A Probabilistic Approach to Online Eye Gaze Tracking Without Personal Calibration

Jixu Chen, *Member, IEEE,* Qiang Ji, *Senior Member, IEEE,*

Abstract—Existing eye gaze tracking systems typically require an explicit personal calibration process in order to estimate certain person-specific eye parameters. For natural human computer interaction, such a personal calibration is often cumbersome and unnatural. In this paper, we propose a new probabilistic eye gaze tracking system without explicit personal calibration. Unlike the traditional eye gaze tracking methods, which estimate the eye parameter deterministically, our approach estimates the probability distributions of the eye parameter and eye gaze. By using an incremental learning framework, the subject doesn't need personal calibration before using the system. His/her eye parameter estimation and gaze estimation can be improved gradually when he/she is naturally interacting with the system. The experimental result shows that the proposed system can achieve less than three degrees accuracy for different people without calibration.

Index Terms—Gaze estimation, gaze calibration, dynamic Bayesian network.

1 Introduction

Gaze tracking is the procedure of determining the point-of-gaze on the monitor, or the visual axis of the eye in 3D space. Gaze tracking systems are primarily used in the Human Computer Interaction (HCI) and in the analysis of visual scanning patterns. In HCI, the eye gaze can serve as an advanced computer input [1] to replace traditional input devices such as a mouse pointer [2]. Also, the graphic display on the screen can be controlled by the eye gaze interactively [3]. Since visual scanning patterns are closely related to the attentional focus, cognitive scientists use the gaze tracking system to study human's cognitive processes [4], [5].

Numerous techniques [6], [7], [8], [9], [10], [11], [12], [13], [3] have been proposed to estimate the eye gaze. Earlier eye gaze trackers are fairly intrusive in that they require physical contacts with the user, such as the attachment of a number of electrodes around the eye [6]. In addition, most of these technologies also require the user's head to be motionless during eye tracking. Nowadays, gaze tracking technology based on video analysis of eye movements has been widely explored. Since it does not require physical contact with the user, video technology opens the most promising direction for building a non-intrusive eye gaze tracker. Various techniques [7], [14], [15], [8], [9], [10], [11], [12], [13], [16], [3] have been proposed to perform the eye gaze estimation based on eye images captured by video cameras. In general, these video-based eye gaze estimation algorithms can be classified into two groups: 2D mapping-based gaze estimation methods [7], [8], [3], [17] and 3D gaze estimation methods [12], [15], [9], [13], which estimate the 3D visual axis of the subjects. A survey of eye tracking techniques may be found in [18].

Recently, 3D methods are becoming more popular because of their high accuracy under free head movement. However, current advanced 3D gaze estimation systems require a calibration procedure for each subject in order to estimate his/her specific eye parameters.

In this work, we propose a novel method to estimate eye gaze without any explicit calibration procedure. In contrast to the traditional calibration procedure which asks the subject to fixate on several points on the screen, we estimate eye parameters and track the eye gaze when the subject naturally looking at the screen without prompting. Our method is based on exploiting the prior eye gaze distribution, which is estimated either from a saliency map of an image or from a generic Gaussian distribution. By combining the prior eye gaze distribution with 3D eye model, our method incrementally estimates the eye parameters and the eye gaze without any explicit personal calibration.

2 Related Work

The related works on video-based gaze estimation algorithms can be classified into two groups: 2D mapping based gaze estimation methods and 3D gaze estimation methods.

In traditional gaze estimation method, a 2D mapping approach learns a polynomial mapping function

- *J. Chen is with the Computer Vision Lab, GE Global Research Center, One Research Circle, KW-C410 Niskayuna, NY 12308. E-mail: chenji@ge.com.*
- *Q. Ji is with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590. E-mail: jiq@rpi.edu.*

from the 2D features (e.g. 2D pupil glint vector) [8], [11], [3], [17] or 2D eye images [7] to the gaze point on the screen. For example, the widely used Pupil Center Corneal Reflection(PCCR) technique [19], [8], [20], [21], [22] is based on the relative position between the centers of corneal reflection (glint) generated by the light source and the pupil. After the pupil and the glint are extracted from the image, the 2D pupil-glint vector is mapped to the gaze point on the screen by a polynomial mapping function.

However, the 2D mapping approach has two common drawbacks. First, in order to learn the person-specific mapping function, the user has to perform a complex experiment to calibrate the parameters of the mapping functions. For example, in the calibration procedure of [20], the subject needs to gaze at nine evenly distributed points on the screen, or twelve points for greater accuracy. Secondly, because the extracted 2D eye image features change significantly with head position, the gaze mapping function is very sensitive to head motion. Morimoto and Mimica [8] reported detailed data showing how the gaze tracking systems decay as the head moves away from the original calibration position. Hence, the user has to keep his head unnaturally still in order to achieve good performance. Methods have also been proposed to handle head pose changes using Neural Networks [3] or SVM [17]. These methods, however, either only consider the in-plane head translation [3] or need complex stereo cameras to obtain the 3D eye position [17].

