. 1646–1653.
[79] M. Mansouryar, J. Steil, Y. Sugano, A. Bulling, 3d gaze estimation from 2d pupil positions on monocular head-mounted eye trackers, in: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA, 2016, pp. 197–200.
[80] Z. Wan, X. Wang, L. Yin, K. Zhou, A method of free-space point-of-regard estimation based on 3d eye model and stereo vision, Applied Sciences 8 (10) (2018) 1769.
[81] R. Jafari, D. Ziou, Eye-gaze estimation under various head positions and iris states, Expert Systems with Applications 42 (1) (2015) 510–518.
[82] K. Cen, M. Che, Research on eye gaze estimation technique base on 3d model, in: 2011 International Conference on Electronics, Communications and Control (ICECC), IEEE, Ningbo, China, 2011, pp. 1623–1626.
[83] K. Wang, Q. Ji, 3d gaze estimation without explicit personal calibration, Pattern Recognition 79 (2018) 216–227.
[84] S.Y. Han, S.H. Lee, N.I. Cho, Gaze estimation using 3-d eyeball model under hmd circumstance, in: 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), IEEE, Luton, UK, 2017, pp. 1–4.
[85] L. El Hafi, K. Takemura, J. Takamatsu, T. Ogasawara, Model-based approach for gaze estimation from corneal imaging using a single camera, in: 2015 IEEE/SICE International Symposium on System Integration (SII), IEEE, Nagoya, Japan, 2015, pp. 88–93.
[86] Q. Wen, D. Bradley, T. Beeler, S. Park, F. Xu, Accurate realtime 3d gaze tracking using a lightweight eyeball calibration, Computer Graphics Forum 39 (2) (2020) 475–485.
[87] B. Brousseau, J. Rose, M. Eizenman, Accurate model-based point of gaze estimation on mobile devices, Vision 2 (3) (2018) 35.
[88] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, A. Torralba, Eye tracking for everyone, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016, pp. 2176–2184.
[89] T. Fischer, H.J. Chang, Y. Demiris, Rt-gene: Real-time eye gaze estimation in natural environments, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 334–352.
[90] P. Kellnhofer, A. Recasens, S. Stent, W. Matusik, A. Torralba, Gaze360: Physically unconstrained gaze estimation in the wild, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, Korea (South), 2019, pp. 6912–6921.
[91] B.A. Smith, Q. Yin, S.K. Feiner, S.K. Nayar, Gaze locking: Passive eye contact detection for human-object interaction, in: Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, Association for Computing Machinery, New York, NY, USA, 2013, pp. 271–280.
[92] K.A. Funes Mora, F. Monay, J.-M. Odobez, 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, Association for Computing Machinery, New York, NY, USA, 2014, pp. 255–258.
[93] Y. Sugano, Y. Matsushita, Y. Sato, Learning-by-synthesis for appearance-based 3d gaze estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2014, pp. 1821–1828.
[94] X. Zhang, Y. Sugano, M. Fritz, A. Bulling, Appearance-based gaze estimation in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, 2015, pp. 4511–4520.
[95] Q. He, X. Hong, X. Chai, J. Holappa, M. Pietikinen, Omeg: Oulu multi-pose eye gaze dataset, in: Scandinavian Conference on Image Analysis, volume 9127, Springer, New York, NY, USA, 2015, pp. 418–427.
[96] M. Tonsen, J. Steil, Y. Sugano, A. Bulling, Invisibleeye: Mobile eye tracking using multiple low-resolution cameras and learning-based gaze estimation, Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 1 (3) (2017) 1–21.
[97] Q. Huang, A. Veeraraghavan, A. Sabharwal, Tabletgaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets, Machine Vision and Applications 28 (5-6) (2017) 445–461.
[98] J. Kim, M. Stengel, A. Majercik, S.D. Mello, D. Luebke, Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–12.
[99] D. Lian, L. Hu, W. Luo, Y. Xu, S. Gao, Multiview multitask gaze estimation with deep convolutional neural networks, IEEE Transactions on Neural Networks and Learning Systems 30 (10) (2019) 3010–3023.
[100] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, O. Hilliges, 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 (ECCV), volume 12350, Springer, Cham, 2020, pp. 365–381.
[101] S. Park, E. Aksan, X. Zhang, O. Hilliges, Towards end-to-end video-based eye–tracking, in: European Conference on Computer Vision (ECCV), volume 12357, Springer, Cham, 2020, pp. 747–763.
