In the Eye of the Beholder: A Survey of Models for Eyes and Gaze

Dan Witzner Hansen, *Member*, *IEEE*, and Qiang Ji, *Senior Member*, *IEEE*

Abstract—Despite active research and significant progress in the last 30 years, eye detection and tracking remains challenging due to the individuality of eyes, occlusion, variability in scale, location, and light conditions. Data on eye location and details of eye movements have numerous applications and are essential in face detection, biometric identification, and particular human-computer interaction tasks. This paper reviews current progress and state of the art in video-based eye detection and tracking in order to identify promising techniques as well as issues to be further addressed. We present a detailed review of recent eye models and techniques for eye detection and tracking. We also survey methods for gaze estimation and compare them based on their geometric properties and reported accuracies. This review shows that, despite their apparent simplicity, the development of a general eye detection technique involves addressing many challenges, requires further theoretical developments, and is consequently of interest to many other domains problems in computer vision and beyond.

Index Terms—Eye, eye detection, eye tracking, gaze estimation, review paper, gaze tracking, object detection and tracking, human-computer interaction.

1 Introduction

As one of the most salient features of the human face, eyes and their movements play an important role in expressing a person's desires, needs, cognitive processes, emotional states, and interpersonal relations [141]. The importance of eye movements to the individual's perception of and attention to the visual world is implicitly acknowledged as it is the method through which we gather the information necessary to negotiate our way through and identify the properties of the visual world. Robust nonintrusive eye detection and tracking is, therefore, crucial for the development of human-computer interaction, attentive user interfaces, and understanding human affective states.

The unique geometric, photometric, and motion characteristics of the eyes also provide important visual cues for face detection, face recognition, and understanding facial expressions. For example, one of the primary stages in the Viola and Jones face detector is a Haar feature corresponding to the eye region [147]. This demonstrates the importance of the eyes for face detection. Additionally, the distance between the eyes is often utilized for face normalization, for the localization of other facial landmarks, as well as in filtering out structural noise. Gaze estimation and tracking are important for many applications including human attention analysis, human cognitive state analysis, gaze-based interactive user interfaces, gaze contingent graphical displays, and human factors. A gaze tracker is a device for analyzing eye movements. As the eye scans the environment or fixates on particular objects in the scene, a gaze tracker simultaneously localizes the eye position in the image and tracks its movement over time to determine the direction of gaze.

Research in eye detection and tracking focuses on two areas: eye localization in the image and gaze estimation. There are three aspects of eye detection. One is to detect the existence of eyes, another is to accurately interpret eye positions in the images, and finally, for video images, the detected eyes are tracked from frame to frame. The eye position is commonly measured using the pupil or iris center. The detected eyes in the images are used to estimate and track where a person is looking in 3D, or alternatively, determining the 3D line of sight. This process is called gaze estimation. In the subsequent discussion, we will use the terms "eye detection" and "gaze tracking" to differentiate them, where eye detection represents eye localization in the image while gaze tracking means estimating gaze paths.

This paper focuses on eye detection and gaze tracking in video-based eye trackers (a.k.a. video-oculography). A general overview of the components of eye and gaze trackers is shown in Fig. 1. Video-oculography systems obtain information from one or more cameras (*Image data*). The eye location in the image is detected and is either used directly in the application or subsequently tracked over frames. Based on the information obtained from the eye region and possibly head pose, the direction of gaze can be estimated. This information is then used by gaze-based applications, e.g., moving the cursor on the screen. The outline of this paper follows the components shown in Fig. 1 and is organized as follows: In Section 2, we categorize eye models and review eye detection techniques using the eye models. An eye model can be used to determine gaze and models for gaze estimation

- *D.W. Hansen is with the IT University of Copenhagen, Rued Langaardsvej 7, 2300 Copenhagen S., Denmark. E-mail: witzner@itu.dk.*
- *Q. Ji is with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, JEC 7004, Troy, NY 12180-3590. E-mail: jiq@rpi.edu.*

Manuscript received 16 Apr. 2008; revised 11 Nov. 2008; accepted 14 Jan. 2009; published online 23 Jan. 2009.
Recommended for acceptance by P.J. Phillips.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2008-04-0217.
Digital Object Identifier no. 10.1109/TPAMI.2009.30.

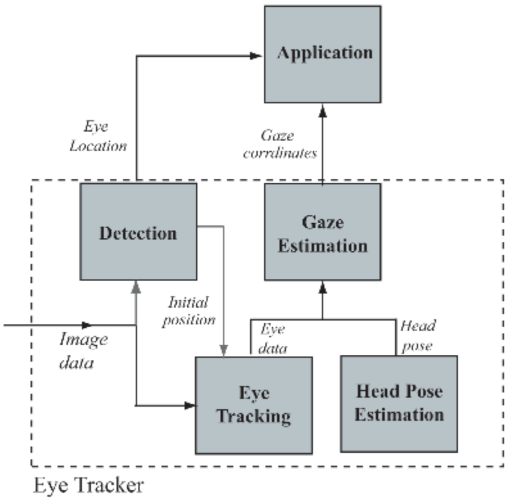

Fig. 1. Components of video-based eye detection and gaze tracking.

are reviewed in Section 3. Applications of eye tracking are versatile and a summary is presented in Section 4. We summarize and conclude the paper in Section 5 with additional perspectives on eye tracking.

2 Eye Models for Eye Detection

In eye detection, it is essential to identify a model of the eye which is sufficiently expressive to take account of large variability in the appearance and dynamics, while also sufficiently constrained to be computationally efficient. The appearance of eye regions shares commonalities across race, illumination, and viewing angle, but, as illustrated in Fig. 2, even for the same subject, a relatively small variation in viewing angles can cause significant changes in appearance. Despite active research, eye detection and tracking remains a very challenging task due to several unique issues, including occlusion of the eye by the eyelids, degree of openness of the eye, variability in either size, reflectivity or head pose, etc. Applications of computer vision, such as people tracking, face detection, and various medical applications, encounter occlusions and shape variations, but rarely of the same order of magnitude and frequency as seen with the eyes.

The eye image may be characterized by the intensity distribution of the pupil(s), iris, and cornea, as well as by their shapes. Ethnicity, viewing angle, head pose, color, texture, light conditions, the position of the iris within the eye socket, and the state of the eye (i.e., open/close) are issues that heavily influence the appearance of the eye. The intended application and available image data lead to different prior eye models. The prior model representation is often applied at different positions, orientations, and scales to reject false candidates.

Being either rigid or deformable, the taxonomy of eye detection techniques consists of *shape-based* [138], [47], [68], [166], [86], [167], [71], [36], [37], [36], [111], [44], [57], [76], [75], [78], [77], [138], [149], [120], [129] [130], *appearance-based* [117], [102], [61], [82], [148], [35], and *hybrid* methods [67], [46], [98], [158], [50], [169].

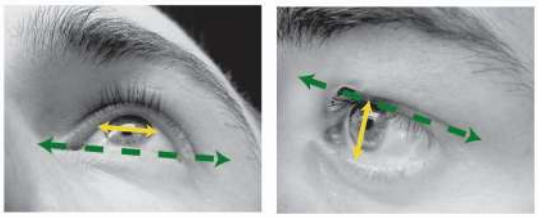

Fig. 2. The shape of the eye may change drastically when viewed from different angles. For example, the eyelids may appear straight from one view but highly curved from another. The iris contour also changes with viewing angle. The dashed lines indicate when the eyelids appear straight, while the solid yellow lines represent the major axis of the iris ellipse.

Shape-based methods can be subdivided into *fixed shape* and *deformable shape*. The methods are constructed from either the local point *features* of the eye and face region or from their *contours*. The pertinent features (Section 2.2) may be edges, eye corners, or points selected based on specific filter responses. The limbus and pupil are commonly used features. While the shape-based methods use a prior model of eye shape and surrounding structures (Section 2.1), the appearance-based methods rely on models built directly on the appearance of the eye region (Section 2.3). The appearance-based approach (the holistic approach) conceptually relates to template matching by constructing an image patch model and performing eye detection through model matching using a similarity measure. The appearance-based methods can be further divided into intensity- and subspace-based methods. The intensity-based methods use the intensity or filtered intensity image directly as a model, while the subspace methods assume that the important information of the eye image is defined in a lower dimensional subspace. Hybrid methods combine feature, shape, and appearance approaches to exploit their respective benefits (Section 2.4).

2.1 Shape-Based Approaches

The open eye is well described by its shape, which includes the iris and pupil contours and the exterior shape of the eye (eyelids). Categorization of shape-based approaches depends on whether the prior model is simple elliptical or of a more complex nature. Shape models usually constitute two components: a geometric eye model and a similarity measure. The parameters of the geometric model define the allowable template deformations and contain parameters for rigid (similarity) transformations and parameters for nonrigid template deformations. Deformable shape models often rely on a generic deformable template by which the eye is located by deforming the shape model through an energy minimization. An important property of these methods is their general ability to handle shape, scale, and rotation changes.

2.1.1 Simple Elliptical Shape Models

Many eye tracking applications (e.g., gaze estimation described in Section 3) only need the detection and tracking of either the iris or the pupil. Depending on the viewing angle, both the iris and pupil appear elliptical and consequently can be modeled by five shape parameters.

Simple ellipse models consist of *voting-based methods* [79], [84], [111], [118], [142], [165] and *model fitting methods* [24], [47], [89]. Voting methods select features that support a given hypothesis through a voting or accumulation process, while model fitting approaches fit selected features to the model (e.g., ellipse). Kim and Ramakrishna [79] and Perez et al. [118] use thresholds of image intensities to estimate the center of the pupil ellipse. Edge detection techniques are used to extract the limbus or the pupil boundaries. Several regions in the image may have a similar intensity profile to the iris and pupil regions and thresholds are therefore only applicable to constrained settings. The Hough transform can be used effectively to extract the iris or the pupil [111], [165], but requires explicit feature detection. Often a circularity shape constraint is employed for efficiency reasons, and consequently, the model only works on near-frontal faces. The computational demand may be reduced by observing the fact that the iris variability can be modeled with two degrees of freedom corresponding to pan and tilt [165].

Kothari and Mitchell [84] propose an alternative voting scheme that uses spatial and temporal information to detect the location of the eyes. They use the gradient field, knowing that the gradient along the iris boundary points outward from the center of the iris. Heuristic rules and a large temporal support are used to filter erroneous pupil candidates. A similar voting scheme is suggested by Valenti and Gevers [142]. Their method is based on isophote curvatures in the intensity image and uses edge orientation directly in the voting process. The approach relies on a prior face model and anthropomorphic averages to limit false positives. Since these models rely on maxima in feature space, they may mistake other features for eyes (e.g., eyebrows or eye corners) when the number of features in the eye region decreases. These methods are typically used when a constrained search region is available.

Daugman [24] proposes a different approach for pupil and iris detection. His technique uses optimization of the curve integral of gradient magnitudes under an elliptical shape model. This model does not take the contour neighborhood into account and may therefore disregard useful information. Witzner and Pece [47] also model the iris as an ellipse, but the ellipse is locally fitted to the image through an EM and RANSAC optimization scheme. They propose a likelihood model that incorporates neighboring information into the contour likelihood model and furthermore also avoids explicit feature detection (such as strongest gray-level gradient and thresholds). This method allows for multiple hypothesis tracking using a particle filter. The aim is to use the method in cases where thresholds are difficult to set robustly. Similarly Li et al. [89] also address low-cost eye tracking and propose the *Starburst* algorithm for detecting the iris through an elliptical shape model. The algorithm locates the strongest gray-level differences along rays and recursively sparkles new rays at previously found maxima. The maximum likelihood estimate of the pupil location is found through RANSAC. While framed differently, the Starburst algorithm is essentially an active shape model, like Cootes and Taylor [18], but allowing for several features to be used along each normal. Simple shape models are usually efficient and they can model features such as iris and pupil well under many viewing angles. However, the simple models are not capable of capturing the variations and intervariations of eye features such as eyelids, eye corners, and eyebrows. High contrast images and thresholds are often used for feature extraction.

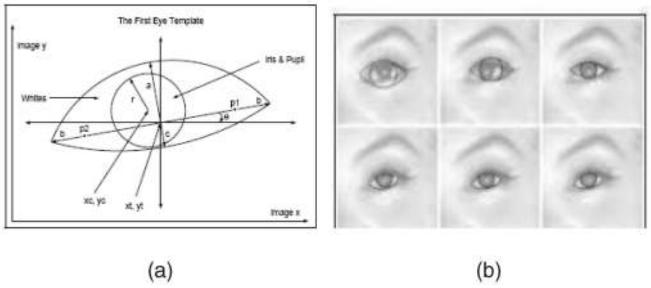

Fig. 3. (a) Yuille et al. model [166], (b) eye detection results (courtesy of Yuille and Hallinan).

2.1.2 Complex Shape Models

Complex shape-based methods allow, by definition, for more detailed modeling of the eye shape [166], [33], [18], [158], [86]. A prominent example is the deformable template model proposed by Yuille et al. [166]. The deformable eye model consists of two parabolas representing the eyelids (modeled with 11 parameters) and a circle for the iris as illustrated in Fig. 3. The model is fitted to the image through an update rule, which incorporates energy functions for valleys, edges, image peaks, and internal forces. Experimental research finds that the initial position of the template is critical. For instance, the algorithm fails to detect the eye when initializing the template above the eyebrow. Another problem lies in the complexity of describing the templates. In addition, the template-based approach may have difficulty with eye occlusions due to either eyelid closure or nonfrontal head pose.

The method proposed by Yuille et al. [166] can be sped up by exploiting the positions of the eye corners [86], [167], [71]. This requires the presence of four corners of each eye: the left and right corners of the eye as well as the corners formed by the iris and the upper eyelid. The four corners are present only if the iris is partially occluded by the upper eyelid. When the eyes are wide open or only occluded by the lower lid, the method fails as these corners do not exist. Using a face model, the eye corner locations are estimated using an eye corner template. The eye corner locations are used to initialize a deformable template that may be used for estimating eye shape. Similarly, Lam and Yan [86] extend Yuille's method for extracting eye features by using corner locations inside the eye windows as initialization points. They use a nonparametric "snake" method to determine the outline of the head. The approximate positions of the eyes are then found by anthropomorphic averages. The detected eye corners are used to reduce the number of iterations in the optimization of the deformable template.

Ivins and Porrill [68] describe a method for tracking the three-dimensional motion of the iris in a video sequence. A five-parameter scalable and deformable model is developed to relate translation, rotation, scaling due to the changes in eye-camera distance, and partial scaling due to the expansion and contraction of the pupil. The method requires high-quality and high-resolution images. Colombo and Del Bimbo [15] propose an eye model with six deformation parameters consisting of two semiellipses that share the same major axis. Coarse estimates of the left and right eye locations and shapes are initially calculated. The templates are then optimized similarly as in Yuille et al.'s method. Combining the elliptical models with complex eye models may speed up the localization and improve accuracy [13], [25].

Deformable template-based methods seem logical and are generally accurate and generic, but they suffer from several limitations. They are:

1. computationally demanding,
2. may require high contrast images, and
3. usually need to be initialized close to the eye for successful localization. For large head movements, they consequently need other methods to provide a good initialization.
4. Deformable contour models may face additionally problems when using IR light as the boundary of the sclera and the face may appear weak (see Section 2.5.1).
5. They may not be able to handle face pose changes and eye occlusions well.

While some deformable models, such as snake models, allow for much shape variability, other deformable models do not take into account the large variability of eye shapes. Further research is needed to produce models that can cope with large shape variations and even handle deformations such as eye closure or inconsistent feature presence (i.e., features appearing and disappearing with changes in scale).

2.2 Feature-Based Shape Methods

Feature-based methods explore the characteristics of the human eye to identify a set of distinctive features around the eyes. The limbus, pupil (dark/bright pupil images), and cornea reflections (see Section 2.5.1) are common features used for eye localization. Compared to the holistic approaches, feature-based methods aim to identify informative local features of the eye and face that are less sensitive to variations in illumination and viewpoint.

2.2.1 Local Features by Intensity

The eye region contains several boundaries that may be detected by gray-level differences. Herpers et al. [57] propose a method that detects local features, such as edges and lines, their orientation, lengths, and scale, and use a prior eye shape model to direct local contour following. The method initially locates a particular edge and then uses steerable Gabor filters to track the edge of the iris or the corners of the eyes. Based on the eye model and the features, a sequential search strategy is initiated in order to locate the eye position, shape, and corners.