In contrast, the 3D gaze estimation is based on high resolution stereo cameras[9], [12], [15], [13] or a single camera with multiple calibrated light sources [14] to estimate 3D eye features (e.g. the corneal center, the pupil center, and the optical axis connecting them) directly using the 3D reconstruction technique. The visual axis is estimated from the 3D features, and the gaze point on the screen is obtained by intersecting the visual axis with the screen. However, this type of method still needs person-specific calibration to estimate the eye parameters. For example, Chen et al. [13] proposed a 3D gaze estimation system with two cameras and one IR light on each camera. This method starts with the reconstruction of the optical axis of the eye. The visual axis can be estimated by adding a constant angle to the optical axis. However, the angle between the visual and optical axes needs to be estimated beforehand through a four-point personal calibration procedure. Guestrin et al. [15] proposed to estimate 3D gaze with two cameras and four IR lights. Their calibration procedure only required the subject to look at one point on the screen.

Most recently, some gaze estimation methods that don't use calibration have been suggested. Model and Eizenman [23] proposed to estimate the eye parameters by exploiting the binocular constraint that assumes that the visual axes of two eyes intersect on the screen. However, because of the noise in the measured optical axis, it is difficult to achieve accurate results. For a standard 40cm$\times$30cm flat monitor, when the optical noise has the error of one degree, this error will propagate and increase to five degrees in the visual axis. Although they propose the use of a larger monitor (160cm$\times$120cm) or a pyramid observation surface to reduce the error, these devices are often not available in real applications. The latest work by Maio, Chen, and Ji[24] improves Model and Eizenman's method by imposing constraints on the range of gaze and on that of the eye parameters. With these constraints, the accuracy and robustness of the method improves, therefore improving its practical utility. The accuracy of the method, however, remains low, compared with the gaze tracking methods that use the traditional personal calibration. In addition, the system requires two cameras to exploit the binocular constraint. The two camera system tends to limit the head movement.

Sugano et al. [25] offered a 2D appearance-based gaze estimation without calibration. They propose to learn a mapping function (Gaussian Process Regresser) between the eye image and the gaze point. In order to collect enough data to train this complex non-linear mapping function, they ask the subject to watch a 10-minute video. For each frame of the shown video, they extract its saliency map [26], which represents the distinctive image features attracting more attention. Finally, by treating the saliency map as the probability distribution of gaze, they generate the training gaze points by sampling from the saliency map. However, watching a movie for 10 minute for training is rather burdensome for the user. Furthermore, since they employ a 2D mapping method which doesn't consider head pose, the user has to fix their head on a chin rest.

In this paper, we propose an incremental probabilistic 3D gaze estimation method which allows free head movement and without explicit calibration. Our method is based on combining prior gaze distribution with 3D eye model. We propose two methods to estimate the gaze distribution: saliency map and Gaussian distribution. The former estimates gaze prior distribution by identifying the salient regions of an image, while the latter assumes gazes follows a Gaussian distribution with its means in the center of the screen. Given the estimated prior gaze distribution, we propose two methods to estimate the final gaze. First, unlike traditional 3D methods, which estimate eye parameter and gaze deterministically, the proposed method estimates the probability of eye parameter and eye gaze, and can better handle the uncertainty in the system. Second, we proposed an incremental learning method to improve estimation result gradually when the subject is naturally using the system. In our system, no explicit calibration process or calibration targets are used.

The experimental result shows that our system

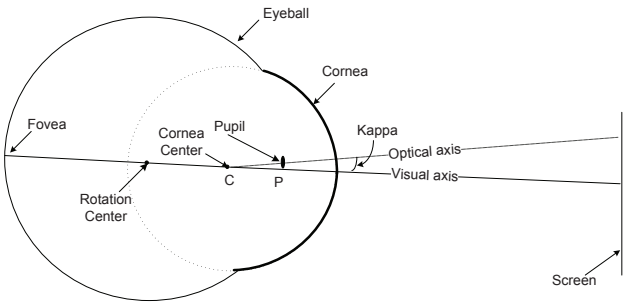


Fig. 1. Structure of the eyeball.

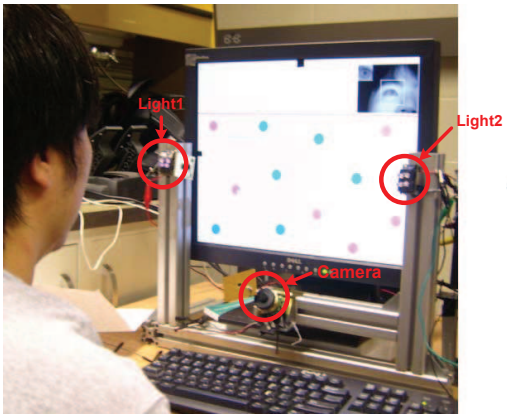


Fig. 2. System Overview. One camera, and two lights are installed on the aluminum framework

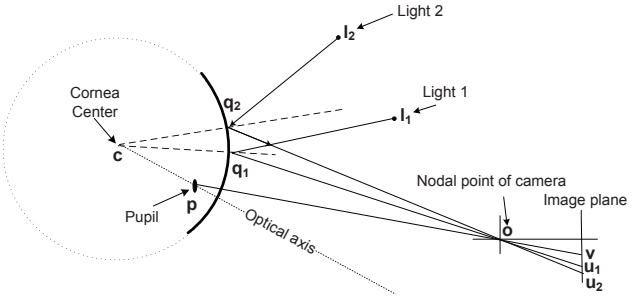


Fig. 3. Ray diagram of our system with one camera and two lights.

achieves less than three degrees average accuracy for different people.