[102] C. Palmero, A. Sharma, K. Behrendt, K. Krishnakumar, S.S. Talathi, Openeds2020: Open eyes dataset, CoRR abs/2005.03876 (2020).
[103] E. Wood, T. Baltruaitis, X. Zhang, Y. Sugano, P. Robinson, A. Bulling, Rendering of eyes for eye-shape registration and gaze estimation, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE, Santiago, Chile, 2015, pp. 3756–3764.
[104] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, A. Bulling, Learning an appearance-based gaze estimator from one million synthesised images, in: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research and Applications, Association for Computing Machinery, New York, NY, USA, 2016, pp. 131–138.
[105] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, R. Webb, Learning from simulated and unsupervised images through adversarial training, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, USA, 2017, pp. 2242–2251.
[106] Y. Yan, Y. Yan, J. Peng, H. Wang, X. Fu, Purifying real images with an attention-guided style transfer network for gaze estimation, Engineering Applications of Artificial Intelligence 91 (May) (2020) 103609.1–103609.9.
[107] F. Lu, Y. Sugano, T. Okabe, Y. Sato, Adaptive linear regression for appearance-based gaze estimation, IEEE Transactions on Pattern Analysis & Machine Intelligence 36 (10) (2014) 2033–2046.
[108] Y. Wang, T. Zhao, X. Ding, J. Peng, J. Bian, X. Fu, Learning a gaze estimator with neighbor selection from large-scale synthetic eye images, Knowledge-Based Systems 139 (2018) 41–49.
[109] A. Kacete, R. Séguier, M. Collobert, J. Royan, Unconstrained gaze estimation using random forest regression voting, in: Asian Conference on Computer Vision, Springer, Cham, Taipei, Taiwan, ROC, 2016, pp. 419–432.
[110] O. Ferhat, F. Vilarino, F. Sánchez, A cheap portable eye-tracker solution for common setups, Journal of Eye Movement Research 7 (3) (2014) 1–10.
[111] M. Shirpour, S.S. Beauchemin, M.A. Bauer, A probabilistic model for visual driver gaze approximation from head pose estimation, in: 2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS), IEEE, Victoria, BC, Canada, 2020, pp. 1–6.
[112] Y.L. Wu, C.T. Yeh, W.-C. Hung, Gaze direction estimation using support vector machine with active appearance model, Multimedia Tools and Applications 70 (3) (2014) 2037–2062.
[113] H.C. Lu, G.L. Fang, C. Wang, Y.W. Chen, A novel method for gaze tracking by local pattern model and support vector regressor, Signal Processing 90 (4) (2010) 1290–1299.
[114] Y. Wang, T. Shen, G. Yuan, J. Bian, X. Fu, Appearance-based gaze estimation using deep features and random forest regression, Knowledge-Based Systems 110 (2016) 293–301.
[115] X. Shan, Z. Wang, X. Liu, M. Lin, L. Zhao, J. Wang, G. Wang, Driver gaze region estimation based on computer vision, in: 2020 12th International Conference on Measuring Technology and Mechatronics Automation, IEEE, Phuket, Thailand, 2020, pp. 357–360.
[116] D. Su, Y.F. Li, H. Chen, Toward precise gaze estimation for mobile head–mounted gaze tracking systems, IEEE Transactions on Industrial Informatics 15 (5) (2019) 2660–2672.
[117] K. Akşit, J. Kautz, D. Luebke, Gaze-sensing leds for head mounted displays, CoRR abs/2003.08499 (2020).
[118] M.C. Chuang, R. Bala, E. Bernal, P. Paul, A. Burry, Estimating gaze direction of vehicle drivers using a smartphone camera, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, IEEE, Columbus, OH, USA, 2014, pp. 165–170.

[119] R.S. Ghiass, O. Arandjelovic, Highly accurate gaze estimation using a consumer rgb-d sensor, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, AAAI Press, 2016, pp. 3368–3374.
[120] F. Lu, X. Chen, Person-independent eye gaze prediction from eye images using patch-based features, Neurocomputing 182 (2016) 10–17.
[121] K.A.F. Mora, J.M. Odobez, Gaze estimation in the 3d space using rgb-d sensors, International Journal of Computer Vision 118 (2) (2016) 194–216.