Waite and Vincent [149] suggest a part-based model, where a part, such as eye corners or eyelid, is called a *microstructure*. They present a multilayer perception method to extract face features by locating eyes within the face image. Based on their work, Reinders et al. [120] propose several improvements by using multiple specialized neural networks. The trained neural network eye detector can detect rotated or scaled eyes and can work under various light conditions, although it is trained on frontal view face images only. A detailed eye model is used subsequently to refine the eye localization.

Bala et al. [4] propose a hybrid approach for eye classification by using an evolutionary algorithm to identify a subset of optimal features (mean intensities, Laplacian, and entropy) to characterize the eye. Feng and Yuen [36], [37] describe an eye model consisting of six landmarks (eye corner points). Initially, the eye landmarks are located and used to guide the localization of the iris and eye boundary. The methods assume the availability of an eye window in which the eye is the only object. The gray-scale face model used for estimating the eye window is described in [37]. The precise eye position is determined and verified by using the variance projection function [36]. Variance projection functions use the variance of intensities within a given eye region to estimate the position and the size of the iris or the positions of the eyelids. The variance projection function can be shown to be orientation and scale invariant. Experiments show that this method fails if the eye is closed or partially occluded by hair or face orientation. It is influenced by shadows and eye movements. In addition, this technique may mistake eyebrows for eyes.

Instead of detecting eye features, Kawato and Ohya [75], [76] detect the area between the two eyes. The between-eyes area has dark parts on its left and right (eyes and eyebrows) and comparably bright on the upper side (forehead) and lower side (nose bridge). The area is argued to be common for most people, viewable for a wide range of angles, and is believed to be more stable and easier to detect than the eyes themselves. They employ a circle-frequency filter to locate candidate points. The spurious points are subsequently eliminated from the candidates based on studying the intensity distribution pattern around the point. To prevent the eyebrows or other hair parts from being taken as eye-like regions, this method is made more robust by constructing a fixed "between-the-eyes" template to identify the true one from within the candidates [78], [77]. Experiments show that the algorithm may fail when hair covers the forehead or when the subject wears black rimmed glasses.

2.2.2 Local Feature by Filter Responses

Filter responses enhance particular characteristics in the image while suppressing others. A filter bank may therefore enhance desired features of the image and, if appropriately defined, deemphasize irrelevant features. The value of the pixels in the image after filtering is related to the similarity of the region to the filter. Regions in the image with particular characteristics can therefore be extracted through the similarity value. Sirohey et al. [129], [130] present methods for eye detection using linear and nonlinear filtering and face modeling. Edges of the eye's sclera are detected with four Gabor wavelets. A nonlinear filter is constructed to detect the left and right eye corner candidates. The eye corners are used to determine eye regions for further analysis. Postprocessing steps are employed to eliminate the spurious eye corner candidates. A voting method is used to locate the edge of the iris. Since the upper part of the iris may not be visible, the votes are accumulated by summing edge pixels in a U-shaped annular region whose radius approximates the radius of the iris. The annulus center receiving the most votes is selected as the iris center (c.f., Section 2.1.1). To detect the edge of the upper eyelid, all edge segments are examined in the eye region and fitted to a third-degree polynomial. Experiments show that the nonlinear filtering method obtains better detection rates than traditional edge-based linear filtering methods. High-quality images are essential for this method. D'Orazio et al. [26] convolve an image with a circular filter intended for gradient directions. The largest value of the convolution provides a candidate center of the iris circle in the image. Symmetry and distance heuristics are used to locate both eyes.

2.2.3 Pupil Detection

When the eye is viewed sufficiently closely, the pupil is a common and fairly reliable feature for eye detection. The pupil and iris may be darker than their surroundings and thresholds may be applied if the contrast is sufficiently large. Yang et al. and Stiefelhagen et al. [162], [132], [133] introduce an iterative threshold algorithm to locate the pupils by looking for two dark regions that satisfy certain anthropometric constraints using a skin-color model. Their method is limited by the results of the skin-color model and it will fail in the presence of other dark regions such as eyebrows and shadows. Even applying the same thresholds for both eyes seems likely to fail, especially considering different face orientations or different light conditions. Simple darkest pixel finding in search-windows centered around the last found eye positions is used for tracking. This scheme fails when there are other regions with similar intensity or during eye closure. Dark region detection may be more appropriate when using IR light than when using visible light (see Section 2.5.1).

The majority of the previously described methods are limited by not being able to model closed eyes. Tian et al. [138] propose a method to track the eye and recover the eye parameters through a dual state model (open/closed eyes) to overcome this limitation. The method requires manual initialization of the eye model. The eye's inner corner and eyelids are tracked using a modified Lucas-Kanade tracking algorithm [94]. The edge and intensity of the iris are used to extract the shape information of the eye using a Yuille et al.-like [166] deformable template. The method, however, requires high contrast images to detect and track eye corners and obtain a good edge image.

The feature-based methods generally report good robustness during illumination changes. For cameras with a wide field of view, eye candidates must be filtered, since several regions may be similar to the eyes. Pupil detection can be made more effective through techniques relying on properties reminiscent of red eye images in flash photography. More details on these methods are given in Section 2.5.1. These techniques work better indoors and even in the dark, but might be more difficult to apply outdoors because the pupils become smaller in bright environments and their intensities vary with illumination changes. Eye tracking and detection methods committed to using explicit feature detection (such as edges) may not be robust due to the change in light, image focus, and occlusion.

2.3 Appearance-Based Methods

While the shape of the eye is an important descriptor, so is its appearance. The appearance-based methods are also known as *image template* or *holistic methods.* The appearance-based methods detect and track eyes directly, based on the photometric appearance as characterized by the color distribution or filter responses of the eye and its surroundings. These methods are independent of the actual object of interest and are, in principle, capable of modeling other objects besides eyes. The term *appearance* may be understood as one or several images (*templates*) defined pointwise with appearance given by the changes of intensity or their filter responses. The appearance-based approaches are carried out either in the spatial or in a transformed domain. One of the main benefits of performing eye detection (object detection in general) in a transformed domain is to alleviate the effect of illumination variation by preserving subbands that are less sensitive to illumination and removing bands that are sensitive to illumination change. Such techniques, however, are in practice only tolerant to moderate illumination change.

Appearance-based methods can be image template-based, where both the spatial and tensity information of each pixel are preserved or holistic in approach, where the intensity distribution is characterized by ignoring the spatial information. Image template-based methods have inherent problems with scale and rotational changes. In addition, single-template models are limited by not modeling interperson variations. Even changes in head pose and eye movements within the same person can negatively influence them.

Holistic approaches use statistical techniques to analyze the intensity distribution of the entire object appearance and derive an efficient representation, defined in a latent space, to handle variations in appearance. Given a test image, the similarity between the stored prototypes and the test view is carried out in the latent space. The appearance-based methods usually need to collect a large amount of training data representing the eyes of different subjects, under different face orientations and different illumination conditions, but the underlying models are essentially independent of the object class. Through the model of pixel variations, a classifier or regression model can then be constructed.

2.3.1 Intensity Domain

Tracking and detecting eyes through template-based correlation maximization is simple and effective [42], [44]. Grauman et al. [42] use background subtraction and anthropomorphic constraints to initialize a correlation-based tracker. Hallinan [44] uses a model consisting of two regions with uniform intensity. One region corresponds to the dark iris region and the other to the white area of the sclera. His approach constructs an idealized eye and uses statistical measures to account for intensity variations in the eye templates. Huang et al. [59] and Zhu et al. [168] detect eyes using support vector machines. Polynomials of second-degree kernels yield the best generalization performance. The natural order in which facial features appear in frontal face images motivated Samaria and Young [124] to employ stochastic modeling, using hidden Markov models (HMMs) to holistically encode frontal facial information. The method assumes size and location normalized images of frontal faces. Only coarse-scale eye location is possible, and thus, further processing is needed to precisely locate the eyes.

Subspace methods may improve detection efficiency and accuracy of eyes using dimensionality reduction. The now standard Eigen analysis (PCA on image vectors) of image templates is capable of modeling variations in the training data such as eyes [58], [102] in a low-dimensional space. Pentland et al. [117] extend the eigenface technique to the description and coding of facial features each called eigeneyes, eigennoses, and eigenmouths. Eye detection is accomplished by projecting hypothetical image patches to the low-dimensional eigeneye-space. Huang and Mariani [102] employ eigeneyes for initial eyes localization. After obtaining the initial eye position, the precise location of the iris is determined by a circle with homogeneous dark intensity.

Image template methods inherently lack size invariance, so either a constant face size or multiscale grid solutions need to be employed. Since no direct model of the eye is present in the image, these methods lack direct access to specific eye parameters.

2.3.2 Filter Responses

The filter response methods for appearance models differ from those for feature-based methods by using the response values directly without making a selection of which features to use. Huang and Wechsler [61] present a method to represent eye images using wavelets in a Radial Basis function classifier. They treat the eye detection as binomial classification. Their experiments show improved performance of the wavelet RBF classifier compared to using intensity images. After eye region detection, they obtain precise eye location information such as the center and radius of the eyeballs by combining contour and region information.

The idealized eye features used by Hallinan [44] are essentially Haar features. The Viola and Jones face detector [147] learns the most discriminative Haar feature set for face detection through Adaboost. Similar approaches are found for eye detection [35], [50]. Witzner and Hansen [50] improve eye detection by combining information from glints (IR) and a Viola and Jones-like eye detector. Fasel et al. [35] use Gentleboost for separately training face and eye models. Using the same fundamental likelihood ratio detection model, they initially locate the faces at multiple scales and then the eyes. The main advantage of Haar features is their computational efficiency. Although the Haar features are easy to compute, their discriminating efficiency may be limited, especially in the final stages of the cascade. For complex patterns, the number of single weak classifiers may be high, where each only deals with a marginal number of negative cases.

The features and the selection procedure used in the Viola and Jones detector are simple and intuitive. However, the feature selection procedure uses brute-force search in a predefined feature pool and requires a significant time and memory consumption. In addition, Haar wavelet features are mainly applicable for detecting eyes on frontal faces. These limitations have inspired Wang et al. [152], [153] to propose *the recursive nonparametric discriminant feature* for face and eye detection using nonparametric discriminant analysis on image patches and Adaboost for training. The method overcomes the limitations of using Haar features. They report good detection and pupil localization results with a reduced number of discriminating features. The use of more complex features comes at the price of decreased runtime performance.

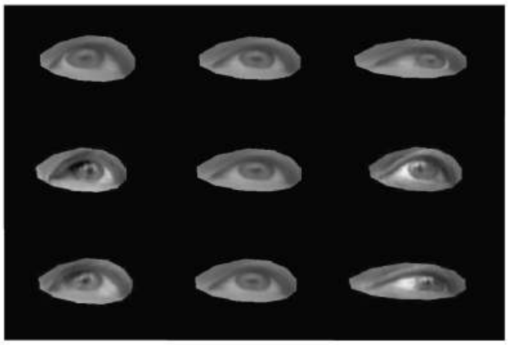

Fig. 4. Modes of variation: First mode of variation for shape (top), texture (middle), and combined models (bottom).

2.4 Hybrid Models

Hybrid methods aim at combining the advantages of different eye models within a single system to overcome their respective shortcomings.

2.4.1 Shape and Intensity

The combination of shape and appearance can, for example, be achieved through part-based methods. Part-based models attempt to build a general model by using a shape model for the location of particular image patches. In this way, a model of the individual part variances can be modeled explicitly while the appearance is modeled implicitly. Xie et al. [158], [159] suggest a part-based model employing a prior shape model consisting of several subcomponents. The eye region is initially detected through thresholding and binary search and is then divided into several parts: the whole eye region, two regions representing the sclera, the whole iris, and the occluded and unoccluded portions of the iris. The irises and the eyelids are modeled by circles and parabolas that have predetermined parameters and intensity distribution. Matsumoto and Zelinsky [98] use 2D image templates to represent facial features located on a 3D facial model. The iris is located by the circular Hough transform. The 2D image templates associated with the 3D model are used for matching purpose. The limitations of the part-based models are that they do not model the image intensities directly in the nonpatch areas and person specific models need to be built.

Other methods combine shape and appearance models more explicitly. Ishikawa et al. [67] and Witzner et al. [46] propose methods that combine shape and appearance models through an Active Appearance Model (AAM) [17]. In these models, both shape and appearance are combined into one generative model. The model can then be fitted to the image by changing the parameters according to a learned deformation model. Fig. 4 shows generated eyes along the first principal directions of the model [45].

For facial feature detection, a modified AAM model is suggested by Cristinacce and Cootes [22]. They use a local

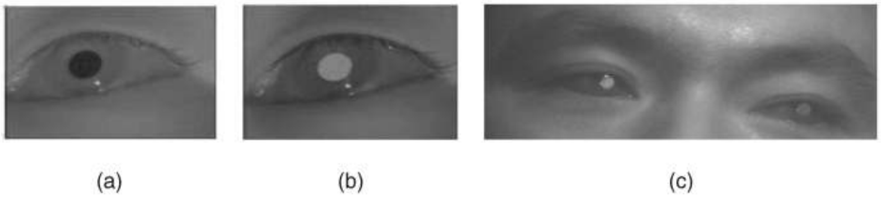

Fig. 5. (a) Dark and (b) bright pupil images. (c) Bright pupil images with varying intensity. Notice the small reflection (often referred to as the glint) on the cornea surface in the dark and bright pupil images.

appearance (patch) model for each landmark point and a global shape constraint for the spatial relationships. The active appearance models and their variants are able to model both shape and texture variations in a fairly low-dimensional subspace. In principle, they should be able to handle significant variations if trained on it. In practice, however, they are strongly influenced by sidelight. The standard active appearance model has relatively high computational demands, but Ishikawa et al. [67] report a modified AAM with very high efficiency. Like deformable models, active appearance models also need to be initialized close to the actual eye position in order to obtain a good fit. This means that these models must rely on another mechanism to handle large head movements. The models also face difficulties modeling the large eye appearance variability as the AAM methods are based on linear decompositions.

2.4.2 Colors and Shape

The color distribution at the eye region is reliably different from its surroundings. Despite this fact, color models of the eye have received very little attention.[1] Colors have mostly been employed for skin-color modeling [5], [38], [162], [132], [133], [76], [75], [78], [77], but there are also some attempts to model the color distribution of eye regions [46], [48]. Sole use of skin color may be prone to errors since skin colors can be similar to other textures in the scene such as certain wood types. Thus, prior eye location data are needed. Witzner et al. [46] use a color model for a mean-shift color tracker [16] for coarse-scale tracking and a gray-scale active appearance model for precise localization. A color-based active appearance model was attempted, but did not improve overall accuracy. The limitations of this approach are that the two models are separate and the active appearance model is dependent on the results from the color tracker.

2.5 Other Methods

A few methods, such as symmetry operators (Section 2.5.2), methods employing temporal information (Section 2.5.3), and active light (e.g., IR, described in Section 2.5.1), are not fully described by the previous model categories. Methods employing IR light are ubiquitous not particular to any eye model category.

2.5.1 Eye Detection under Active IR Illumination

Indoor video eye and gaze tracking systems utilize infrared (IR) light in practically all stages (detection, tracking, and gaze estimation) and its use dominates current eye tracker developments. Methods relying on visible light [89], [47], [166] are called *passive* light approaches; otherwise, the methods are called *active*. Most active light implementations use near IR light sources with wavelength around 780-880 nm. These wavelengths can be captured by many commercially available cameras and are invisible to the human eye and therefore do not distract the user or cause the pupil to contract. The amount of light emitted by current systems, whether IR light or visible light, is subject to international safety standards currently under development.

If a light source is located close to the optical axis of the camera (*on-axis* light), the captured image shows a bright pupil since most of the light reflects back to the camera. This effect is reminiscent to the red-eye effect when using a flash in photography. When a light source is located away from the optical axis of the camera (*off-axis*), the image shows a dark pupil. The use of *IR* illumination is shown in Fig. 5.