Compared to Sugano's method [25], our method has the following advantages. First, we propose a systematic probabilistic framework to derive the analytic solutions to eye parameter and eye gaze, while their method is primarily numerical via sampling. Second, thanks to the incremental learning, our method does not need an explicit training procedure. It keeps improving the estimation as the user continues using the system. Third, since computing the saliency map is time consuming, we propose to use a more efficient Gaussian distribution as an alternative. Finally, because of the use of 3D gaze estimation method, our method is more accurate and allows free head movement.

3 3D Gaze Estimation

Before introducing our method, we briefly summarize the 3D gaze estimation techniques.

3.1 3D Eyeball structure

As shown in Figure 1, the eyeball is made up of the segments of two spheres of different sizes [27]. The smaller anterior segment is the cornea. The cornea is transparent, and the pupil is inside the cornea. The optical axis of the eye is defined as the 3D line connecting the center of the pupil ($\mathbf{p}$) and the center of the cornea ($\mathbf{c}$). The visual axis is the 3D line connecting the corneal center ($\mathbf{c}$) and the center of the fovea (i.e. the highest acuity region of the retina). Since the gaze point is defined as the intersection of the visual axis rather than the optical axis with the scene, the relationship between these two axes has to be modeled. The angle between the optical axis and visual axis is named *kappa* (κ), which is a constant value for each person. In traditional gaze estimation methods, κ is estimated through a personal calibration.

3.2 3D Gaze Estimation

Here, we implement the 3D gaze estimation system in [14], where the cornea center $\mathbf{c}$ and optical axis $\mathbf{o}$ are directly estimated from a single camera and two infrared lights, as shown in Figure 2. A simplified ray diagram of this system is shown in Fig. 3. Here, the cornea surface is modeled as a convex mirror with radius R. $\mathbf{l}_{1,2}$ are two IR lights and $\mathbf{q}_{1,2}$ are their reflections (glints) on the cornea surface. $\mathbf{p}$ is the pupil center and the distance between $\mathbf{p}$ and $\mathbf{c}$ is a constant K. Here, the positions of light and the camera parameters are fixed and estimated through a one-time system calibration. The two constant values R and K are fixed. We use their typical value in [14]. In gaze estimation, given the pupil $\mathbf{v}$ and glints $\mathbf{u}_{1,2}$ positions in the image, $\mathbf{p}$ and $\mathbf{c}$ and the 3D optical axis can be estimated directly by solving a system of equations. Because of the image noise, the noise on the final optical axis $\mathbf{o}$ estimation is about one degree. We refer the reader to [14] for details.

3.3 Personal Calibration

The estimated 3D optical axis can be represented by horizonal and vertical angles ($\mathbf{o} = (\theta, \varphi)$) as shown in Figure. 4. The unit vector of the optical axis is represented as:

$$\mathbf{v}_o = \begin{pmatrix} \cos(\varphi)\sin(\theta) \\ \sin(\varphi) \\ -\cos(\varphi)\cos(\theta) \end{pmatrix} \quad (1)$$

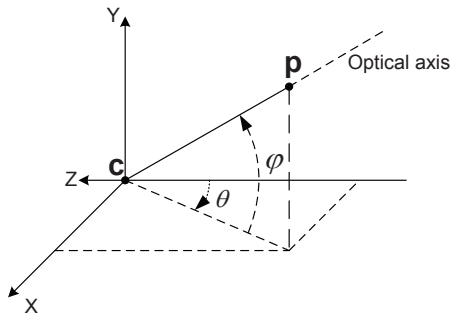


Fig. 4. Orientation of optical axis.

The subject's visual axis is estimated by adding $\kappa = (\alpha, \beta)$ to the optical axis:

$$\mathbf{v}_g = \begin{pmatrix} \cos(\varphi+\beta)\sin(\theta+\alpha) \\ \sin(\varphi+\beta) \\ -\cos(\varphi+\beta)\cos(\theta+\alpha) \end{pmatrix} \quad (2)$$

Finally, the gaze point $\mathbf{g}$ on the screen is estimated by intersecting $\mathbf{v}_g$ with the screen. Here, we use a coordinate frame affixed on the screen, with the screen plane as $Z = 0$, thus $\mathbf{g}$ can be written as $\mathbf{g} = (g_x, g_y, 0)^T$. This gaze point is determined by the optical axis and κ :

$$\begin{aligned} \mathbf{g} &= \mathbf{g}(\mathbf{o}, \kappa) = \mathbf{g}(\varphi, \theta, \alpha, \beta) \\ &= \mathbf{c} + k_c \cdot \begin{pmatrix} \cos(\varphi+\beta)\sin(\theta+\alpha) \\ \sin(\varphi+\beta) \\ -\cos(\varphi+\beta)\cos(\theta+\alpha) \end{pmatrix}. \end{aligned} \quad (3)$$