[122] J. Pi, B.E. Shi, Task-embedded online eye-tracker calibration for improving robustness to head motion, in: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA, 2019.
[123] R. Valenti, N. Sebe, T. Gevers, Combining head pose and eye location information for gaze estimation, IEEE Transactions on Image Processing 21 (2) (2012) 802–815.
[124] G. Yuan, Y. Wang, T. Zhao, X. Ding, Z. Mi, X. Fu, Eye gaze region estimation via multi-scale sparse dictionary learning, in: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, Xi'an, China, 2019, pp. 1459–1463.
[125] Y. Zhuang, Y. Zhang, H. Zhao, Appearance-based gaze estimation using separable convolution neural networks, in: 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), volume 5, IEEE, Chongqing, China, 2021, pp. 609–612.
[126] H. Deng, W. Zhu, Monocular free-head 3d gaze tracking with deep learning and geometry constraints, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, Italy, 2017, pp. 3143–3152.
[127] Z. Wang, J. Zhao, C. Lu, H. Huang, F. Yang, L. Li, Y. Guo, Learning to detect head movement in unconstrained remote gaze estimation in the wild, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE, Snowmass, CO, USA, 2020, pp. 3443–3452.
[128] K. Wang, H. Su, Q. Ji, Neuro-inspired eye tracking with eye movement dynamics, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019, pp. 9831–9840.
[129] C. Palmero, J. Selva, M.A. Bagheri, S. Escalera, Recurrent cnn for 3d gaze estimation using appearance and shape cues, CoRR abs/1805.03064 (2018).
[130] S. Gu, L. Wang, L. He, X. He, J. Wang, Gaze estimation via a differential eyes' appearances network with a reference grid, Engineering 7 (6) (2021) 777–786.
[131] S. Park, X. Zhang, A. Bulling, O. Hilliges, Learning to find eye region landmarks for remote gaze estimation in unconstrained settings, in: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Association for Computing Machinery, New York, NY, USA, 2018.
[132] G. Liu, Y. Yu, K.A.F. Mora, J.M. Odobez, A differential approach for gaze estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (3) (2021) 1092–1099.
[133] B. Klein Salvalaio, G. de Oliveira Ramos, Self-adaptive appearance-based eye--tracking with online transfer learning, in: 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), IEEE, Salvador, Brazil, 2019, pp. 383–388.
[134] Z. Wu, S. Rajendran, T. Van As, V. Badrinarayanan, A. Rabinovich, Eyenet: A multi-task deep network for off-axis eye gaze estimation, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), IEEE, Seoul, Korea (South), 2019, pp. 3683–3687.
[135] C. Akinlar, H.K. Kucukkartal, C. Topal, Accurate cnn-based pupil segmentation with an ellipse fit error regularization term, Expert Systems with Applications 188 (2022) 116004.
[136] W. Fuhl, M. Tonsen, A. Bulling, E. Kasneci, Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art, Machine Vision & Applications 27 (2016) 1275–1288.
[137] S.Y. Han, H.J. Kwon, Y. Kim, N.I. Cho, Noise-robust pupil center detection through cnn-based segmentation with shape-prior loss, IEEE Access 8 (2020) 64739–64749.
[138] Y. Cheng, X. Zhang, F. Lu, Y. Sato, Gaze estimation by exploring two-eye asymmetry, IEEE Transactions on Image Processing 29 (2020) 5259–5272.
[139] Y. Cheng, L. Feng, X. Zhang, Appearance-based gaze estimation via evaluation-guided asymmetric regression, in: Proceedings of the European Conference on Computer Vision (ECCV), Springer, Cham, 2018, pp. 100–115.
[140] Y. Yu, G. Liu, J.M. Odobez, Improving few-shot user-specific gaze adaptation via gaze redirection synthesis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019, pp. 11937–11946.
[141] E. Lindén, J. Sjöstrand, A. Proutiere, Learning to personalize in appearance-based gaze tracking, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, South Korea, 2019.
[142] J. He, K. Pham, N. Valliappan, P. Xu, V. Navalpakkam, On-device few-shot personalization for real-time gaze estimation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, South Korea, 2019.
[143] K. Wang, R. Zhao, H. Su, Q. Ji, Generalizing eye tracking with bayesian adversarial learning, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019, pp. 11899–11908.
[144] Y. Xiong, H.J. Kim, V. Singh, Mixed effects neural networks (menets) with applications to gaze estimation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019, pp. 7735–7744.