Several investigations have been made on the relationship between the intensity of the bright pupil and parameters such as head pose, gaze direction, and ethnic background [1], [101], [110]. Their studies show that bright pupil responses vary significantly between subjects and ethnic groups. Changes in head position or head pose affect the apparent brightness of the pupil. The brightest pupil responses occur when the eye is turned away from the light source.

Several objects in the background may generate patterns similar to the dark and bright pupil images, but the pupil effects rarely occur simultaneously for objects other than eyes. Eye models based on active remote IR illumination may therefore use the difference of dark and bright pupil images by switching between on and off-axis light sources, and often need to be synchronized with the camera through a video decoder [19], [31], [30], [32], [69], [107], [106], [53], [115], [49], [164]. The major advantages of the image difference methods are their robustness to global light changes, simplicity, and efficiency.

Manually defined thresholds are straightforward and fairly effective when differential lighting schemes are employed [31], [69], but should be made adaptive to the variations in pupil response. Ji and Yang [69] use the Kullback-Leibler Information distance for setting the threshold. Geometric and temporal criteria are used to filter blob candidates. Larger and fast head movements cause larger differences in the dark-bright pupil images. Methods to compensate for these effects have been suggested, using limited precision ultrasound to range the user's face, mirrors, and pan and tilt [134], [96]. To ensure that the eyes are within the view of the camera, other methods employ several cameras with pan and tilt [8], [116]. Tomono et al. [139], [34] propose systems in which three CCD cameras and two near IR light sources of different wavelengths are used. In

[1] This may be due to the common use of IR light.

addition, filters are used to control the captured information according to its polarization. Two cameras (one with a polarizing filter) are sensitive to only one wavelength while the third is sensitive to the second wavelength, effectively exploiting the dark and bright pupil images. Amir et al. propose a hardware solution to meet the requirements of fast eye pupil candidate detection [2]. Reflection of IR light sources on glasses is a generic and challenging research problem, which has only been partially solved, e.g., through pupil brightness stabilization techniques [30].

Many existing eye trackers are based on active light schemes. These systems are particularly efficient indoors and in dim environments, where ambient light is less of a complication. Most of these methods require distinct bright/dark pupil effects to work well. The success of such a system strongly depends on the brightness and size of the pupils. The brightness is affected by several factors, including eye closure, eye occlusion due to the face rotation, external illumination interferences, the distance of the subject to the camera, and the intrinsic properties of the eyes (i.e., the bright pupil reflection tends to be darker for older people). Furthermore, thick eye glasses tend to disturb the infrared light so much that the pupils appear very weak and often with many reflections. Conditions under which bright pupils are not necessarily reliable include eye closure and oblique face orientations, the presence of other bright objects (due to either eye glasses glare or motion), and external illumination interference. As discussed by Nguyen et al. [110] and shown in Fig. 5, even minor off-plane head rotation for the same subject may cause the bright pupil intensity to vary.

In order to overcome some of these challenges, Haro et al. [53] propose pupil tracking based on combining eye appearance, the bright pupil effect, and motion characteristics so that pupils can be distinguished from other equally bright objects in the scene. To do so, they verify the pupil blobs using conventional appearance-based matching methods and the motion characteristics of the eyes. Their method cannot track closed or occluded eyes or eyes with weak pupil intensity due to disadvantageous ambient light levels.

Zhu et al. propose a real-time, robust method for eye tracking under variable lighting conditions and face orientations [169]. The bright pupil effect and appearance of eyes (intensity distribution) are utilized simultaneously for eye detection and tracking. Support Vector Machines and mean-shift object tracking are employed for appearance-based pupil detection and tracking, which is combined with the bright pupil effect so that the pupil can be detected and tracked under variable head position and illumination. Witzner and Hammoud [49] propose a similar strategy by formulating a likelihood model to be used in a particle filter. They propose (either through mean-shift or directly) to weigh the contributions of the image patch before constructing the intensity distribution as to preserve some spatial location while maintaining flexibility to spatial variations.

Droege et al. [27] compare the accuracy of several dark pupil detection algorithms under relatively stable indoor conditions. Their study revealed only marginal performance differences. However, future work may show a larger performance variation under more challenging conditions.

A further discussion on the purpose of IR and cameras for gaze estimation is given in Section 3, where gaze estimation is discussed.

2.5.2 Symmetry Operators

Symmetry is an important cue for human perception [91], [92] and has been investigated for the purpose of automated eye and face detection [121], [90], [126], [85], [81], [39]. A well-known symmetry operator is Reisfeld and Yeshurun's *generalized symmetry transform*, which highlights regions of high contrast and local radial symmetry [121]. Their symmetry operator is based more on intuition than on formal grounds. It involves analyzing the gradient in a neighborhood for each point. Within this neighborhood, the gradients at pairs of points symmetrically arranged about the central pixel are used as evidence of radial symmetry and a contribution to the symmetry measure of the central point is computed. Rather than determining the contribution each pixel makes to the symmetry of pixels in its neighborhood, Loy and Zelinsky [93] propose the *Fast Radial Symmetry Transform* by considering the contribution of a local neighborhood to a central pixel. Their approach has a time complexity lower than those previously outlined. A study on the comparative complexity of symmetry operators was conducted by Loy and Zelinsky [93]. Gofman and Kiryati [40] introduced a global optimization approach similar to an evolutionary algorithm for the detection of local reflection symmetry using 2D Gabor decomposition. The use of symmetry operators for eye detection and tracking is limited by the need for thresholds to perform feature selection and a time complexity that scales with the size of the radius of the feature.

2.5.3 Blinks and Motion

Blinks are involuntary and periodic and usually simultaneous in both eyes. Blinking is necessary in order to keep the eyes moist, cool, and clean. These dynamic characteristics may be exploited for eye detection. Recently, eye motion and eye blinks have been used as a cue to detect eyes and faces [5], [23], [42], [77], [138]. Grauman et al. [42] locate eyes by assuming a fixed head position. Hypothetical eye positions are extracted based on the thresholded differences of successive frames. The most likely set of eye regions is chosen through anthropomorphic heuristics. Bala et al. [5] extract a face region based on a combination of background subtraction and skin-color information by analyzing luminance differences between successive images in the face region in order to extract eye blinking. On the successful localization of the eye regions, a dark circle-like region (pupil) is searched within each eye area. The center of the pupil is then taken as the center of the eye pattern and stored for the following matching process. A similar work was proposed by Crowley and Berard [23], where eye blink detection is based on luminance differences in successive images in small boundary areas of the eye.

Both of the above methods, however, assume static head, at least between two successive images where blinks occur. Kawato and Tetsutani [77] use eye blinks for initializing a between-the-eyes template. Their approach uses the differences between successive images, which distinguishes eyelid movements from head movement in order to detect blinks even while the head is moving. Blink detection may be achieved through relatively simple measures of the eye region (e.g., template correlation or the variation from the intensity mean). However, fast blinking and head movements make reliable blink detection challenging. Keeping track of eye characteristics during blinks may be necessary, thus, noneye features may be more reliable (e.g., using between-the-eyes templates). Furthermore, eye detection based on blinking is currently limited to detecting eyes in near-frontal faces. One possible solution to detecting blink during head movements is to track the motion of a few rigid feature points on the face, and subtract their motion from that of the eye motion to minimize the effect of the head movement.

2.6 Discussion

In this section, we summarize different techniques for eye detection and tracking. Based on their geometric and photometric properties, the techniques can be classified as shape-based, feature-based, appearance-based, and hybrid. Alternative techniques may exploit motion and symmetry. Active IR illumination may be employed by various techniques. Each technique has its advantages and limitations, but the optimal performance of any technique also implies that its particular optimal conditions with regard to image quality are met. These conditions relate to illumination, head pose, ethnicity, and degree of eye occlusion. For example, the techniques based on active IR illumination work well indoors, while techniques based on shape and appearances can work reasonably well both indoors and outdoors. The existing methods are to a large extent only applicable to near frontal view angles, fully open eyes, and under relatively constrained light conditions. In addition, the eye appearance may change significantly with changes in scale. Features defined on one scale do not exist or have changed dramatically in another scale. It is therefore challenging to apply a single-scale eye model to multiple scales. It would therefore be instructive to determine distributions of features and feature responses for the class of eyes, as in natural image statistical approaches, so as to be better able to control for changes in eye appearance over scale. It remains a challenge to detect and track the eyes due to wide, complex variations in the eye image properties due to ethnicity, illumination conditions, scale, head pose, and eye state (open/closing eyes). Recently, patch-based methods for object detection, recognition, and categorization have received significant attention as they show promising results. As a feature-based method, they tend to be more discriminative, more robust to face pose, and illumination variant than the holistic eye detection approaches.

Fig. 6 summarizes and qualitatively compares various eye detection methods presented in this section. The table categorizes techniques and summarizes their relative performance under various image conditions. The intention is that readers can determine suitable techniques for their particular applications.

Before concluding this section, we also want to discuss a few related issues: 1) In order to develop effective eye detection techniques, the training and testing of eye data are essential. Various eye and face databases such as BioID and Yale [123] can be used to validate eye detection techniques but others are also available [14]. 2) The eye image requirements differ among the methods discussed in this section. While hardware choice plays an important role, we have so far avoided placing too much emphasis on these issues, but rather chosen to describe hardware-independent eye detection techniques. Some applications use fairly high-quality cameras with variable lenses and sometimes with pan and tilt heads in order to perform accurate and robust eye detection under conditions of large head movements. These applications incur high cost. On the other hand, some applications aim to use low-quality consumer cameras to minimize cost to the consumer. Low-cost solutions with a standard lens may require the camera to be close to the eye and the head to be relatively stationary. 3) The detection techniques discussed in this section are specifically developed for eye detection, but some techniques can easily be extended to detect and track other objects. While the simple ellipse-based methods can be used to detect and track any circular or elliptical object, the complex deformable shape models can be used to detect and track complex objects like the hand and human organs in medical imaging. The appearance-based methods of both intensity and subspace domains have been widely applied to face detection, animal detection (e.g., horses), and vehicles. The local feature-based methods have been applied to the detection and tracking of other facial features including mouth corners and nose corners.

3 Gaze Estimation

The primary task of gaze trackers is to determine gaze. Gaze should in this context be understood as either the *gaze direction* or the *point of regard (PoR)*.[2] Gaze modeling consequently focuses on the relations between the image data and the point of regard/gaze direction.

Basic categorizations of eye movements include saccades and fixations. A fixation occurs when the gaze rests for some minimum amount of time on a small predefined area, usually within 2-5 degrees of central vision, usually for at least 80-100 ms. Saccades are fast, jump-like rotations of the eye between two fixated areas, bringing objects of interest into the central few degrees of the visual field. Smooth pursuit movements are a further categorization that describe the eye following a moving object. Saccadic eye movements have been extensively investigated for a wide range of applications including the detection of fatigue/drowsiness, human vision studies, diagnosing neurological disorders, and sleep studies [28]. Fixations are often analyzed in vision science, neuroscience, and psychological studies to determine a person's focus and level of attention. Properties of saccades and fixations may provide diagnostic data for the identification of neurological, vision or sleep disorders. Eye positions are restricted to a subset of anatomically possible positions described in Listing's and Donder's laws [140]. According to Donder's law, gaze direction determines the eye orientation uniquely and the orientation is furthermore independent of the previous

[2] Keep in mind that gaze information does not necessarily mean that the person is in an attentive state nor that the estimated point of regard is coincident with the monitor.

Method Type	Info (P,I,C,E,BE)	Light (I,O,IR)	Invariance (H,S,O)	Requirements (H,C,T,G)	References
Circular Shape	P	IR	S, H°	H,C	[79], [89], [118], [111]
Ellipse Shape	I, P	I, O, (IR)	H, S	C	[24], [47], [84], [142], [165]
Ellipse Shape	P	IR	H, S	C, T	[19], [30], [31], [32], [49], [53] [69], [107], [106], [115], [164]
Complex Shape	P, I, C	I, O	$H^{\bullet}$,S	H,G	[13], [15], [25], [68], [71] [86], [167], [166]
Feature	I, C	I	—	—	[36], [37], [57], [120], [149]
Feature	I	I	—	C	[5], [26]
Feature	I,C	I	—	C	[129], [130]
Feature	E	I	—	C	[138], [132], [133]
Feature	BE	I	$S^{\bullet}$, $H^{\bullet}$, O		[74]-[77]
Feature	P	I	—	C	[162], [132], [133], [38], [42], [44], [58] [59], [102], [124], [168]
Appearance	E	I,O	$H^{\bullet}$,$S^{\bullet}$	—	[22], [35], [60], [49]
Symmetry	I,P	S	H,C		[121], [90], [126], [85], [81], [39].
Motion	E	I,O	—	T	[5], [23], [42], [77], [138]
Hybrid	P,I,C,E,	I,O,IR	H,S	G	[22], [46], [67], [98], [158], [159]

Fig. 6. Eye Detection models: The "Method Type" column corresponds to the method category. The "Info" column refers to the information that can be obtained directly from the model: Pupil (P), Iris (I), Corners (C), Entire Eye (E), Between-the-Eyes (BE). The "Light" column indicates under which light conditions the method operates: Indoor(I), Outdoor (O), or under IR light (IR). The "Invariance" column considers the robustness to scale (S), head pose (H) changes, and to occlusion (O) due to the eye blinks or closed eyes. The "requirements" column includes high-resolution (H) eye images, High contrast (C), Temporal (T) dependencies, Good Initialization (G). Superscript $^{\bullet}$ indicates robustness to some degree and $^{\circ}$ indicates robustness to a minor degree. Methods may consequently possess properties not reported here (e.g., capable of using IR in both indoor and outdoor conditions). Values given in parentheses are optional.

positions of the eye. Listing's law describes the valid subset of eye positions as those which can be reached from the so-called primary position through a single rotation about an axis perpendicular to the gaze direction.

When light falls on the curved cornea of the eye (see Fig. 7), some of it is reflected back in a narrow ray pointing directly toward the light source. Several reflections occur on the boundary between the lens and the cornea, producing the so-called *Purkinje images* [28]. The first Purkinje image or *corneal reflection* is often referred to as the glint.

To exemplify the importance of light sources for gaze estimation, consider looking directly at a light source. The distance between the glint and the center of the pupil is small. However, looking away increases this distance. This implies that if the sole purpose of the gaze tracker is to determine whether a person is looking at a specific light source, all that is needed is to make a simple classification (threshold) on the length of the pupil-glint vector. This also illustrates that high-accuracy gaze estimation may not be necessary for all applications.

People move their heads when using a gaze tracker. A person's gaze is determined by the head pose (position and orientation) and eyeball orientation. A person can change gaze direction by rotating the eyeball (and consequently also the pupil) while keeping the head stationary. Similarly, a person can change gaze direction by moving the head while keeping the eye stationary relative to the head. Usually, a person moves the head to a comfortable position before orienting the eye. Head pose, therefore, determines the coarse-scale gaze direction while the eyeball orientation

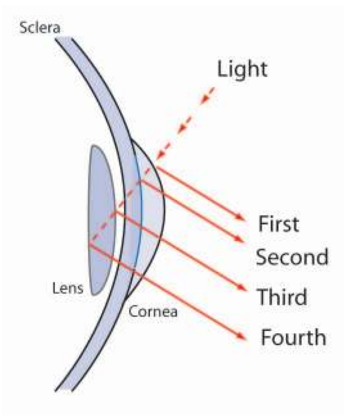

Fig. 7. Light is reflected on the eye and results in various Purkinje images (first, second, etc.).

determines the local and detailed gaze direction. Gaze estimation therefore needs to (either directly or implicitly) model both head pose and pupil/iris position. The problem of ensuring head pose invariance in gaze trackers is important and constitutes a challenging research topic. Head pose invariance may be obtained through various hardware configurations and prior knowledge of the geometry and cameras. Information on head pose is rarely used directly in the gaze models. It is more common to incorporate it implicitly either through the mapping function (regression-based method described in Section 3.1) or through the use of reflections on the cornea (3D model-based approaches described in Section 3.2).

All gaze estimation methods need to determine a set of parameters through calibration. We clarify the calibration procedures into the following:

1. *camera-calibration*: determining intrinsic camera parameters;
2. *geometric-calibration*: determining relative locations and orientations of different units in the setup such as camera, light sources, and monitor;
3. *personal calibration*: estimating cornea curvature, angular offset between visual and optical axes; and
4. *gazing mapping calibration*: determining parameters of the eye-gaze mapping functions.