, where $\mathbf{c}$ is the cornea center. Because the z-component of $\mathbf{g}$ equals 0, the value of k_c is :

$$k_c = \frac{c_z}{\cos(\varphi+\beta)\cos(\theta+\alpha)}. \quad (4)$$

However, because κ varies for different subjects, it needs to be estimated beforehand through calibration. In traditional methods [14], [13], [15], the subject is asked to look at N specific calibration points on the screen: $\mathbf{g}_i^*, i = 1, .., N$. The eye parameter can then be estimated by minimizing the distance between the estimated gaze points and these ground-truth gaze points:

$$\kappa^* = \arg\min_{\kappa} \sum_i \|\mathbf{g}_i^* - \mathbf{g}(\mathbf{o}_i^*, \kappa)\| \quad (5)$$

where $\mathbf{o}_i^*$ is the estimated optical axis when subject is looking at the ith gaze point $\mathbf{g}_i^*$. The traditional gaze estimation method can be represented as Figure 5. During calibration, the eye parameter κ^* is estimated from the calibration gaze point $\mathbf{g}^*$ and the predicted optical axis $\mathbf{o}^*$. During gaze estimation, the eye parameter κ^* is fixed, and a new optical axis $\mathbf{o}$ is estimated from the camera. The gaze point is determined by $\mathbf{o}$ and κ^* through Eq. 3.

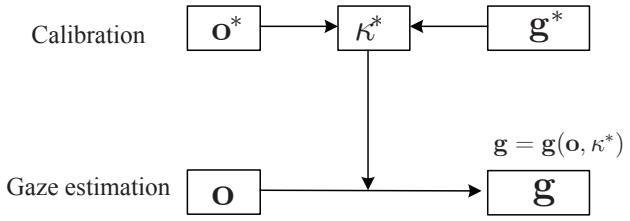


Fig. 5. Diagram of traditional 3D gaze estimation. where $\mathbf{g}^*$ is ground-truth gaze.

4 Probabilistic Gaze Estimation

In the traditional methods, in order to acquire the ground-truth gaze points to estimate κ, the subject has to look at some specific points. This procedure is often cumbersome and unnatural. Here, we propose a new framework to estimate the probability of κ and eye gaze without requiring the subject to look at specific calibration points.

4.1 Proposed Probabilistic Framework

The basic idea is to replace the need of looking at specific gaze points on the screen with a probability distribution of the gaze. Given this idea, the subject naturally interact with the images on the screen (in full-screen mode), while his/her personal parameters are implicitly estimated. We introduce two methods to estimate the gaze probability distribution.

4.1.1 Gaze probability distribution from the saliency map

First, we utilized the method in [26] to estimate the saliency map of each image, which represents the distinctive features in the image. Some examples of saliency maps are shown in Figure 6. The experimental results in [26] shows remarkable consistency between the saliency map and the gaze. Thus, given the image I on the screen, the gaze distribution can be represented as the conditional probability of the gaze position $p(\mathbf{g}|I)$. Here, the only assumption we have is that the user has a higher probability of looking at the salient regions of the image.

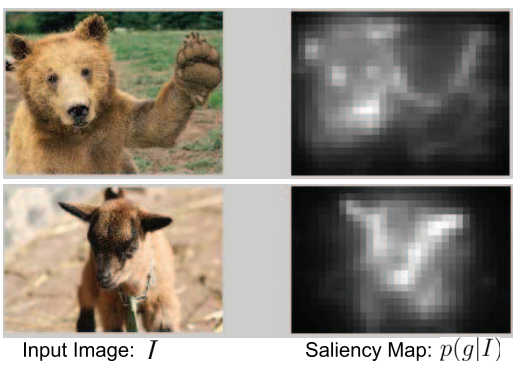


Fig. 6. Examples of saliency map ($p(g|I)$)

4.1.2 Gaussian Gaze Distribution

In the above section, we approximate the gaze probability distribution with the saliency map. The saliency map was extracted from the image shown to the user. While the saliency map can effectively approximate the gaze distribution, computing the saliency map for each image is time consuming. To alleviate this limitation, in this section, we further extend our method to more general scenarios where the subject is watching a video or movie. Under such scenarios, we relax the need of computing saliency map. Instead, we assume the subject is naturally watching the computer screen, and that most of the gazes (fixations) are concentrated on the center of the screen, with peripheral vision on the margins of the screen. This assumption simply means the probability of a gaze is located near the center is higher than away from the screen center. This assumption is quite weak, but it works well when people are watching videos or movies because the movie cameraman usually capture the videos with objects of interest in the center. This assumption, however, may not work well for window desktop user or for a person who surfs the Internet. To empirically verify this, figure 7 shows the examples of gaze distribution when different subjects are naturally watching a randomly selected movie for 5 minutes. It is clear that most gazes are focused on the center region, while much fewer gaze fixations are in other regions.

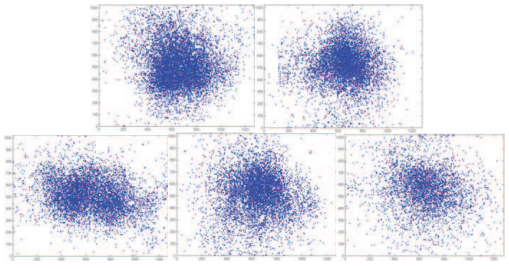


Fig. 7. Examples of natural gaze distribution when people watch a video

Given this understanding, we can characterize the gaze probability distribution as a simple Gaussian distribution $\mathcal{N}(\mu, \mathbf{\Sigma})$ (Figure 8). Its mean is located in the center of the display ($x = 640, y = 512$), and its variances can be empirically estimated based on historical data.

Thus, the gaze probability can be either computed from the saliency map $p(g) = p(g|I)$ or from the assumption of Gaussian gaze distribution, i.e, $p(g) = \mathcal{N}(\mu, \mathbf{\Sigma})$, where we have omitted the image "I" from $p(g)$ since $p(g)$ is the same for all images.