[145] D. Kononenko, V. Lempitsky, Semi-supervised learning for monocular gaze redirection, in: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), IEEE, Xi'an, China, 2018, pp. 535–539.
[146] R. Kothari, S. De Mello, U. Iqbal, W. Byeon, S. Park, J. Kautz, Weakly-supervised physically unconstrained gaze estimation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, 2021, pp. 9975–9984.
[147] Z. Guo, Z. Yuan, C. Zhang, W. Chi, Y. Ling, S. Zhang, Domain adaptation gaze estimation by embedding with prediction consistency, ACCV 2020, 2020.
[148] Y. Yu, J.M. Odobez, Unsupervised representation learning for gaze estimation, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, 2020, pp. 7314–7324.
[149] N. Dubey, S. Ghosh, A. Dhall, Unsupervised learning of eye gaze representation from the web, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, Budapest, Hungary, 2019, pp. 1–7.
[150] Y. Yu, G. Liu, J.-M. Odobez, Deep multitask gaze estimation with a constrained landmark-gaze model, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Springer, Cham, Munich, Germany, 2018, pp. 456–474.
[151] S. Park, A. Spurr, O. Hilliges, Deep pictorial gaze estimation, in: Proceedings of the European Conference on Computer Vision (ECCV), volume 11217, Springer, Cham, 2018, pp. 741–757.
[152] C. Zhang, R. Yao, J. Cai, Efficient eye typing with 9-direction gaze estimation, Multimedia Tools & Applications 77 (2018) 19679–19696.
[153] X. Cha, X. Yang, Z. Feng, T. Xu, X. Fan, J. Tian, Calibration-free gaze zone estimation using convolutional neural network, in: 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), IEEE, Jinan, China, 2018, pp. 481–484.
[154] J. Lemley, A. Kar, A. Drimbarean, P. Corcoran, Convolutional neural network implementation for eye-gaze estimation on low-quality consumer imaging systems, IEEE Transactions on Consumer Electronics 65 (2) (2019) 179–187.
[155] X. Zhang, Y. Sugano, A. Bulling, O. Hilliges, Learning-based region selection for end-to-end gaze estimation, in: British Machine Vision Virtual Conference (BMVC), 2020.
[156] S. Park, S.D. Mello, P. Molchanov, U. Iqbal, O. Hilliges, J. Kautz, Few-shot adaptive gaze estimation, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Seoul, Korea (South), 2019, pp. 9367–9376.
[157] S. Jyoti, A. Dhall, Automatic eye gaze estimation using geometric & texture-based networks, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, Beijing, China, 2018, pp. 2474–2479.
[158] R. Ogusu, T. Yamanaka, Lpm: Learnable pooling module for efficient full-face gaze estimation, in: 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), IEEE, Lille, France, 2019, pp. 1–5.
[159] Y. Cheng, S. Huang, F. Wang, C. Qian, F. Lu, A coarse-to-fine adaptive network for appearance-based gaze estimation, Proceedings of the AAAI Conference on Artificial Intelligence 34 (07) (2020) 10623–10630.
[160] Y. Bao, Y. Cheng, Y. Liu, F. Lu, Adaptive feature fusion network for gaze tracking in mobile tablets, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, Milan, Italy, 2021, pp. 9936–9943.
[161] Z. Chen, B.E. Shi, Appearance-based gaze estimation using dilated-convolutions, in: Asian Conference on Computer Vision, volume 11366, Springer, Cham, 2018, pp. 309–324.
[162] Y. Zhang, X. Yang, Z. Ma, Driver's gaze zone estimation method: A four-channel convolutional neural network model, in: 2020 2nd International Conference on Big-Data Service and Intelligent Computation, Association for Computing Machinery, New York, NY, USA, 2020, pp. 20–24.
[163] Z. Chen, B.E. Shi, Offset calibration for appearance-based gaze estimation via gaze decomposition, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, IEEE, Snowmass, CO, 2020, pp. 270–279.
[164] Y. Cheng, F. Lu, Gaze estimation using transformer, CoRR abs/2105.14424 (2021).
[165] X. Cai, B. Chen, J. Zeng, J. Zhang, Y. Sun, X. Wang, Z. Ji, X. Liu, X. Chen, S. Shan, Gaze estimation with an ensemble of four architectures, CoRR abs/2107.01980 (2021).