Some parameters may be estimated for each session by letting the user look at a set of predefined points on the monitor; others need only be calculated once (e.g., human specific parameters) and yet other parameters are estimated prior to use (e.g., camera parameters, geometric and physical parameters such as angles and location between camera and monitor). A system where the camera parameters and geometry are known is termed *fully calibrated*. This classification will be used to differentiate the assumptions made in various methods.

Desirable attributes in a gaze tracker include minimal intrusiveness and obstruction, allowing for free head movements while maintaining high accuracy, easy and flexible setup, and low cost. A more detailed description of eye tracker preferences is given by Scott and Findlay [125]. Only a few years ago, the standard eye tracker was intrusive, requiring, for example, a reflective white dot placed directly onto the eye or attaching a number of electrodes around the eye [63]. Use of headrests, bite-bars, or making the eye tracker head mounted were common approaches to accommodate significant head movements. Head movements are typically tracked using either a magnetic head tracker, another camera, or additional illuminators. Head and eye information is fused to produce gaze estimates [122], [151].

Compared to the early systems, video-based gaze trackers have now evolved to the point where the user is allowed much more freedom of head movements while maintaining good accuracy (0.5 degree or better). As reviewed in this section, recent studies show that using specific reflections from the cornea allows gaze trackers to be easily and cheaply produced, and enhances stable and head pose invariant gaze estimation. However, commercial eye trackers remain regrettably expensive.

Current gaze estimation methods are mostly feature-based (described in the subsequent section), but we will later review others such as appearance-based methods.

3.1 Feature-Based Gaze Estimation

Gaze estimation methods using extracted local features such as contours, eye corners, and reflections from the eye image are called *Feature-based* methods. The primary reasons for using feature-based methods are that the pupil and glints (under active light models) are relatively easy to find and that these features can, as indicated above, be formally related to gaze. This encompasses aspects related to the geometry of the system as well as to eye physiology. For these reasons, they have become the most popular approach for gaze estimation.

Two types of feature-based approaches exist: the *model-based* (geometric) and the *interpolation-based* (regression-based). The interpolation-based methods [11], [31], [46], [47], [69], [104], [157] that assume the mapping from image features to gaze coordinates (2D or 3D) have a particular parametric form such as a polynomial [104], [131] or a nonparametric form such as in neural networks [69], [45]. These methods avoid explicitly calculating the intersection between the gaze direction and gazed object. The *3D model-based methods*, on the other hand, directly compute the gaze direction from the eye features based on a geometric model of the eye. The point of gaze is estimated by intersecting the gaze direction with the object being viewed [113], [144], [151], [100], [8].

In the following sections, we first describe regression-based methods (Section 3.1), and follow with a review of the 3D model-based approaches in Section 3.2, which are further subdivided based on their hardware requirements. A discussion of gaze estimation methods and a table summarizing the models are given in Section 3.3.

3.1.1 2D Regression-Based Gaze Estimation

Early gaze tracking systems employed a single IR light source to improve contrast and obtain stable gaze estimation results. The erroneous assumption implicitly made by many single-glint methods is that the corneal surface is a perfect mirror, so, if the head is kept fixed even when the cornea surface is rotated, the glint remains stationary. The glint is therefore considered the origin of a glint-centered coordinate system. The difference between the glint and the pupil center is in this view used to estimate gaze direction. A mapping from the pupil-glint difference vector to the screen is often conducted.

As early as 1974, Merchant et al. [99] proposed a real-time video-based eye tracker employing IR light (dark-bright pupil images) using a single camera. A collection of mirrors and galvanometers allows for head movements. They use the pupil-glint vector and a linear mapping to estimate the point of regard (POS) and notice nonlinearities with large pupil-glint angles. They compensate for these using polynomial regression. Similarly, and much later, Morimoto et al. [103] also used a single camera and utilized one second order polynomial for the x and y directions separately to represent a direct mapping of the glint-pupil difference vector to the point of regard. Unfortunately, the calibration mapping decays as the head moves away from its original position [104]. A similar approach, but without using glint information, is described by Stampe [131]. He additionally proposes polynomial functions to model the correlation between pupil centers.

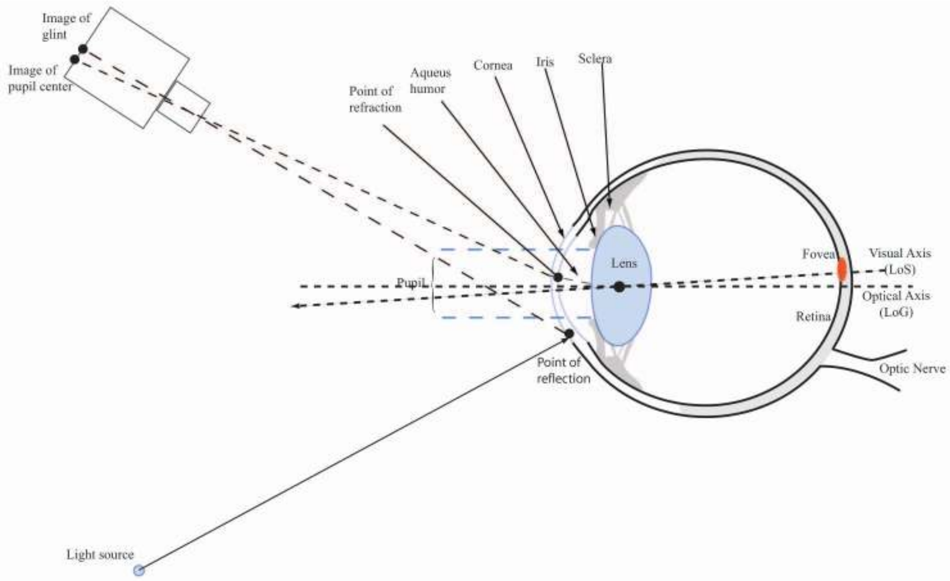

Fig. 8. General model of the structures of the human eye, light, light sources, and projections.

White et al. [156] assume a flat cornea surface and propose a polynomial regression method for PoR estimation in a similar way as Morimoto et al. and Merchant et al. [103], [99]. They additionally propose using a first order linear regression to account for gaze imprecision resulting from lateral head movements. During calibration, a set of four calibration mappings for different head locations is estimated by exploiting spatial symmetry. Head positions are accurately located by creating a second glint using another IR light source. Using two light sources as points of reference and exploiting spatial symmetries, a single static calibration can be adjusted as the head moves. They mention that, in practice, higher order polynomial functions do not provide better calibration and argue that gaze estimation can be done independently of eye rotation and head translation—a fact that was later generalized and proved to be true [43], [128].

Neural networks and their variants are popular tools for regression tasks. Ji and Zhu [70] suggest a generalized regression neural network-based method in which the pupil parameters, pupil-glint displacement, orientation and ratio of the major and minor axes of the pupil ellipse, and glint coordinates are used to map the screen coordinates. The intention and advantage of the method is that no calibration is necessary after initial training. This method only improves head movement moderately. It is reported that the method handles head movements while still producing accuracies of about 5 degree. Zhu et al. [171] suggest the use of Support Vector Machines to describe the mapping from the pupil and single glint to screen coordinates.

Most gaze estimation methods do not offer a way of knowing when the current inputs are no longer compatible with the calibration data. Witzner et al. [45], [46] use Gaussian process interpolation to exploit the covariance of the training data and new inputs as an indicator of when gaze predictions deviate from the inputs of the calibration data (e.g., a head movement) and to make predictions.

Two-dimensional interpolation methods do not handle head pose changes well. Helmets may be of some help, but, contrary to the intentions behind mounting the eye trackers on the head, they may still move after calibration and thus influence accuracy. Kolakowski and Pelz propose a set of heuristic rules for adjusting minor slippage of head mounts [83].

Using a single camera, the 2D regression methods model the optical properties, geometry, and the eye physiology indirectly and may therefore be considered as approximate models, which may not strictly guarantee head pose invariance. They are, however, simple to implement, do not require camera or geometric calibration, and may still provide good results under conditions of small head movements. More recent 2D regression-based methods attempt to improve performance under larger head movements through compensation, or by adding additional cameras [171], [170]. Zhu and Ji introduce a 2D regression-based method [170] using two cameras to estimate 3D head position. They use the 3D eye position to modify the regression function to compensate for head movements. However, contrary to other regression methods, the method of Zhu and Ji [170] needs a prior stereo calibration of the cameras.

3.1.2 3D Model-Based Gaze Estimation

Three-dimensinoal model-based approaches model the common physical structures of the human eye geometrically so as to calculate a 3D gaze direction vector. By defining the gaze direction vector and integrating it with information about the objects in the scene, the *point of regard* is computed as the intersection of the gaze direction vector with the nearest object of the scene (e.g., the monitor).

Fig. 8 shows the structures of the eye used in gaze tracking. The eyeball is approximately spherical, with a radius of about 12-13 mm. The parts of the eye that are visible from the outside are the *pupil*, the *iris* (colored part), and the *sclera* (the white part of the eye). The boundary between the iris and the sclera is called the *limbus*. The pupil is the *aperture* located in the center of the iris and it regulates the amount of light entering the eye by continuously expanding and contracting. The *cornea* is a protective transparent membrane on the surface of the eye in front of the iris with a radius of about 7.8 mm [43]. Behind the iris is the biconvex multilayered structured *lens*. The shape of the lens changes so as to focus objects at various distances on the retina, which is a layer coating the back of the eye containing photosensitive cells. The *fovea* is a small region in the center of the retina, in line with the central 5 or so degrees of vision. The fovea contains the majority of color sensitive cells, and these cells are more tightly packed and more differentially connected to the optic nerve than cells in peripheral areas of the retina. The fovea is responsible for the perception of fine details. *Gaze direction* is either modeled as the optical axis or as the visual axis. The optical axis (a.k.a. line of gaze (LoG)) is the line connecting the pupil center, cornea center, and the eyeball center. The line connecting the fovea and the center of the cornea is known as visual axis (a.k.a. the line of sight (LoS)). The line of sight is assumed to be the true direction of gaze. The visual and optical axes intersect at the cornea center (a.k.a. nodal point of the eye) with subject-dependent angular offsets. In a typical adult, the fovea is located about 4-5 degrees horizontally and about 1.5 degrees below the point of the optic axis and the retina and may vary up to 3 degrees between subjects [12], [43]. Knowledge of the 3D location of the eyeball center or the corneal center is a direct indicator for the head location in 3D space and may obviate explicit head location models. The estimation of these points is therefore the cornerstone of most head pose invariant models.

The parameters used for geometric modeling of the eye can be divided into *extrinsic*, *fixed eye intrinsic,* and *variable* categories [8]. The *extrinsic parameters* model 3D eye position (center of the eyeball) and optical axis. The *fixed eye intrinsic* parameters include cornea radii, angles between visual and optical axes, refraction parameters (e.g., in aqueous humor), iris radius, and the distance between the pupil center and cornea center. They remain fixed during a tracking session, but may change slowly over the years. Parameters such as the visual axis, refraction indexes,[3] the distance between the cornea center and pupil center, and the angles of the visual axis and optical axis are subject-specific, some of which are difficult to measure directly. The *variable parameters* change the shape of the eye model and include the pupil radius.

Most 3D model-based (or geometric) approaches [151], [144], [145], [156], [43], [105], [100], [112], [109], [128], [127] rely on metric information and thus require camera calibration and a global geometric model (external to the eye) of light sources, camera and monitor position, and orientation. Exceptions to this are methods that use projective invariants [164], [20] or simplifying assumptions [47]. It is out of the scope of this paper to provide mathematical details of these methods, but most follow the same fundamental principles. Euclidean relations such as angles and lengths can be employed through calibrated cameras. The general approach is to estimate the center of the cornea, and thus, the optical axis in 3D. Points on the visual axis are not directly measurable from the image. By showing at least a single point on the screen, the offset to the visual can be estimated. The intersection of the screen (known in fully calibrated setups) and the visual axis yield the point of regard.

Model-based approaches typically estimate gaze direction by assuming spherical eyeball and cornea surfaces. Only a few methods model these structures as ellipsoid [8]. The spherical models of the cornea may not be suitable for modeling the boundary area of the cornea and often lead to greater inaccuracies when the user moves the eye to the extremities of the display (e.g., glints move on a non-spherical surface). Both the optical and visual axes intersect at the cornea center making the cornea center an important parameter to estimate in a geometric approach.

With a known cornea curvature, it is possible to find the cornea center using one camera and two light sources. However, estimation of the cornea center requires at least two light sources and two cameras when the eye-specific parameters are unknown [128]. Instances of applying these results in a fully calibrated setup have been proposed by several authors [105], [127], [43]. Estimation of the angles between the optical and visual axes is also needed to find the direction of gaze, requiring at least a single point of calibration [127]. For simplicity, anthropomorphic averages for the cornea curvature are frequently used [112], [51].

The following sections describe methods and formal relations of point features such as center of pupil with varying geometry and cameras. Starting with one light source and a single camera, each subsequent section reviews models with increasing number of light sources and cameras. Calculating the importance of refraction requires additional work and quantification and will only be quantified when applicable.

B1: Single camera and single light. Due to their simplicity and effectiveness, single camera and single glint approaches were quite common in early gaze tracker implementations. Regression-based methods (Section 3.1) mostly employ a single camera and a single light source; but for the remainder of this section, attention is only given to the geometric approach, i.e., the 3D model-based approach. A few geometric methods use a single light source [115], [43]. Ohno et al. [115] describe a model-based approach using a single calibrated camera and single glint. They use population averages for the cornea curvature, distance between pupil center and the center of cornea, as well as a constant refraction index (1.336) to estimate the optical axis. Later, Ohno argues that personal calibration can be reduced to two fixation points using two light sources and a single camera [114].

Shih et al. [128] prove that the use of a single glint and the pupil center cannot lead to head pose invariant gaze estimation. Their results explain the need for additional constraints such as additional cameras to compensate for

[3] Refraction occurs when light passes from one transparent medium to another; it changes speed and bends. The degree of bending depends on the refractive index of the mediums and the angle between the light ray and the normal to the surface separating the two mediums. The consequences of refraction are a nonlinear displacement of the observed location of features such as the pupil and may, thus, influence the estimation of the 3D location.

head pose changes in single glint systems, using population averages or a spherical eyeball model to obtain the optical axis [115].

Guestrin and Eizenman [43] generalize these results for calibrated systems, showing that the gaze direction can be determined by using only a single glint given the distance between the eye and the monitor or keeping the head fixed. Much of the theory behind geometric models using fully calibrated setups has been formalized by Guestrin and Eizenman [43]. Their model covers a variable number of light sources and cameras, human specific parameters, light source positions, and camera parameters. The model, however, is limited by requiring euclidean information (e.g., fully calibrated setups).

The majority of the results of Guestrin and Eizenman [43] regard point sources such as center of the pupil, center of glint etc. Villanueva and Cabeza [146] show that, in fully calibrated setups, the ellipse information of the pupil (disregarding refraction) can be used to reduce the number of light sources to one and still provide head pose invariance. Consequently, there may be several unrevealed results using pupil or iris contours.

A common misconception of single glint systems is that the pupil-glint difference vector remains constant when the eye or the head moves. The glint will clearly change location when the head moves, but it is less obvious that the glint shifts position when changing gaze direction. The eye rotates around the eyeball center and not around the cornea center. This means that a change of gaze direction also moves the cornea in space, and thus, the glint will not remain fixed. Second, minor changes of glint position may also be due to a nonspherical cornea. The influence of small head movements on the difference vector is indeed minimal. The technique is used somewhat successfully in gaze trackers, where the camera is fixed relative to the eye to compensate for small amounts of slippage. However, larger head movements cause significant changes in the difference vector.