4.1.3 Probabilistic Gaze Estimation

Based on this gaze probability, we propose the new gaze estimation framework shown in Figure 9. Notice

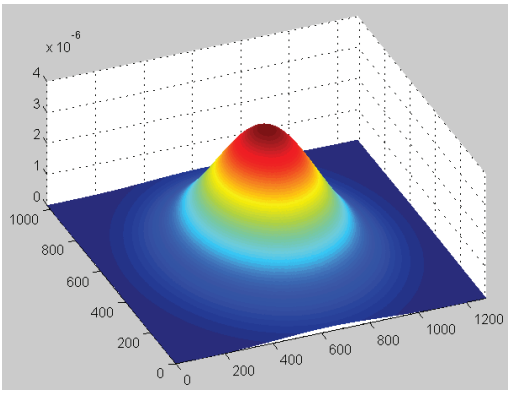


Fig. 8. The Gaussian distribution as the prior distribution of the gaze. The display size is 1280×1024

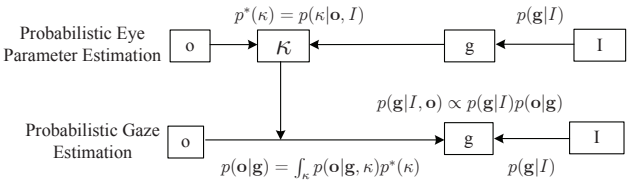


Fig. 9. Diagram of the probabilistic gaze estimation.

the differences between our method and the traditional method in Figure 5 :

1) Firstly, the traditional method needs to collect the ground-truth gaze ($\mathbf{g}^*$), when the subject is looking at specific points during calibration, while our method only needs the gaze probability $p(\mathbf{g}|I)$, when the subject is looking at the image I.
2) Secondly, the traditional method estimates the eye parameter κ^* deterministically. However, without ground-truth gaze, we cannot deterministically estimate the value of κ^*. Instead, we estimate κ^* probabilistically through the probability distribution of its measurement κ, i.e., $p^*(\kappa)$. Notice that the groundtruth value of κ^* is a constant, but its measurement κ is a random variable following $p^*(\kappa)$ [1]. The final distribution of $p^*(\kappa)$ will have one peak, which represents our estimate of the true κ^*.
3) Thirdly, during gaze estimation, the traditional method estimates gaze only from the optical axis and κ^*, while our method first estimates the gaze likelihood $p(\mathbf{o}|\mathbf{g})$ from the optical axis and $p^*(\kappa)$, then combines it with the gaze's prior probability $p(\mathbf{g}|I)$ (e.g. from the saliency map) to estimate gaze posterior probability.

This framework is mainly composed of two parts: *probabilistic eye parameter estimation* and *probabilistic gaze estimation*. We discuss them separately in the following two sections.

1. The measurement κ follows conditional probability $p^*(\kappa|\kappa^*)$. Because κ^* is a constant, we use $p^*(\kappa)$ for notational clarity.

4.2 Probabilistic Eye Parameter Estimation

In this section, we discuss the method to estimate eye parameter (κ) probability from gaze probability (e.g the saliency map). Firstly, we introduce a general graphical model to represent the relationships between the shown image (I), eye gaze ($\mathbf{g}$), optical axis ($\mathbf{o}$), and the eye parameters (κ).

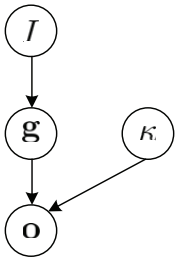


Fig. 10. Probabilistic relationships in BN.

Figure 10 is the Bayesian Network (BN) [28] that represents the probabilistic relationships. The nodes in the BN represent random variables, and the links represent the conditional probability distributions (CPDs) of nodes given their parents. Based on the gaze probability map and the eye model, we define the CPDs as follows:

1) $p(\mathbf{g}|I)$: $\mathbf{g}$ is a two dimensional vector $\mathbf{g} = (x, y)$, which represents the location of the gaze on the screen (According to the resolution of the monitor, the gaze position is discretized in the range: $0 < x < 1280, 0 < y < 1024$). The link $I \rightarrow \mathbf{g}$ is quantified by $p(\mathbf{g}|I)$ which is the gaze probability distribution estimated from image.
2) $p(\mathbf{o}|\mathbf{g}, \kappa)$: $\mathbf{o}$ has two parents $\mathbf{g}$ and κ. As discussed above, the camera in a gaze system cannot directly observe the visual axis and the gaze. It can only observe the optical axis ($\mathbf{o}$) as the measurement of gaze ($\mathbf{g}$). In the traditional method, $\mathbf{o}$ is a deterministic function of $\mathbf{g}$ by subtracting a constant bias κ, ignoring any uncertainties. In our proposed method, considering the noise in the gaze system, we model the conditional probability as a Gaussian distribution:

$$p(\mathbf{o}|\mathbf{g}, \kappa) = \mathcal{N}(f(\mathbf{g}, \kappa), \Sigma) \tag{6}$$

where $\mathbf{o} = (\theta, \varphi)^T$ is a 2-D vector. $f(\mathbf{g}, \kappa)$ is the inverse function of Eq. 3, which estimates the optical axis by subtracting κ from the visual axis. Based on Eq. 3, the directional vector of visual axis can be computed as $\mathbf{d} = \mathbf{g} - \mathbf{c}$, and the horizontal and vertical angles of the visual axis are $\arctan(d_x/d_z)$ and $\arctan(d_y/\sqrt{(d_x^2 + d_z^2)})$ respectively. Finally, the optical axis is $\mathbf{o} = (\arctan(d_x/d_z) - \alpha, \arctan(d_y/\sqrt{(d_x^2 + d_z^2)}) - \beta)^T$. Σ models the noise of the optical axis which is estimated from 3D gaze estimation system in [14] which is discussed in section 3.2. (According to previous tests of this system, we set the standard deviation of the optical axis as one degree on both θ and φ.)