[166] Y. Sugano, Y. Matsushita, Y. Sato, H. Koike, Appearance-based gaze estimation with online calibration from mouse operations, IEEE Transactions on Human–Machine Systems 45 (6) (2015) 750–760.
[167] N. Iqbal, H. Lee, S.-Y. Lee, Smart user interface for mobile consumer devices using model-based eye-gaze estimation, IEEE Transactions on Consumer Electronics 59 (1) (2013) 161–166.
[168] S. Wyder, F. Hennings, S. Pezold, J. Hrbacek, P.C. Cattin, With gaze tracking toward noninvasive eye cancer treatment, IEEE Transactions on Biomedical Engineering 63 (9) (2016) 1914–1924.
[169] J. Steil, P. Müller, Y. Sugano, A. Bulling, Forecasting user attention during everyday mobile interactions using device-integrated and wearable sensors, in: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, Association for Computing Machinery, New York, NY, USA, 2018.
[170] Z. Hu, C. Lv, P. Hang, C. Huang, Y. Xing, Data-driven estimation of driver attention using calibration-free eye gaze and scene features, IEEE Transactions on Industrial Electronics 69 (2) (2022) 1800–1808.
[171] J. O'Dwyer, R. Flynn, N. Murray, Continuous affect prediction using eye gaze and speech, in: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, Kansas City, MO, USA, 2017, pp. 2001–2007.
[172] Q. Tong, X. Hua, J. Qiu, K. Luo, L. Peng, P. Han, A new mapping function in table-mounted eye tracker, in: 2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/spectroscopy & Signal Processing Technology, volume 10620, International Society for Optics and Photonics, 2018, p. 106200B.

[173] K. Harezlak, P. Kasprowski, Application of eye tracking in medicine: A survey, research issues and challenges, Computerized Medical Imaging and Graphics 65 (2017) 176–190.
[174] M.J. Reale, S. Canavan, L. Yin, K. Hu, T. Hung, A multi-gesture interaction system using a 3-d iris disk model for gaze estimation and an active appearance model for 3-d hand pointing, IEEE Transactions on Multimedia 13 (3) (2011) 474–486.
[175] S. Tsutsumi, W. Tamashiro, M. Sato, M. Okajima, T. Ogura, K. Doi, Frequency analysis of gaze points with ct colonography interpretation using eye gaze tracking system, SPIE Medical Imaging, volume 10140, SPIE, Orlando, Florida, United States, 2017.
[176] G. Yuan, Y. Wang, H. Yan, X. Fu, Self-calibrated driver gaze estimation via gaze pattern learning, Knowledge-Based Systems 235 (2022) 107630.
[177] L. Yang, K. Dong, A.J. Dmitruk, J. Brighton, Y. Zhao, A dual-cameras-based driver gaze mapping system with an application on non-driving activities monitoring, IEEE Transactions on Intelligent Transportation Systems 21 (10) (2020) 4318–4327.

Jiahui Liu, received her Ph.D. degree in Control Science and Engineering from the University of Science and Technology Beijing (USTB), Beijing, China, in Jan. 2021. She is devoted to the research on the theoretical algorithms and the practical applications of the gaze tracking systems.

Jiannan Chi, is an Associate Professor in the Department of Instrument Science and Technology, USTB. He received his B.S. degree from Tianjin University, Tianjin, China, in 1990. He received the M.S. and Ph.D. degrees from Northeastern University, Boston, MA, USA, in 2002 and 2005, respectively. He was a visiting scholar with Nanyang Technological University, Singapore, in 2011. His current research interests include computer vision, human-computer interaction, and optical measurement.

Huijie Yang, received his B.S. degree in Measurement and Control Technology and Instruments from USTB, Beijing, China, in Jun, 2020, where he is currently pursuing the M.S. degree in Instrumentation Engineering. His main work is appearance-based gaze estimation.

Xucheng Yin, is a Full Professor and the Director of the Pattern Recognition and Information Retrieval Laboratory, Department of Computer Science and Technology, USTB. He received the Ph.D. degree in Pattern Recognition and Intelligent Systems from the Institute of Automation, Chinese Academy of Sciences, in 2006. He was a Visiting Professor with the College of Information and Computer Sciences, University of Massachusetts at Amherst, Amherst, MA, USA, for three times from January 2013 to January 2014, from July 2014 to August 2014, and from July 2016 to September 2016.