B2: Single camera and multiple lights. Adding light sources to the setup is a small step from the previous methods, but, as it turns out, a giant leap for obtaining head pose invariance. Shih et al. [128] and Guestrin and Eizenman [43] show that the cornea center and, in turn, also gaze direction can be estimated in fully calibrated settings using two or more light sources and known cornea curvature. Guestrin and Eizenman's system[4] allows for only small head movements, but it appears that their well-founded approach would allow for greater head movements with a higher resolution camera [43]. They also make simplifying assumptions on refraction in the aqueous humor. Several authors follow this approach with minor adjustments to the model [100], [144], [127], [55]. In fact, only one calibration point is needed to estimate the cornea curvature, cornea center, and visual axis (single angle) when using two light sources [146]. These methods are usually accurate, but ongoing collection and maintenance of geometric and camera calibrations usually entails errors.

The gaze tracking systems relying on this approach are consequently inflexible when attempting to change the geometry of light sources, camera (e.g., zoom), and screen to particular needs. They may also result in heavy systems. This type of approach seems to be the foundation of several commercial systems.

Contrary to the previous methods, Yoo and Chung [164] describe a method which is capable of determining the point of regard based solely on the availability of light source positions information (e.g., no camera calibration) by exploiting the cross-ratio of four points (light sources) in projective space. Yoo and Chung [164] use two cameras and four IR light sources placed around the screen to project these corners on the corneal surface, but only one camera is needed for gaze estimation. When looking at the screen, the pupil center should ideally be within the four glint areas. A fifth IR light emitter is placed on axis to produce bright pupil images and to be able to account for nonlinear displacements of the glints. In order to account for the nonlinear displacements of the projected glints on the cornea, they learn four α_i parameters, initially asking the user to look at the light sources. Coutinho and Morimoto [20] extend the model of Yoo and Chung [164] by using the LoS-LoG offset as an argument for learning a constant on-screen offset. Based on this, they argue that a simpler model can be made by learning a single α value rather than four different values as originally proposed. They show significant accuracy improvements compared to the original paper, provided that the user does not change their distance to the camera and monitor. The method is not robust to depth-scale changes since a constant LoS-LoG offset does not yield a constant offset on the screen when changing the distance of the eye to the screen. The model of cross-ratios is also an approximation since the pupil is located on a different plane from that determined by the (corrected) corneal reflections. The advantage of the method is that it does not require a calibrated camera. It only requires light source position data relative to the screen. One limitation is that the light sources should be placed right on the corners of the screen—a task which is not entirely trivial. In practice, the method is highly sensitive to the individual eye, and formal analysis of the method is presented by Kang et al. [72].

The intersection of gaze direction vectors from two eyes provides information about the 3D point of gaze and has recently motivated researchers to propose methods reminiscent of stereo vision for 3D PoR estimation [29],[56]. Methods for 3D PoR seem to obtain fairly reliable results, but are still in an early stage of development.

In general, multiple light sources are faced with increased chance that one of the glints might disappear. There may therefore need to be physical restraints on the actual head locations in order to ensure that all glints appear in the image.

B3: Multiple cameras and multiple lights. Fixed single camera systems are faced with the dilemma of trading head movements against high-resolution eye images. A large field of view is required to allow for free head motion, but a limited field of view is needed to capture sufficiently high-resolution eye images to provide reliable gaze estimates. Multiple cameras are utilized to achieve these goals either through wide-angle lens cameras or a movable narrow-angle lens cameras. Multiple cameras also allow for 3D eye modeling. The first remote eye tracking systems appearing in the literature that use multiple cameras, either have separate cameras for each eye or use one camera for head

[4] A particular instance of their model with a single camera and two light sources.

location tracking to compensate for head pose changes and another camera for close-up images of the eye [70], [171], [151]. Whenever the eye moves outside the range of the narrow field of view camera, some systems mechanically reorient the narrow field of view camera toward the new eye position using a pan and tilt head [135], [113]. Acquisition time of pan and tilt cameras can be improved by replacing them with mirrors [112]. Only recently have the geometric constraints known from stereo been used effectively [8], [11], [113], [128].

1. *Head pose compensation using multiple cameras.* Regression-based gaze estimation methods are sensitive to head pose changes. A direct solution to compensate for minor head movement is to use one camera for observing the head orientation and another camera for eye images and then combine the information [70], [73], [171]. The methods are more complex due to the need for additional geometric calibrations and it is not obvious how to fuse observed head orientations and regression parameters into gaze coordinates. Applying multiple cameras in this way does not use the available stereo constraints effectively, since eye information is only coarsely defined in one of the cameras. In the following section, we describe methods where multiple cameras are used in a more common stereo setup.
2. *Stereo and Active Cameras:* Stereo makes 3D eye modeling directly applicable [8], [11], [80], [113], [127]. In fact, it can be shown that information of the optical axis can be estimated in fully calibrated stereo systems without any session calibration [128] and only one calibration point is needed when also modeling the visual axis (one angle) [127]. A recent implementation of this is suggested by Zhu and Ji [170]. Differently from Shih's method, their method can estimate gaze when the optical axis of the eye intersects or is close to the line connecting the nodal points of the two cameras.

Tomono et al. [139] discuss a setup consisting of three cameras and two light sources and mirrors. Even though stereo is used, they employ a simplified face model (rather than modeling the center of the eye) together with an eye model to estimate LoS.

Beymer and Flickner present an elaborate system modeling the eye in 3D and estimate LoS with four cameras: two stereo wide-angle cameras and two stereo narrow field of view cameras [8]. A separate stereo system is used to detect the face in 3D and to direct galvanometer motors to orient the narrow field of view cameras. They use dark-bright pupil principle, but do not exploit information about the light sources. Inspired by Beymer and Flickner, Brolly and Mulligan [11] use a mirror galvanometer system for rapid head movement tracking, but only using a single narrow field camera. Rather than explicitly modeling the eye and the mappings from the stereo and the galvo-coordinates, they propose learning the polynomial regression model. In spite of a lower resolution of the eye images as well as a simpler modeling problem, they obtain accuracies similar to those obtained by Beymer and Flickner [8].

Combinations of stereo systems with pan-tilt have been suggested [135], [113]. Talmi and Liu [135] suggest combining a stereo system for face modeling with a pan-tilt for detailed eye images [135]. Ohno and Mukawa utilize three cameras, two fixed stereo wide-angle cameras and a narrow-angled camera mounted on a pan-tilt unit [113]. Their main result, however, is that two calibration points are necessary in order to estimate the visual axis.

Noureddin et al. [112] suggest a two-camera solution, where a fixed wide-angle camera uses a rotating mirror to direct the orientation of the narrow-angled camera. They show that the rotating mirror speeds up acquisition in comparison to a pan-tilt setup.

Multiple camera solutions have also been successfully applied with head mounts, where one or more cameras are oriented toward the user and one pointing away. The camera that is pointing away is synchronized with the gaze direction [9]. The use of multiple cameras seems to produce robust results, but requires stereo calibration. They are faced with the usual problems of stereo (e.g., point matching, occlusion, and more data to process).

3.2 Other Methods

IR light and feature extraction are important for most current gaze estimation methods. This section reviews methods that follow another path. These alternative approaches include use of visible light [15], [150], [157], [45], the appearance-based approaches [6], [160], [157], and methods that only use the reflections from the layers of the eye avoiding extraction of pupil and iris features (dual-purkinje methods [21], [108]).

3.2.1 Appearance-Based Methods

Feature-based methods require detection of pupils and glints, but the extracted features may be prone to errors. Besides, there may be latent features conveying information about gaze that is not modeled by the chosen features. Similarly to the appearance models of the eyes, appearance-based models for gaze estimation do not explicitly extract features, but rather use the image contents as input with the intention of mapping these directly to screen coordinates (PoR). Consequently, the hope is that the underlying function for estimating the point of regard, relevant features, and personal variation can be extracted implicitly, without requirements of scene geometry and camera calibration. One such approach employs cropped images of eyes to train regression functions, as seen in multilayer network [6], [133], [160] or Gaussian processes [157] or manifold learning [136]. Images are high-dimensional representations of data, which are defined on a lower dimensional manifold. Tan et al. employ Locally Linear Embedding to learn the eye image manifold [136]. They use a significantly lower number of calibration points while improving accuracy as compared to Baluja and Pomerleau [6]. Williams et al. [157] use a sparse Gaussian process interpolation method on filtered visible spectrum images, and consequently, obtain gaze predictions and associated error measurements.

Appearance-based methods typically do not require calibration of cameras and geometry data since the mapping is made directly on the image contents. Thus, they resemble the interpolation-based methods described in Section 3.1. The appearance models have to infer both the geometry and the relevant features from the images, and therefore, tend to require a significant number of calibration points. The relatively high number of calibration points are for some applications less of a problem. It may be more relevant to avoid processing the image (e.g., in the case of a low-resolution eye region) or not requiring glints (e.g., for outdoor use). While appearance methods intend to model the geometry implicitly, no method has reported head pose invariance. The reason is that the appearance of the eye region may look the same under different poses and gaze directions. In addition, given the same pose, change in illumination will also alter the eye appearance and possibly lead to less accuracy. Future methods may reveal how to place geometric priors on appearance models.

3.2.2 Natural Light Methods

Natural light methods are a natural alternative to the use of IR. Natural light approaches face several new challenges such as light changes in the visible spectrum, lower contrast images, but are not as sensitive to the IR light in the environment, and may thus, be potentially better suited when used outdoor [109], [89], [15], [47], [46], [157], [151].

Colombo and Del Bimbo [15] model the visible portions of the user's eyeball as planar surface and regard any gaze shift due to an eyeball rotation as a translation of the pupil in the face plane. Knowing the existence of a one-to-one mapping of the hemisphere and the projective plane, Witzner and Pece [47] model the point of regard as a homographic mapping from the iris center to the monitor. This is only an approximation as the nonlinear one-to-one mapping is not considered. These methods are not head pose invariant. Newman et al. [109] and Wang and Sung [150], [151] propose two separate systems employing stereo and face models to estimate gaze direction. Newman et al. [109] model the eyes as spheres and estimate the point of regard by intersecting the two estimates of line of gaze for each eye. The eyeball center is estimated from a head pose model. Personal calibration is also employed. Wang and Sung [150], [151] also combine a face pose estimation system with a narrow field of view camera to compute the line of gaze through the two irises [150] and one iris [151], respectively. They assume that the iris contour is a circle to estimate its normal direction in 3D through novel eye models.

Gaze estimation methods using rigid facial features have also been proposed [54], [67], [161]. The location of the iris and the eye corners are tracked with a single camera, and by imposing structure-from-motion-like algorithms, the visual axis is estimated. To estimate the point-of-gaze, Matsumoto et al. [97] propose the use of stereo cameras.

These methods work without IR lights, but accuracy is low (about 5 degrees); however, they are in an early stage of development and thus are nonetheless promising for use in a wide range of scenarios. Single camera models are currently limited by the same degenerate configurations as structure-from-motion algorithms, with the implication being that the scale of the head must remain constant.

Methods using visible may also employ corneal reflections since the results obtained using IR are also applicable to visible light. The difference is that the required image features are less accurately depicted in the images, and that visible light may disturb the user and cause the pupil to contract.

3.2.3 Dual Purkinje

A single light source may produce several glints due to reflections from different layers of the eye [21], [108]. When the eye undergoes translation, both the first and fourth reflections (see Fig. 7) move together, but, during rotation, the interdistance of the reflections changes. This interdistance provides a measure of the angular orientation of the eye. Methods using the difference between these reflections (purkinje-images) are called dual-purkinje methods. The accuracy of the Dual-Purkinje-Image technique is generally high, but, since the fourth purkinje image is weak, heavily controlled light conditions are necessary.

3.3 Discussion

Several alternative approaches to gaze estimation have been presented of which the feature-based methods encompass the majority. We have reviewed current techniques expressing the relationships among gaze, eye features (pupil and glints), hardware choice (light sources and cameras), prior geometry information, and pose. Calibration of cameras and geometry, in-session calibration, and the presence of glints are often needed for these techniques to be effective. Additional human-specific parameters may require further calibration. These methods often obtain head pose invariance through the use of glints. Explicit modeling of head pose is most common when glints are not available, e.g., in visible spectrum methods.

Two-dimensional interpolation-based approaches, often used with single camera setups, are relatively simple but are mainly effective when the head remains motionless with reference to the camera(s), either physically restrained or used with head-mounted gaze trackers. Changes in head movements may be compensated for in single glint systems by using additional (explicit) head models, pan-tilt cameras, or by incorporating a rotating mirror. Since the main advantage of a single camera system is low cost and simplicity, these methods seem to complicate matters unnecessarily and disregard the relatively accurate 3D eye models usually obtained by stereo cameras or with additional lights. Methods relying on fully calibrated setups are most common in commercial systems but are limited for public use unless placed in a rigid setup. Any change (e.g., placing the camera differently or changing the zoom of the camera) requires a tedious recalibration. A procedure for effectively performing accurate and automatic system calibration has not yet been reported. Head pose invariance is obtained using at least two light sources in a fully calibrated setup. In the partially calibrated case, a good approximation to the PoR, which is robust to head pose changes, can be obtained with multiple light sources (known position w.r.t. monitor). The 3D model-based approaches, while involving more complex setup and algorithms, can handle head movement robustly and with good accuracy. Stereo approaches obtain 3D measurements only in the overlapping areas of the two visual fields and so the model, and hence, user movement is constrained to this region. Pan-tilt camera solutions allow for greater movement, but have to be reoriented mechanically, which may slow them down. Mirrors have been used to speed up acquisition.

A comparison of the accuracy of different trackers, both research and commercial systems, and a short description of the main characteristics of each system are provided in Fig. 9. Although speed of computation and the number of points necessary for a calibration are important attributes in an eye tracking system, they are not discussed in this review due to both a general lack of formal data and available data being outdated. Note that the numbers reported in the table refer to

Cameras	Lights	Gaze Info	Head pose	Calibration	Accuracy (deg)	References	Comments
1	0	PoR	—	—	$2-4$	[47], [46], [157]	web-camera
1	0	LoG/LoS	—	Fully	$1-2$	[151], [144], [145]	
1	0	LoG	≈	—	< 1	[79]	$*^a$
1	1	PoR	—	—	1-2	[103], [156], [70]	$*^b$
1	2	PoR	✓	Fully	$1-3$	[105], [100], [43]	
1+1	1	PoR	✓	Fully	3	[112]	Mirrors
1(+1)	4	PoR	✓	—	$< 1-2.5$	[164], [20]	
2	0	PoR	✓	—	1	[109]	$*^c$
2+1	1	LoG	✓	—	0.7-1	[135]	pan/tilt
2+2	2	PoR	✓	Fully	0.6	[8]	Mirrors
2	2(3)	PoR	✓	Fully	$< 1-2$	[128], [127]	$*^d$
3	2	PoR	✓	Fully	—	[139][11]	
1	1	PoR	—	—	0.5-1.5	[6], [133], [136], [160]	$*^e$

Fig. 9. Comparison of gaze estimation methods with respective prerequisites and reported accuracies (e.g., based on different data and scenarios). The "cameras" column shows the number of cameras necessary for the methods. An additional "+1" means that an extra pan and tilt camera is used. If this is given in parenthesis, the pan and tilt is used in the implementation, but not necessary by the method. The column "Lights" indicates the number of light sources needed and with an additional set of parenthesis to indicate if extra lights have been used in the implementations. "Gaze info" describes the type of gaze information being inferred by the method (PoR), optical (LoG), or visual axes (LoS). When LoG/LoS is used, it is implicitly assumed that an additional 3D scene model is needed to get the point of regard. The column "Head pose" shows if the methods are head pose invariant (√), if approximate solutions are proposed (≈), or an external head pose unit is needed (—). The "Calibration" column indicates if explicit calibration of scene geometry and cameras are needed prior to use. [a]Additional markers, iris radius, parallel with screen. [b]Polynomial approximation. [c]3D face model. [d]Experiments have been conducted with three glints, but two ought to be sufficient. [e]Appearance-based.

the publications and one should be careful when comparing accuracy since these data come from various sources and because gaze estimates might have been temporally regularized (i.e., smoothing the output).

Theoretical aspects of feature-based gaze tracking based on point sources in a fully calibrated setup are to a large extent understood [43]. However, the estimated fixation points at the border of the monitor tend to be less accurate than those at the central portion of the monitor. There are several identifiable causes for this inaccuracy.