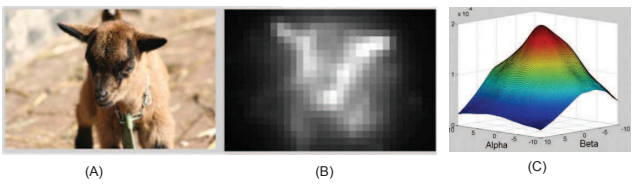
Now, based on the BN model, eye parameter estimation is solved as an inference problem in the BN, which estimates the posterior probability $p(\kappa|\mathbf{o}, I)$ given the optical axis and the shown image. Based on the conditional independencies in the BN model in Figure 10, the probability of κ can be written as:

$$\begin{aligned} p(\kappa|\mathbf{o}, I) &= \textstyle\int_{\mathbf{g}} p(\kappa, \mathbf{g}|\mathbf{o}, I) = \int_{\mathbf{g}} \frac{p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g},\kappa)\cdot p(\kappa)\cdot p(I)}{p(\mathbf{o},I)} \\ &\propto \textstyle\int_{\mathbf{g}} p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g}, \kappa) \end{aligned} \tag{7}$$

$p(\mathbf{g}|I)$ is the gaze probability map; $p(\mathbf{o}|\mathbf{g}, \kappa)$ is the Gaussian distribution as defined in Eq.6. Notice that, the prior probability $p(\kappa)$ is initially assumed to be a uniform distribution, thus $p(\kappa)$ is a constant. $p(I)$ and $p(\mathbf{o}, I)$ are constant because I and $\mathbf{o}$ are given. Here, Eq.7 is a one-step belief propagation that propagates the probability from the gaze to κ given one optical axis. The gaze position is discrete in limited range; thus the integral in the above equation can be approximated by summation.

Figure 11(C) shows an example of the estimated eye parameter probability. Here, we collected 40 optical axes when the subject was looking at the image in 11(A). i.e., the training optical axes are $\mathbf{o}_{1,..,40}$ and their corresponding shown images $I_{1,..,40}$ are the same. Assuming these optical axes $\mathbf{o}_{1,..,40}$ are conditionally independent to each other, we can estimate the κ probability as the product of each single probability:

$$p^*(\kappa) = p(\kappa|\mathbf{o}_{1,..,40}, I_{1,..,40}) \propto \prod_{i=1}^{40} \int_{\mathbf{g}_i} p(\mathbf{g}_i|I_i)p(\mathbf{o}_i|\mathbf{g}_i, \kappa) \tag{8}$$



(A) (B) (C)

Fig. 11. Probabilistic Eye Parameter Estimation. (A) is the shown image. (B) is the gaze probability map $p(\mathbf{g}|I)$ of the image estimated from the saliency map. (C) is the estimated probability distribution of eye parameter $p^*(\kappa)$. (x-axis represents α and y-axis represents β.)

Based on the biological study, eye parameters should be in a limited range for normal eyes. Here we restricted the eye parameter in the range $-10^o < \alpha < 10^o$ and $-10^o < \beta < 10^o$.

4.3 Probabilistic Gaze Estimation

Given the estimated eye parameter probability $p^*(\kappa)$, we can estimate the gaze probability. For consistency,

this derivation is based on the same BN model in Figure 10. Unlike the eye parameter estimation, the estimated $p^*(\kappa)$ is now used as the prior probability of the κ node. Then, the probability of the gaze, given the optical axis and the shown image, can be written as:

$$p(\mathbf{g}|\mathbf{o}, I) \propto p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g}) \tag{9}$$

where $p(\mathbf{g}|I)$ is the prior probability of gaze from either the saliency map of the shown image I or the Gaussian gaze distribution, and $p(\mathbf{o}|\mathbf{g})$ is the gaze likelihood, which can be derived from $p^*(\kappa)$ as:

$$p(\mathbf{o}|\mathbf{g}) = \int_{\kappa} p(\mathbf{o}|\mathbf{g}, \kappa)p^*(\kappa) \tag{10}$$

Note that all the derivations above are only valid based on the conditional independencies in the BN model.