1. The fovea is modeled as a point, but physically, it exists over a small area on the retina.
2. The angle between the line of sight and the optical axis may vary from one fixation point to the next, but the angle is usually modeled as a constant.
3. A spherical model of the eyeball may be sufficient for the central part of the cornea, but it is not representable enough for its periphery, i.e., tracking accuracy may be degraded if the curvature of the cornea varies greatly between subjects.
4. Fixations are used to measure accuracy, but they are, contrary to their name, not stable as the eye jitters due to drift, tremor, and involuntary microsaccades [163].

Refraction and glasses may nonlinearly change the appearance and reflective properties of the eyes as well as the locations of reflections. Refraction causes points presumed to be located on 3D to appear on different lines. The image of the pupil is also altered nonlinearly. Villanueva and Cabeza [146] point out that refraction is an important parameter when modeling pupil images. The difference in gaze accuracy may differ more than 1 degree depending on whether refraction is accounted for. It would therefore be valuable to compare methods using the iris (which is less influenced by refraction) with similar models for the pupil. The use of glasses may likewise confound the physical assumptions of such models (e.g., reflections come from glasses and not from the cornea). We are not aware of any models that geometrically (explicitly) model glasses. Appearance-based interpolation methods implicitly model these nonlinearities.

Eye tracking hardware can be produced and sold with predefined configurations. In this case, applying models where fixed geometry is assumed may be viable. However, these systems do not allow the tailoring of hardware arrangements to particular needs (e.g., in wheelchairs) and may be costly since 1) time-consuming, precise hardware calibration is needed and 2) rigid, purpose-built frames need to be constructed to keep the hardware fixed. Relaxing the prior assumptions of the systems or using low-grade cameras may decrease gaze accuracy and require more session calibration. Note that even the best methods do not guarantee head pose invariance. However, they may, in practice, produce good results if only under optimal working conditions. Depending on the intended application, high accuracy may not be needed. For some cases, it may be more important to lower the price by using Web cameras, allowing for easy and flexible hardware configurations, and avoid IR light and feature detection (e.g., for outdoor use). For example, high accuracy may be required when using gaze for analyzing Web pages or in clinical experiments, while a lower accuracy is required in applications such as environmental control or eye typing, where only a few buttons need to be activated. Similarly, it may be that, for some applications, it is acceptable to use multiple session calibration points, while for others, it is necessary to have only a few (e.g., working with children). Uncalibrated or partially calibrated setups allow for more flexibility but lead to a more difficult modeling problem. Future research may reveal the potential of partially calibrated gaze trackers (e.g., unknown position of light sources, but calibrated cameras) or identify other eye features that provide sufficient information to ensure head pose invariance. Several approaches address the uncalibrated scenario by relying on approximations, and the use of multiple glints is an obvious choice for obtaining robust solutions. However, glints may disappear as the eye or head is moving, and thus, using multiple light sources may restrict user movements in practice. A combination of geometry-based feature methods and appearance-based methods could potentially benefit from the relative strengths of both types of methods.

Both 2D regression-based and appearance-based methods map image data directly to the point of regard. Contrary to the 3D feature-based methods, this method requires multiple in-session calibration points. The gaze estimation problem has an inherent set of parameters associated with it. In fully calibrated settings, the majority of these parameters have been calibrated prior to use, and thus, only a few session calibration points are needed to infer the remaining parameters. Appearance-based models make few assumptions on image features and geometry, and therefore, need to obtain the parameters through session calibration. Comparing methods based on the required number of calibration points should only be done with care. The choice of model depends on multiple factors: required accuracy, hardware cost, image quality/eye region resolution, available information in the image (e.g., glints), and flexibility of setup.

There are several possible important findings, which could benefit current models. For example, Donder's and Listing's laws and recent discoveries about the cognitive or perceptual processes underlying eye movements are rarely used explicitly. Besides this, features such as the iris and pupil contour as well as other facial features provide useful additional information that could be used to reduce the required number of light sources [146].

4 Eye Detection and Gaze Tracking Applications

Eye detection and gaze tracking have found numerous applications in multiple fields. Eye detection is often the first and one of the most important steps for many computer vision applications such as facial recognition, facial feature tracking, facial expression analysis, as well as in iris detection and iris recognition. The accuracy of eye detection directly affects the subsequent processing and recognition. In addition, automatic recovery of eye position and eye status (open/close) from image sequences is one of the important topics for model-based coding of videophone sequences and driver fatigue applications.

Gaze tracking offers a powerful empirical tool for the study of real-time cognitive processing and information transfer. Gaze tracking applications include two main fields of application, namely, *diagnostic* and *interactive* [28]. Diagnostic eye trackers provide an objective and quantitative method for recording the viewer's point of regard. This information is useful when examining people watching commercials, using instruments in plane cockpits, and interacting with user interfaces, and in the analysis and understanding of human attention [3], [41], [119], [155].

In contrast, gaze-based interactive user interfaces react to the user's gaze either as a *control* input [7], [10], [52], [87], [154] or as the basis of *gaze-contingent* change in display. Gaze-contingent means that the system is aware of the user's gaze and may adapt its behavior based on the visual attention of the user, e.g., for monitoring human vigilance [28], [66], [65], [69], [143]. Thus, the system tends to adapt its behavior according to the gaze input, which, in turn, reflects the person's desires. This property of eye movements, as well as the fact that eye tracking facilitates hands-free interaction with little muscle strain, makes gaze tracking systems a unique and effective tool for disabled people, where eye movements may be the only available means of communication and interaction with the computer and other people. Specifically, early work on interactive eye tracking applications focused primarily on users with disabilities [62], [88], [137]. Among the first applications were "eye typing" systems, where the user could produce text through gaze inputs (for a review, see [95]). For some applications, eye movements are more natural, fast, and comfortable means of communication, and the tendency now is to develop gaze-based applications for the benefit of all. For example, anyone reading foreign languages could be provided with suggestions for words and sentences based on eye movement patterns as they read [64]. Using eye trackers ubiquitously may require gaze-based application designers to be more conscious of current challenges such as higher noise levels on gaze estimates. A new approach addressing navigation and selections in large information spaces with noisy inputs is suggested by Witzner et al. [51]. Nonintrusive gaze tracking may be used for interaction with computer in a similar way to using the mouse, or in game-like interaction with videos, where the viewer seamlessly interacts and defines the narrative of the video. Eye tracking also seems to be gaining interest in the vehicle industry for driver vigilance and safety. Another important and yet perhaps least investigated application involves using eye movements to gain some insight into the way that people view synthesized images and animations, with the dual purpose of optimizing perceived quality and developing more efficient algorithms.

5 Summary and Conclusions

We have presented a review categorizing eye tracking systems from numerous angles; from the different methods of detecting and tracking eye images to computational models of eyes for gaze estimation and gaze-based applications. Specifically, for eye detection and tracking, we have discussed various techniques using different properties of the eyes including appearance, shape, motion, or some combination.

While these methods have been successful in improving eye detection and tracking, there remains significant potential for further developments. Reliably detecting and tracking eyes in conditions of variable face pose and variable ambient lighting remains largely problematic. It appears that an integrated approach exploiting several available attributes is the promising direction for further development. While eyes are nonrigid, their spatial relation to other parts of the face is relatively stable. These relations are of potential interest for eye detection models, for example, through patch-based approaches.

We have reviewed several categories of techniques for gaze estimation. While the regression-based methods using a single glint are simple and fairly accurate, they are only suited to particular applications due to their restrictions with regard to head movements. This restriction can be relaxed by using a wide-angle face camera and a pan-tilt-controlled eye camera. However, this setup increases both the complexity and cost. The 3D model-based eye tracking systems can tolerate natural head movements, but they usually require a one-time system and geometric calibration. In the fully calibrated setup, considering possible light and camera configurations, the one-camera and two-light-source configuration appears to be an interesting choice (especially for commercial systems) since its setup is simple and robust to head pose changes. Adding an additional camera may reduce the number of calibration points since the assumption of the known cornea curvature is no longer needed. When only modeling the visual and not the optical axes, the number of calibration points can be reduced from two to one, in addition to the system calibration. A low number of calibration points are preferable, but this requires complete knowledge of the geometric arrangement of system parts and the eye/face. The desire for a simple calibration procedure with few calibration points therefore implies decreased flexibility and increased price. Appearance-based methods, on the other hand, are not based on known parameters of feature extraction, setup, and light conditions, and may therefore be more flexible and simple. Since they must infer more parameters, they require more session calibration and do not ensure head pose invariance.

In summary, future gaze tracking systems should still be low cost, easy to setup, minimal or no calibration, and good gaze estimation accuracy under varying illumination conditions and natural head movements. Some of these requirements are currently conflicting, for example, flexibility and a low number of calibration points.

Future directions for the eye and gaze trackers include:

- **Limit the use of IR**: IR light is useful for eye trackers, mainly because it is not only invisible to the user but also it can be used for controlling light conditions, obtaining higher contrast images, and stabilizing gaze estimation. A practical limitation of systems using IR light is that they are not necessarily reliable when used outdoors. Future eye tracking systems should also be able to function outdoors. Current efforts in this direction employ structure-from-motion methods on facial feature points. These techniques remain in an early stage of development and further research is needed.
- **Head mounts**: While significant emphasis has been placed on remote gaze tracking, head-mounted gaze trackers could be experiencing a renaissance due to both the challenges facing remote eye trackers, and to the increased interest in mobile eye tracking and tiny head-mounted displays. In these cases, eye trackers could facilitate hands-free interaction and improve the quality of the displays. Head-mounted eye trackers may also be more accurate since they are less affected by external changes (head pose, lights, etc.) and the simplified geometry may allow for more constraints to be applied. For example, the use of glints may become unnecessary.
- **Flexible setup**: Many current gaze trackers require calibration of the camera(s) and the geometric arrangement. In some situations, it would be convenient for the light sources, cameras, and monitor to be as needed, but without requiring explicit calibration of geometry and cameras. For example, this would benefit eye trackers intended for mobility and the mass market, as the rigid frames can be avoided, resulting in more compact, lightweight, adaptable, and cheap eye trackers.
- **Limit calibration**: Current gaze models either use a strong prior model (hardware calibration) with little session calibration or weak prior model, but more calibration points. Another future direction will be to develop methods that do not require any calibration. This does not seem to be possible given the current eye and gaze models. New eye models and theories need to be developed to achieve calibration-free gaze tracking.
- **Costs**: The costs of current eye tracking systems remain too high for general public use. The main reason for this is the cost of parts, especially high-quality cameras and lenses, the cost of development, and the relatively small market. Alternatively, systems may opt for standard or even off-the-shelf components such as digital video and Web cameras and exploit the fast development in this area [46], [47], [89]. While advances in new camera and sensor technology may add to the continuing progress in the fields, new theoretical developments are needed in order to perform accurate gaze tracking with low-quality images.
- **Higher degree of tolerance**: Tolerance toward glasses and contact lenses is a practical problem that has been solved only partially. The use of several light sources, synchronized according to the users' head movement relative to the camera and light source, may remove some of the associated problems. However, more detailed modeling such as modeling glass themselves may be needed if eye trackers are to be used outdoors, where light conditions are less controllable.

  The tendency to produce mobile and low-cost systems may increase the ways in which eye tracking technology can be applied to mainstream applications, but may also lead to less accurate gaze tracking. While high accuracy may not be needed for such applications, mobile systems must be able to cope with higher noise levels than eye trackers indoor use.
- **Interpretation of gaze**: Besides the technical issues of localizing the eye and determining gaze, the interpretation that the cognitive and affective states underlying gaze behavior is also important. The analysis of eye movement behavior involves

understanding human visual perception and cognition, as well as the emotional and cognitive states associated with the task. Development of applications that exploit a combination of gaze with other gestures and known neuropsychological correlates of human eye movements certainly provides sufficient material for long-term research.

While the techniques surveyed in this paper focus on eye detection and gaze tracking, many of the same techniques can be useful for detection and tracking of other objects (e.g., faces). Despite the fact that describing the structure of the eye is relatively simple, the complexities in its appearance make the eye a challenging research topic.

Eye and gaze tracking and their applications involve unique and clearly defined problems, which have already spawned new models, influencing research beyond eye tracking [157], [166], but eyes could also be seen as a primary case for future models in image analysis, geometry, and machine learning due to the inherent challenging properties of the eye as a trackable subject. For this reason, research in the area of eye tracker development is of increasing interest to a wide variety of research fields.

ACKNOWLEDGMENTS

This paper is partially funded by the European network of Excellence, COGAIN, supported by the European Commission's IST 6th Framework Program. Qiang Ji is partially funded by the US Air Force Office of Sceintific Research (F49620-03-1-0160) and the US Defense Advanced Research Projects Agency (DARPA) (N00014-03-01-1003). The authors would like to thank Arantxa Villanueva, Fiona Mulvey, Martin Böhme, and Javier San Agustin for fruitful comments and discussions.

REFERENCES

[1] J.S. Agustin, A. Villanueva, and R. Cabeza, "Pupil Brightness Variation as a Function of Gaze Direction," *Proc. 2006 Symp. Eye Tracking Research and Applications,* pp. 49-49, 2006.

[2] A. Amir, L. Zimet, A. Sangiovanni-Vincentelli, and S. Kao, "An Embedded System for an Eye-Detection Sensor," *Computer Vision and Image Understanding,* vol. 98, no. 1, pp. 104-123, 2005.

[3] G. Anders, "Pilot's Attention Allocation during Approach and Landing—Eye- and Head-Tracking Research," *Proc. 11th Int'l Symp. Aviation Psychology,* 2001.

[4] J. Bala, K. De Jong, J. Huang, H. Vafaie, and H. Wechsler, "Visual Routine for Eye Detection Using Hybrid Genetic Architectures," *Proc. Int'l Conf. Pattern Recognition,* 1996.

[5] L.-P. Bala, K. Talmi, and J. Liu, "Automatic Detection and Tracking of Faces and Facial Features in Video Sequences," *Proc. Picture Coding Symp.,* Sept. 1997.

[6] S. Baluja and D. Pomerleau, "Non-Intrusive Gaze Tracking Using Artificial Neural Networks," *Advances in Neural Information Processing Systems,* J.D. Cowan, G. Tesauro, and J. Alspector, eds., vol. 6, pp. 753-760, Morgan Kaufmann Publishers, 1994.

[7] P. Baudisch, D.D. Andrew, T. Duchowski, and W.S. Geisler, "Focusing on the Essential: Considering Attention in Display Design," *Comm. ACM,* vol. 46, no. 3, pp. 60-66, 2003.

[8] D. Beymer and M. Flickner, "Eye Gaze Tracking Using an Active Stereo Head," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. II, pp. 451-458, 2003.

[9] G. Boening, K. Bartl, T. Dera, S. Bardins, E. Schneider, and T. Brandt, "Mobile Eye Tracking as a Basis for Real-Time Control of a Gaze Driven Head-Mounted Video Camera," *Proc. 2006 Symp. Eye Tracking Research and Applications,* p. 56, 2006.

[10] R. Bolt, "Gaze-Orchestrated Dynamic Windows," *Proc. ACM SIGGRAPH,* pp. 109-119, 1981.

[11] X.L.C. Brolly and J.B. Mulligan, "Implicit Calibration of a Remote Gaze Tracker," *Proc. 2004 Conf. Computer Vision and Pattern Recognition Workshop,* vol. 8, p. 134, 2004.

[12] R.H.S. Carpenter, *Movements of the Eyes.* Pion Limited, 1988.

[13] G. Chow and X. Li, "Towards a System for Automatic Facial Feature Detection," *Pattern Recognition,* vol. 26, pp. 1739-1755, 1993.

[14] COGAIN, http://www.cogain.org/, 2007.

[15] C. Colombo and A.D. Bimbo, "Real-Time Head Tracking from the Deformation of Eye Contours Using a Piecewise Affine Camera," *Pattern Recognition Letters,* vol. 20, no. 7, pp. 721-730, July 1999.

[16] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 5, pp. 564-577, May 2003.

[17] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *Proc. European Conf. Computer Vision,* vol. 2, pp. 484-498, 1998.

[18] T.F. Cootes and C.J Taylor, "Active Shape Models—'Smart Snakes'," *Proc. British Machine Vision Conf.,* pp. 266-275, 1992.

[19] J. Coughlan, A. Yuille, C. English, and D. Snow, "Efficient Deformable Template Detection and Localization without User Initialization," *Computer Vision and Image Understanding,* vol. 78, no. 3, pp. 303-319, 2000.

[20] F.L. Coutinho and C.H. Morimoto, "Free Head Motion Eye Gaze Tracking Using a Single Camera and Multiple Light Sources," *Proc. ACM SIGGRAPH,* M.M. de Oliveira Neto and R. Lima Carceroni, eds., Oct. 2006.

[21] H. Crane and C. Steele, "Accurate Three-Dimensional Eye Tracker," *J. Optical Soc. Am.,* vol. 17, no. 5, pp. 691-705, 1978.

[22] D. Cristinacce and T. Cootes, "Feature Detection and Tracking with Constrained Local Models," *Proc. 17th British Machine Vision Conf.,* pp. 929-938, 2006.

[23] J.L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications," *Proc. 1997 IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 640-645, 1997.

[24] J. Daugman, "The Importance of Being Random: Statistical Principles of Iris Recognition," *Pattern Recognition,* vol. 36, no. 2, pp. 279-291, 2003.

[25] J. Deng and F. Lai, "Region-Based Template Deformation and Masking for Eye-Feature Extraction and Description," *Pattern Recognition,* vol. 30, pp. 403-419, 1997.

[26] T. D'Orazio, M. Leo, G. Cicirelli, and A. Distante, "An Algorithm for Real Time Eye Detection in Face Images," *Proc. 17th Int'l Conf. Pattern Recognition,* vol. 3, no. 0, pp. 278-281, 2004.

[27] D. Droege, C. Schmidt, and D. Paulus, "A Comparison of Pupil Centre Estimation Algorithms," *Proc. Conf. COGAIN 2008—Comm., Environment, and Mobility Control by Gaze,* H. Istance, O. Stepankova, and R. Bates, eds., pp. 23-26, 2008.

[28] A. Duchowski, *Eye Tracking Methodology: Theory and Practice.* Springer-Verlag, 2003.

[29] A.T. Duchowski, E. Medlin, N. Cournia, A. Gramopadhye, B. Melloy, and S. Nair, "3D Eye Movement Analysis for VR Visual Inspection Training," *Proc. 2002 Symp. Eye Tracking Research and Applications,* pp. 103-110, 2002.

[30] Y. Ebisawa, "Improved Video-Based Eye-Gaze Detection Method," *IEEE Trans. Instrumentation and Measurement,* vol. 47, no. 2, pp. 948-955, Aug. 1998.

[31] Y. Ebisawa and S. Satoh, "Effectiveness of Pupil Area Detection Technique Using Two Light Sources and Image Difference Method," *Proc. 15th Ann. Int'l Conf. IEEE Eng. in Medicine and Biology Soc.,* pp. 1268-1269, 1993.

[32] Y. Ebisawa, "Realtime 3D Position Detection of Human Pupil," *Proc. 2004 IEEE Symp. Virtual Environments, Human-Computer Interfaces and Measurement Systems,* pp. 8-12, 2004.

[33] G.J. Edwards, T.F. Cootes, and C.J. Taylor, "Face Recognition Using Active Appearance Models," *Proc. Fifth European Conf. Computer Vision,* vol. 2, pp. 581-95, 1998.

[34] A. Tomono et al., "Pupil Extraction Processing and Gaze Point Detection System Allowing Head Movement," *Trans. Inst. of Electronics, Information, and Comm. Eng. of Japan,* vol. J76-D II, no. 3, 1993.

[35] I.R. Fasel, B. Fortenberry, and J.R. Movellan, "A Generative Framework for Real Time Object Detection and Classification," *Computer Vision and Image Understanding,* vol. 98, no. 1, pp. 182-210, Apr. 2005.

[36] G.C. Feng and P.C. Yuen, "Variance Projection Function and Its Application to Eye Detection for Human Face Recognition," *Int'l J. Computer Vision,* vol. 19, pp. 899-906, 1998.

[37] G.C. Feng and P.C. Yuen, "Multi-Cues Eye Detection on Gray Intensity Image," *Pattern Recognition,* vol. 34, pp. 1033-1046, 2001.

[38] R.S. Feris, T.E. de Campos, and R.M. Cesar Jr., "Detection and Tracking of Facial Features in Video Sequences," *Proc. Conf. Medical Image Computing and Computer-Assisted Intervention,* pp. 127-135, 2000.

[39] V. Di Gesu and C. Valenti, "Symmetry Operators in Computer Vision," *Vistas Astronomy,* vol. 40, no. 4, pp. 461-468, 1996.

[40] Y. Gofman and N. Kiryati, "Detecting Symmetry in Gray Level Images: The Global Optimization Approach," *Proc. Int'l Conf. Pattern Recognition,* 1996.

[41] J.H. Goldberg and A.M. Wichansky, *Eye Tracking in Usability Evaluation: A Practitioner's Guide,* pp. 493-516. Elsevier Science, 2003.

[42] K. Grauman, M. Betke, J. Gips, and G.R. Bradski, "Communication via Eye Blinks: Detection and Duration Analysis in Real Time," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. I, pp. 1010-1017, 2001.

[43] E.D. Guestrin and M. Eizenman, "General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections," *IEEE Trans. Biomedical Eng.,* vol. 53, no. 6, pp. 1124-1133, June 2006.

[44] P.W. Hallinan, "Recognizing Human Eyes," *Geometric Methods in Computer Vision,* pp. 212-226, SPIE, 1991.

[45] D.W. Hansen, "Comitting Eye Tracking," PhD thesis, IT Univ. of Copenhagen, 2003.

[46] D. Witzner Hansen, J.P. Hansen, M. Nielsen, A.S. Johansen, and M.B. Stegmann, "Eye Typing Using Markov and Active Appearance Models," *Proc. IEEE Workshop Applications on Computer Vision,* pp. 132-136, 2003.

[47] D.W. Hansen and A.E.C. Pece, "Eye Tracking in the Wild," *Computer Vision and Image Understanding,* vol. 98, no. 1, pp. 182-210, Apr. 2005.

[48] D.W. Hansen, "Using Colors for Eye Tracking," *Color Image Processing: Methods and Applications,* pp. 309-327, CRC Press, 2006.

[49] D.W. Hansen and R. Hammoud, "An Improved Likelihood Model for Eye Tracking," *Computer Vision and Image Understanding,* 2007.

[50] D.W. Hansen and J.P. Hansen, "Robustifying Eye Interaction," *Proc. Conf. Vision for Human Computer Interaction,* pp. 152-158, 2006.

[51] D.W. Hansen, H.H.T. Skovsgaard, J.P. Hansen, and E. Møllenbach, "Noise Tolerant Selection by Gaze-Controlled Pan and Zoom in 3D," *Proc. 2008 Symp. Eye Tracking Research and Applications,* pp. 205-212, 2008.

[52] J.P. Hansen, K. Itoh, A.S. Johansen, K. Tørning, and A. Hirotaka, "Gaze Typing Compared with Input by Head and Hand," *Proc. Eye Tracking Research and Applications Symp. 2004,* pp. 131-138, 2004.

[53] A. Haro, M. Flickner, and I. Essa, "Detecting and Tracking Eyes by Using Their Physiological Properties, Dynamics, and Appearance," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2000.

[54] J. Heinzmann and A. Zelinsky, "3D Facial Pose and Gaze Point Estimation Using a Robust Real-Timetracking Paradigm," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* 1998.

[55] C. Hennessey, B. Noureddin, and P. Lawrence, "A Single Camera Eye-Gaze Tracking System with Free Head Motion," *Proc. Symp. Eye Tracking Research and Applications,* pp. 87-94, 2006.

[56] C. Hennesy and P. Lawrence, "3D Point-of-Gaze Estimation on a Volumetric Display," *Proc. 2008 Symp. Eye Tracking Research and Applications,* 2008.

[57] R. Herpers, M. Michaelis, K. Lichtenauer, and G. Sommer, "Edge and Keypoint Detection in Facial Regions," *Proc. Int'l Conf. Automatic Face and Gesture-Recognition,* pp. 212-217, 1996.

[58] P.M. Hillman, J.M. Hannah, and P.M. Grant, "Global Fitting of a Facial Model to Facial Features for Model-Based Video Coding," *Proc. Third Int'l Symp. Image and Signal Processing and Analysis,* vol. 1, pp. 359-364, 2003.

[59] J. Huang, D. Ii, X. Shao, and H. Wechsler, "Pose Discrimination and Eye Detection Using Support Vector Machines (SVMs)," *Proc. Conf. NATO-ASI on Face Recognition: From Theory to Applications,* pp. 528-536, 1998.

[60] J. Huang and H. Wechsler, "Eye Location Using Genetic Algorithms," *Proc. Second Int'l Conf. Audio and Video-Based Biometric Person Authentication,* 1999.

[61] J. Huang and H. Wechsler, "Eye Detection Using Optimal Wavelet Packets and Radial Basis Functions (RBFs)," *Int'l J. Pattern Recognition and Artificial Intelligence,* vol. 13, no. 7, 1999.

[62] T.E. Hutchinson, "Human-Computer Interaction Using Eye-Gaze Input," *IEEE Tran. Systems, Man, and Cybernetics,* vol. 19, no. 6, pp. 1527-1534, Nov./Dec. 1989.

[63] K. Hyoki, M. Shigeta, N. Tsuno, Y. Kawamuro, and T. Kinoshita, "Quantitative Electro-Oculography and Electroencephalography as Indices of Alertness," *Electroencephalography and Clinical Neurophysiology,* vol. 106, pp. 213-219, 1998.

[64] A. Hyrskykari, P. Majaranta, A. Aaltonen, and K.-J. Räihä, "Design Issues of iDict: A Gaze-Assisted Translation Aid," *Proc. Symp. Eye Tracking Research and Applications 2000,* pp. 9-14, 2000.

[65] A. Hyrskykari, P. Majaranta, and K.-J. Räihä, "Proactive Response to Eye Movements," *Proc. IFIP TC13 Int'l Conf. Human-Computer Interaction,* M. Rauterberg, M. Menozzi, and J. Wesson, eds., pp. 129-136, 2003.

[66] A. Hyrskykari, P. Majaranta, and K.-J. Räihä, "From Gaze Control to Attentive Interfaces," *Proc. 11th Int'l Conf. Human-Computer Interaction,* 2005.

[67] T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, "Passive Driver Gaze Tracking with Active Appearance Models," *Proc. 11th World Congress Intelligent Transportation Systems,* Oct. 2004.

[68] J.P. Ivins and J. Porrill, "A Deformable Model of the Human Iris for Measuring Small 3-Dimensional Eye Movements," *Machine Vision and Applications,* vol. 11, no. 1, pp. 42-51, 1998.

[69] Q. Ji and X. Yang, "Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance," *Real-Time Imaging,* vol. 8, no. 5, pp. 357-377, 2002.

[70] Q. Ji and Z. Zhu, "Eye and Gaze Tracking for Interactive Graphic Display," *Proc. Second Int'l Symp. Smart Graphics,* pp. 79-85, 2002.

[71] M. Kampmann and L. Zhang, "Estimation of Eye, Eyebrow and Nose Features in Videophone Sequences," *Proc. Int'l Workshop Very Low Bitrate Video Coding,* 1998.

[72] J.J. Kang, E.D. Guestrin, and E. Eizenman, "Investigation of the Cross-Ratio Method for Point-of-Gaze Estimation," *Trans. Biometrical Eng.,* 2008.

[73] F. Karmali and M. Shelhamer, "Compensating for Camera Translation in Video Eye Movement Recordings by Tracking a Landmark Selected Automatically by a Genetic Algorithm," *Proc. Ann. Int'l Conf. IEEE Eng. in Medicine and Biology,* pp. 5298-5301 and 4029752, 2006.

[74] S. Kawato and N. Tetsutani, "Detection and Tracking of Eyes for Gaze-Camera Control," *Proc. Int'l Conf. Vision Interface,* p. 348, 2002.

[75] S. Kawato and J. Ohya, "Real-Time Detection of Nodding and Head-Shaking by Directly Detecting and Tracking the Between-Eyes," *Proc. IEEE Fourth Int'l Conf. Automatic Face and Gesture Recognition,* pp. 40-45, 2000.

[76] S. Kawato and J. Ohya, "Two-Step Approach for Real-Time Eye Tracking with a New Filtering Technique," *Proc. Int'l Conf. System, Man and Cybernetics,* pp. 1366-1371, 2000.

[77] S. Kawato and N. Tetsutani, "Detection and Tracking of Eyes for Gaze-Camera Control," *Proc. 15th Int'l Conf. Vision Interface,* 2002.

[78] S. Kawato and N. Tetsutani, "Real-Time Detection of Between-the-Eyes with a Circle Frequency Filter," *Proc. Asian Conf. Computer Vision '02,* vol. II, pp. 442-447, 2002.

[79] K.-N. Kim and R.S. Ramakrishna, "Vision-Based Eye-Gaze Tracking for Human Computer Interface," *Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics,* 1999.

[80] S. Kim and Q. Ji, "Non-Intrusive Eye Gaze Tracking under Natural Head Movements," *Proc. 26th Ann. Int'l Conf. IEEE Eng. in Medicine and Biology,* Sept. 2004.

[81] C. Kimme, D. Ballard, and J. Sklansky, "Finding Circles by an Array of Accumulators," *Comm. ACM,* vol. 18, no. 2, pp. 120-122, Feb. 1975.

[82] I. King and L. Xu, "Localized Principal Component Analysis Learning for Face Feature Extraction and Recognition," *Proc. Workshop 3D Computer Vision,* pp. 124-128, 1997.

[83] S.M. Kolakowski and J.B. Pelz, "Compensating for Eye Tracker Camera Movement," *Proc. 2006 Symp. Eye Tracking Research and Applications,* pp. 79-85, 2006.

[84] R. Kothari and J.L. Mitchell, "Detection of Eye Locations in Unconstrained Visual Images," *Proc. Int'l Conf. Image Processing,* vol. 3, pp. 519-522, 1996.

[85] P. Kovesi, "Symmetry and Asymmetry from Local Phase," *Proc. 10th Australian Joint Conf. Artificial Intelligence,* pp. 185-190, 1997.

[86] K. Lam and H. Yan, "Locating and Extracting the Eye in Human Face Images," *Pattern Recognition,* vol. 29, pp. 771-779, 1996.

[87] C. Lankford, "Effective Eye-Gaze Input into Windows," *Proc. Eye Tracking Research and Applications Symp. '00,* pp. 23-27, 2000.

[88] J.L. Levine, "An Eye-Controlled Computer," Technical Report RC-8857, IBM T.J. Watson Research Center, 1982.

[89] D. Li, D. Winfield, and D.J. Parkhurst, "Starburst: A Hybrid Algorithm for Video-Based Eye Tracking Combining Feature-Based and Model-Based Approaches," *Proc. Vision for Human-Computer Interaction Workshop, IEEE Computer Vision and Pattern Recognition Conf.,* 2005.

[90] C.-C. Lin and W.-C. Lin, "Extracting Facial Features by an Inhibitory Mechanism Based on Gradient Distribution," *Pattern Recognition,* vol. 29, no. 12, pp. 2079-2101, 1996.

[91] P.J. Locher and C.F. Nodine, "Symmetry Catches the Eye," *Eye Movements—from Physiology to Cognition,* pp. 353-361, 1987.

[92] P.J. Locher and C.F. Nodine, "The Perceptual Value of Symmetry," *Computers and Math. Applications,* vol. 17, pp. 475-484, 1989.

[93] G. Loy and A. Zelinsky, "Fast Radial Symmetry for Detecting Points of Interest" *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 8, pp. 959-973, Aug. 2003.

[94] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. Int'l Joint Conf. Artificial Intelligence,* 1981.

[95] P. Majaranta and K.-J. Räihä, "Twenty Years of Eye Typing: Systems and Design Issues," *Proc. Symp. Eye Tracking Research and Applications,* pp. 15-22, 2002.

[96] T. Marui and Y. Ebisawa, "Eye Searching Technique for Video-Based Eye-Gaze Detection," *Proc. 20th Ann. Int'l Conf. IEEE Eng. in Medicine and Biology Soc. '98,* vol. 2, pp. 744-747, 1998.

[97] Y. Matsumoto, T. Ogasawara, and A. Zelinsky, "Behaviour Recognition Based on Head Pose and Gaze Direction Measurement," *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems,* pp. 2127-2132, 2000.