Thus, the probabilistic gaze estimation is composed of the following steps:

1) First, we estimate the gaze prior probability distribution $p(\mathbf{g}|I)$ either from the saliency map or from the Gaussian gaze distribution.
2) Then, we estimate the likelihood gaze map $p(\mathbf{o}|\mathbf{g})$, given the current optical axis and the eye parameter prior $p^*(\kappa)$, based on Eq. 10.
3) Finally, the product $p(\mathbf{g}|I)p(\mathbf{o}|\mathbf{g})$ represents the gaze posterior probability map $p(\mathbf{g}|\mathbf{o}, I)$. Given the posterior probability, the maximum posterior point is selected as the gaze point $\mathbf{g}^*$:

$$\mathbf{g}^* = \arg\max_{\mathbf{g}} p(\mathbf{g}|\mathbf{o}, I) \tag{11}$$

The results of the three steps are shown in Figure. 12. Here we compare our method with the traditional gaze estimation method, which uses 9-point calibration to calibration the eye parameter. The traditional method can achieve one degree of accuracy. The peak in our posterior probability map is very close to the estimated gaze of the traditional method as shown in Figure 12, but our method does not need any explicit calibration.

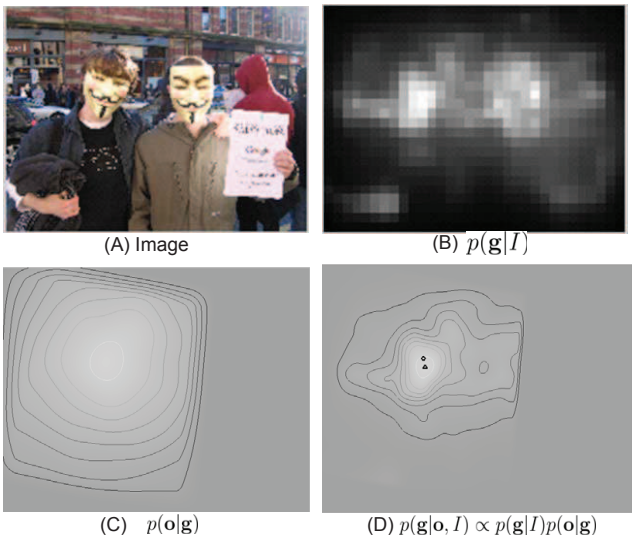


Fig. 12. Probabilistic Gaze Estimation. (A) is the shown image. (B) is the saliency map $p(\mathbf{g}|I)$ of the image. (C) is the gaze likelihood map given the optical axis. (D) is the gaze posterior probability map. The triangle shows the maximum posterior point. The circle shows the estimated gaze using the traditional method.

5 Incremental Learning for Gaze Estimation

The above probabilistic framework includes two stages. First, $p^*(\kappa)$ is estimated when the subject is looking at the training images. Then, his/her gaze is estimated when he/she is looking at the test images.

In order to provide a more natural user experience, we propose an incremental learning algorithm for our probabilistic framework. This new framework does not need any prior training. It can quickly adapt to the user, and incrementally improves gaze estimation accuracy as the subject uses the system.

We first assume the initial distribution of κ as uniform. When the subject starts to use the system, we record a sequence of his optical axes $\mathbf{o}_{t,..,1}$. Given the corresponding shown image sequence $I_{t,..,1}$, the incremental learning framework continually updates the estimations of κ and gaze given all previous information, i.e. estimating $p(\kappa_t|I_{t,..,1}, \mathbf{o}_{t,..,1})$ and $p(\mathbf{g}_t|I_{t,..,1}, \mathbf{o}_{t,..,1})$. We employ a recursive updating procedure detailed as follows.

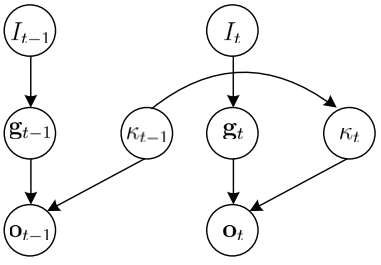


Fig. 13. DBN for incremental learning.

For incremental learning, we first extend the BN to a dynamic BN (DBN) model as shown in Figure 13. In general, a DBN is comprised of interconnected time slices of static BNs. One important assumption of DBN is first-order Markovian, i.e., given the state of the closest previous time slice, the current time slice is independent from other past time slices. Thus, in Figure 13, we only show the DBN of the current and previous time slices. It includes two kinds of links. *Intra-frame links* in the current time slice are the same as the BN model we set before and *inter-frame link* from κ_{t-1} to κ_t captures the temporal relationships. Base on the anatomy, κ cannot vary much over time.

Thus, we model it as a Gaussian distribution:

$$p(\kappa_t|\kappa_{t-1}) = \mathcal{N}(\kappa_{t-1}, \Sigma_k) \tag{12}$$

where Σ_k is the covariance matrix which allows κ_t to vary in a small range around the previous estimation κ_{t-1}. It depends on the uncertainty in our system. Here we empirically set the standard deviations of α_t and β_t to one degree , i.e. Σ_k is an identity matrix.

Given the above temporal relationship in the DBN, the probability of κ can be updated recursively. Firstly, we predicted the prior probability of the current κ_t based on its previous probability, as shown in Eq. 13.

$$\begin{aligned} &p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) \\ &= \int_{\kappa_{t-1}} p(\kappa_t|\kappa_{t-1}) p(\kappa_{t-1}|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) \end{aligned} \tag{13}$$

where $p(\kappa_{t-1}|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$ is the κ probability from the previous time frame. Since the temporal CPD $p(\kappa_t|\kappa_{t-1})$ is a Gaussian distribution, this integral is implemented by a convolution of previous κ probability map with a Gaussian kernel. (In the first time frame, when no prior information of κ is available, we assume $p(\kappa_1)$ is uniformly distributed.)