[98] Y. Matsumoto and A. Zelinsky, "An Algorithm for Real-Time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 499-504, 2000.

[99] J. Merchant, R. Morrissette, and J. Porterfield, "Remote Measurements of Eye Direction Allowing Subject Motion over One Cubic Foot of Space," *IEEE Trans. Biomedical Eng.,* vol. 21, no. 4, pp. 309-317, July 1974.

[100] A. Meyer, M. Böhme, T. Martinetz, and E. Barth, "A Single-Camera Remote Eye Tracker," *Perception and Interactive Technologies,* pp. 208-211, Springer, 2006.

[101] J.M. Miller, H.L. Hall, J.E Greivenkamp, and D.L. Guyton, "Quantification of the Brückner Test for Strabismus," *Investigative Ophthalmology & Visual Science,* vol. 36, no. 4, pp. 897-905, 1995.

[102] W.M. Huang and R. Mariani, "Face Detection and Precise Eyes Location," *Proc. Int'l Conf. Pattern Recognition,* 2000.

[103] C.H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil Detection and Tracking Using Multiple Light Sources," *Image and Vision Computing,* vol. 18, no. 4, pp. 331-335, 2000.

[104] C.H. Morimoto and M.R.M. Mimica, "Eye Gaze Tracking Techniques for Interactive Applications," *Computer Vision and Image Understanding,* vol. 98, no. 1, pp. 4-24, Apr. 2005.

[105] C. Morimoto, A. Amir, and M. Flickner, "Detecting Eye Position and Gaze from a Single Camera and 2 Light Sources," *Proc. Int'l Conf. Pattern Recognition,* 2002.

[106] C.H. Morimoto and M. Flickner, "Real-Time Multiple Face Detection Using Active Illumination," *Proc. Fourth IEEE Int'l Conf. Automatic Face and Gesture Recognition '00,* 2000.

[107] C.H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil Detection and Tracking Using Multiple Light Sources," Technical Report RJ-10117, IBM Almaden Research Center, 1998.

[108] P. Müller, D. Cavegn, G. d'Ydewalle, and R. Groner, "A Comparison of a New Limbus Tracker, Corneal Reflection Technique, Purkinje Eye Tracking and Electro-Oculography," *Perception and Cognition,* G. d'Ydewalle and J.V. Rensbergen, eds., pp. 393-401, Elsevier Science Publishers, 1993.

[109] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, "Real-Time Stereo Tracking for Head Pose and Gaze Estimation," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 122-128, 2000.

[110] K. Nguyen, C. Wagner, D. Koons, and M. Flickner, "Differences in the Infrared Bright Pupil Response of Human Eyes," *Proc. Symp. Eye Tracking Research and Applications,* pp. 133-138, 2002.

[111] M. Nixon, "Eye Spacing Measurement for Facial Recognition," *Proc. Conf. Soc. Photo-Optical Instrument Eng.,* 1985.

[112] B. Noureddin, P.D. Lawrence, and C.F. Man, "A Non-Contact Device for Tracking Gaze in a Human Computer Interface," *Computer Vision and Image Understanding,* vol. 98, no. 1, pp. 52-82, 2005.

[113] T. Ohno and N. Mukawa, "A Free-Head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction," *Proc. Eye Tracking Research and Applications Symp. 2004,* pp. 115-122, 2004.

[114] T. Ohno, "One-Point Calibration Gaze Tracking Method," *Proc. 2006 Symp. Eye Tracking Research and Applications,* pp. 34-34, 2006.

[115] T. Ohno, N. Mukawa, and A. Yoshikawa, "Freegaze: A Gaze Tracking System for Everyday Gaze Interaction," *Proc. Eye Tracking Research Applications Symp.,* pp. 125-132, 2002.

[116] K.R. Park, J.J. Lee, and J. Kim, "Facial and Eye Gaze Detection," *Biologically Motivated Computer Vision,* pp. 368-376, Springer, 2002.

[117] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 1994.

[118] A. Perez, M.L. Cordoba, A. Garcia, R. Mendez, M.L. Munoz, J.L. Pedraza, and F. Sanchez, "A Precise Eye-Gaze Detection and Tracking System," *J. WSCG,* pp. 105-108, 2003.

[119] K. Rayner, C. Rotello, C. Steward, A. Keir, and A. Duffy, "When Looking at Print Advertisements," *J. Experimental Psychology: Applied,* vol. 7, no. 3, pp. 219-226, 2001.

[120] M. Reinders, R. Koch, and J. Gerbrands, "Locating Facial Features in Image Sequences Using Neural Networks," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition,* 1997.

[121] D. Reisfeld and Y. Yeshurun, "Robust Detection of Facial Features by Generalized Symmetry," *Proc. Int'l Conf. Pattern Recognition,* vol. I, pp. 117-120, 1992.

[122] Arrington Research, http://www.arringtonresearch.com, 2007.

[123] RPI, http://www.ecse.rpi.edu/homepp./cvrl/database/database.html, 2008.

[124] F. Samaria and S. Young, "HMM-Based Architecture for Face Identification," *Image and Vision Computing,* vol. 12, no. 8, pp. 537-543, 1994.

[125] D. Scott and J. Findlay, "Visual Search, Eye Movements and Display Units," human factors report, 1993.

[126] G. Sela and M.D. Levine, "Real-Time Attention for Robotic Vision," *Real-Time Imaging,* vol. 3, pp. 173-194, 1997.

[127] S.-W. Shih and J. Liu, "A Novel Approach to 3D Gaze Tracking Using Stereo Cameras," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 34, no. 1, pp. 234-245, Feb. 2004.

[128] S.-W. Shih, Y.-T. Wu, and J. Liu, "A Calibration-Free Gaze Tracking Technique," *Proc. 15th Int'l Conf. Pattern Recognition,* pp. 201-204, 2000.

[129] S. Sirohey, A. Rosenfeld, and Z. Duric, "A Method of Detecting and Tracking Irises and Eyelids in Video," *Pattern Recognition,* vol. 35, no. 6, pp. 1389-1401, June 2002.

[130] S.A. Sirohey and A. Rosenfeld, "Eye Detection in a Face Image Using Linear and Nonlinear Filters," *Pattern Recognition,* vol. 34, pp. 1367-1391, 2001.

[131] D.M. Stampe, "Heuristic Filtering and Reliable Calibration Methods for Video-Based Pupil-Tracking Systems," *Behaviour Research Methods, Instruments & Computers,* vol. 25, no. 2, pp. 137-142, 1993.

[132] R. Stiefelhagen, J. Yang, and A. Waibel, "A Model-Based Gaze Tracking System," *Proc. IEEE Int'l Joint Symp. Intelligence and Systems,* pp. 304-310, 1996.

[133] R. Stiefelhagen, J. Yang, and A. Waibel, "Tracking Eyes and Monitoring Eye Gaze," *Proc. Workshop Perceptual User Interfaces,* pp. 98-100, 1997.

[134] A. Sugioka, Y. Ebisawa, and M. Ohtani, "Noncontact Video-Based Eye-Gaze Detection Method Allowing Large Head Displacements," *Proc. Ann. Int'l Conf. IEEE Eng. in Medicine and Biology,* vol. 2, pp. 526-528, 1996.

[135] K. Talmi and J. Liu, "Eye and Gaze Tracking for Visually Controlled Interactive Stereoscopic Displays," *Signal Processing: Image Comm.,* vol. 14, no. 10, pp. 799-810, 1999.

[136] K.-H. Tan, D.J. Kriegman, and N. Ahuja, "Appearance-Based Eye Gaze Estimation," *Proc. Sixth IEEE Workshop Applications of Computer Vision '02,* pp. 191-195, 2002.

[137] J.H. Ten Kate, E.E.E. Frietman, W. Willems, B.M. Ter Haar Romeny, and E. Tenkink, "Eye-Switch Controlled Communication Aids," *Proc. 12th Int'l Conf. Medical and Biological Eng.,* 1979.

[138] Y. Tian, T. Kanade, and J.F. Cohn, "Dual-State Parametric Eye Tracking," *Proc. Fourth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* 2000.

[139] A. Tomono, M. Iida, and Y. Kobayashi, "A TB Camera System which Extracts Feature Points for Non-Contact Eye Movement Detection," *Proc. SPIE Optics, Illumination, and Image Sensing for Machine Vision,* pp. 2-12, 1989.

[140] D. Tweed and T. Vilis, "Geometric Relations of Eye Position and Velocity Vectors during Saccades," *Vision Research,* vol. 30, no. 1, pp. 111-127, 1990.

[141] G. Underwood, *Cognitive Processes in Eye Guidance.* Oxford Univ. Press, 2005.

[142] R. Valenti and T. Gevers, "Accurate Eye Center Location and Tracking Using Isophote Curvature," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2008.

[143] R. Vertegaal, I. Weevers, and C. Sohn, "GAZE-2: An Attentive Video Conferencing System," *Proc. CHI '02: Extended Abstracts on Human Factors in Computing Systems,* pp. 736-737, 2002.

[144] A. Villanueva, R. Cabeza, and S. Porta, "Eye Tracking: Pupil Orientation Geometrical Modeling," *Image and Vision Computing,* vol. 24, no. 7, pp. 663-679, July 2006.

[145] A. Villanueva, R. Cabeza, and S. Porta, "Gaze Tracking System Model Based on Physical Parameter," *Int'l J. Pattern Recognition and Artificial Intelligence,* 2007.

[146] A. Villanueva and R. Cabeza, "Models for Gaze Tracking Systems," *J. Image and Video Processing,* vol. 2007, no. 3, 2007.

[147] P. Viola and M. Jones, "Robust Real-Time Face Detection," *Proc. Int'l Conf. Computer Vision,* vol. II, p. 747, 2001.

[148] R. Wagner and H.L Galiana, "Evaluation of Three Template Matching Algorithms for Registering Images of the Eye," *IEEE Trans. Biomedical Eng.,* vol. 12, pp. 1313-1319, Dec. 1992.

[149] J. Waite and J.M. Vincent, "A Probabilistic Framework for Neural Network Facial Feature Location," *British Telecom Technology J.,* vol. 10, no. 3, pp. 20-29, 1992.

[150] J.G. Wang and E. Sung, "Gaze Determination via Images of Irises," *Image and Vision Computing,* vol. 19, no. 12, pp. 891-911, 2001.

[151] J.G. Wang, E. Sung, and R. Venkateswarlu, "Estimating the Eye Gaze from One Eye," *Computer Vision and Image Understanding,* vol. 98, no. 1, pp. 83-103, Apr. 2005.

[152] P. Wang, M.B. Green, Q. Ji, and J. Wayman, "Automatic Eye Detection and Its Validation," *Proc. 2005 IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 3, pp. 164-164, 2005.

[153] P. Wang and Q. Ji, "Learning Discriminant Features for Multi-View Face and Eye Detection," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,,* vol. 1, pp. 373-379, 2005.

[154] D.J. Ward and D.J.C. MacKay, "Fast Hands-Free Writing by Gaze Direction," *Nature,* vol. 418, no. 6900, p. 838, 2002.

[155] M. Wedel and R. Peiters, "Eye Fixations on Advertisements and Memory for Brands: A Model and Findings," *Marketing Science,* vol. 19, no. 4, pp. 297-312, 2000.

[156] K.P. White Jr., T.E. Hutchinson, and J.M. Carley, "Spatially Dynamic Calibration of an Eye-Tracking System," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 23, no. 4, pp. 1162-1168, July/Aug. 1993.

[157] O.M.C. Williams, A. Blake, and R. Cipolla, "Sparse and Semi-Supervised Visual Mapping with the s^3p," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 230-237, 2006.

[158] X. Xie, R. Sudhakar, and H. Zhuang, "On Improving Eye Feature-Extraction Using Deformable Templates," *Pattern Recognition,* vol. 27, no. 6, pp. 791-799, June 1994.

[159] X. Xie, R. Sudhakar, and H. Zhuang, "A Cascaded Scheme for Eye Tracking and Head Movement Compensation," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 28, no. 4, pp. 487-490, July 1998.

[160] L.Q. Xu, D. Machin, and P. Sheppard, "A Novel Approach to Real-Time Non-Intrusive Gaze Finding," *Proc. British Machine Vision Conf.,* 1998.

[161] H. Yamazoe, A. Utsumi, T. Yonezawa, and S. Abe, "Remote Gaze Estimation with a Single Camera Based on Facial-Feature Tracking without Special Calibration Actions," *Proc. 2008 Symp. Eye Tracking Research and Applications,* pp. 140-145, 2008.

[162] J. Yang, R. Stiefelhagen, U. Meier, and A. Waibel, "Real-Time Face and Facial Feature Tracking and Applications," *Proc. Conf. Auditory-Visual Speech Processing,* pp. 79-84, 1998.

[163] A. Yarbus, *Eye Movements and Vision.* Plenum Press, 1967.

[164] D.H. Yoo and M.J. Chung, "A Novel Non-Intrusive Eye Gaze Estimation Using Cross-Ratio under Large Head Motion," *Computer Vision and Image Understanding,* vol. 98, no. 1, pp. 25-51, Apr. 2005.

[165] D. Young, H. Tunley, and R. Samuels, "Specialised Hough Transform and Active Contour Methods for Real-Time Eye Tracking," Technical Report 386, School of Cognitive and Computing Sciences, Univ. of Sussex, 1995.

[166] A. Yuille, P. Hallinan, and D. Cohen, "Feature Extraction from Faces Using Deformable Templates," *Int'l J. Computer Vision,* vol. 8, no. 2, pp. 99-111, 1992.

[167] L. Zhang, "Estimation of Eye and Mouth Corner Point Positions in a Knowledge-Based Coding System," *Proc. SPIE,* pp. 21-18, 1996.

[168] Z. Zhu, K. Fujimura, and Q. Ji, "Real-Time Eye Detection and Tracking under Various Light Conditions," *Proc. Eye Tracking Research and Applications Symp.,* 2002.

[169] Z. Zhu, Q. Ji, and K. Fujimura, "Combining Kalman Filtering and Mean Shift for Real Time Eye Tracking" *Proc. Int'l Conf. Pattern Recognition,* vol. IV, pp. 318-321, 2002.

[170] Z. Zhu and Q. Ji, "Novel Eye Gaze Tracking Techniques under Natural Head Movement," *IEEE Trans. Biomedical Eng.,* vol. 54, no. 12, pp. 2246-2260, Dec. 2007.

[171] Z. Zhu, Q. Ji, and K.P. Bennett, "Nonlinear Eye Gaze Mapping Function Estimation via Support Vector Regression," *Proc. 18th Int'l Conf. Pattern Recognition,* vol. 1, pp. 1132-1135, 2006.

Dan Witzner Hansen received the PhD degree from the IT University of Copenhagen and is currently working as an associate professor in the Innovative Communications Group at that university. He has been a visiting scholar at Cambridge University and an assistant professor at the Technical University of Denmark. He has been supported by the IST EU Sixth framework programme COGAIN and the Danish Ministry of Research. His research interests are in computer vision, pattern recognition, and machine learning, with a bias toward vision-based HCI and eye tracking. He has served as a program committee member for several international conferences and workshops. He is a member of the IEEE.

Qiang Ji received the PhD degree in electrical engineering from the University of Washington. He is currently a professor with the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He is also a program director at the US National Science Foundation (NSF), managing part of the NSF's computer vision and machine learning programs. He has also held teaching and research positions with the Beckman Institute at the University of Illinois at Urbana-Champaign, the Robotics Institute at Carnegie Mellon University, the Department of Computer Science at the University of Nevada at Reno, and the US Air Force Research Laboratory. He currently serves as the director of the Intelligent Systems Laboratory (ISL) at RPI. His research interests are in computer vision, pattern recognition, and probabilistic graphical models. He has published more than 150 papers in peer-reviewed journals and conferences. His research has been supported by major governmental agencies including the US National Science Foundation (NSF), NIH, DARPA, ONR, ARO, and AFOSR, as well as by major companies including Honda and Boeing. He is an editor for several computer vision and pattern recognition related journals and he has served as a program committee member, area chair, and program chair for numerous international conferences/workshops. He is a senior member of the IEEE and a member of the IEEE Computer Society.