Based on the predicted temporal prior probability $p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$, the current probability of $\mathbf{g}_t$ and κ_t can be derived as the filtering problem in the DBN :

$$\begin{aligned} &p(\mathbf{g}_t|I_{t,t-1,..,1}, \mathbf{o}_{t,t-1,..,1}) \\ &\propto p(\mathbf{g}_t|I_t) \cdot \int_{\kappa_t} p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) \end{aligned} \tag{14}$$

$$\begin{aligned} &p(\kappa_t|I_{t,t-1,..,1}, \mathbf{o}_{t,t-1,..,1}) \\ &\propto \int_{\mathbf{g}_t} p(\mathbf{g}_t|I_t) p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) \cdot p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1}) \end{aligned} \tag{15}$$

The current estimation of $p(\kappa_t|I_{t,t-1,..,1}, \mathbf{o}_{t,t-1,..,1})$ is updated recursively from its previous estimation $p(\kappa_{t-1}|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$, based on Eq.13 and Eq. 15.

Letting $p^*(\kappa_t) = p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$ and $p'(\kappa_t) = p(\kappa_t|I_{t,..,1}, \mathbf{o}_{t,..,1})$, the above incremental learning algorithm can be summarized in Algorithm 1.

Compared with the probabilistic framework in Figure 9, the diagram of this incremental learning algorithm is shown as Figure 14.

An example of the incremental learning of $p'(\kappa_t)$ is shown in Figure 15. The estimated $p'(\kappa_1)$ for the first time frame has a high probability in multiple regions. By updating its probability incrementally, it gradually converges to a single peak after twenty time frames.

The estimation of κ is shown in Figure 16. Note that our algorithm used the whole κ probability map rather than a single point. Here, we show the maximum point in the probability map (Fig.15) as our κ estimation. $\kappa = (\alpha, \beta)$ includes two parameters which are shown in separate figures.

Note the relationship between the BN and the DBN models. The only difference between the DBN and the BN is that the DBN considers the temporal prior of κ_t, and continues updating it over time. For example,

Algorithm 1 Incremental Gaze Estimation Algorithm

$t \leftarrow 1$
Set $p^*(\kappa_1)$ as uniform distribution.

Estimate the first gaze:
$p(\mathbf{g}_1|I_1, \mathbf{o}_1) \propto p(\mathbf{g}_1|I_1) \cdot \int_{\kappa_1} p(\mathbf{o}_1|\mathbf{g}_1, \kappa_1) p^*(\kappa_1)$

Update κ probability :
$p'(\kappa_1) \propto \int_{\mathbf{g}_1} p(\mathbf{g}_1|I_1) p(\mathbf{o}_1|\mathbf{g}_1, \kappa_1)$
loop
$t \leftarrow t + 1$
Temporal belief propagation $p'(\kappa_{t-1}) \rightarrow p^*(\kappa_t)$:
$p^*(\kappa_t) = \int_{\kappa_{t-1}} p(\kappa_t|\kappa_{t-1}) p'(\kappa_{t-1})$
Probabilistic gaze estimation:
$p(\mathbf{g}_t|I_{t,..,1}, \mathbf{o}_{t,..,1}) \propto p(\mathbf{g}_t|I_t) \cdot \int_{\kappa_t} p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) p^*(\kappa_t)$

Update κ probability:
$p'(\kappa_t) \propto \int_{\mathbf{g}_t} p(\mathbf{g}_t|I_t) p(\mathbf{o}_t|\mathbf{g}_t, \kappa_t) \cdot p^*(\kappa_t)$
end loop

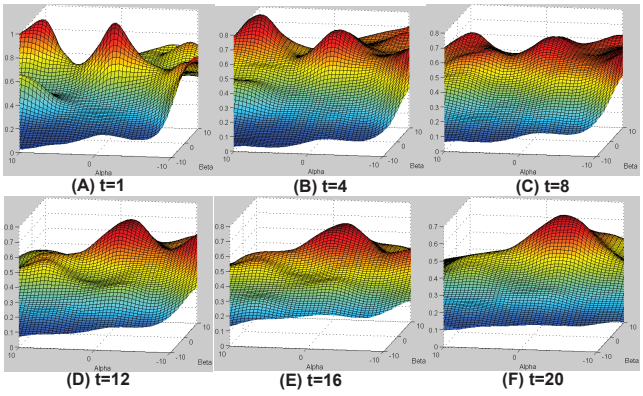
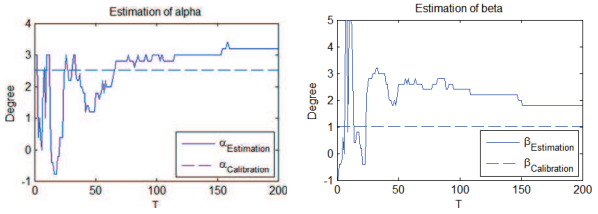


Fig. 15. Incremental learning of $p'(\kappa_t)$.



(a) Estimation of alpha. (b) Estimation of beta.

Fig. 16. Estimation of κ in incremental learning. The estimated α and β are shown as solid lines, and their ground-truth value from calibration are shown as dashed lines. In the beginning, this estimation oscillates significantly because the probability map hasn't converged, and it includes several peaks (Fig.15). Finally, the estimate α converges to the groundtruth values.

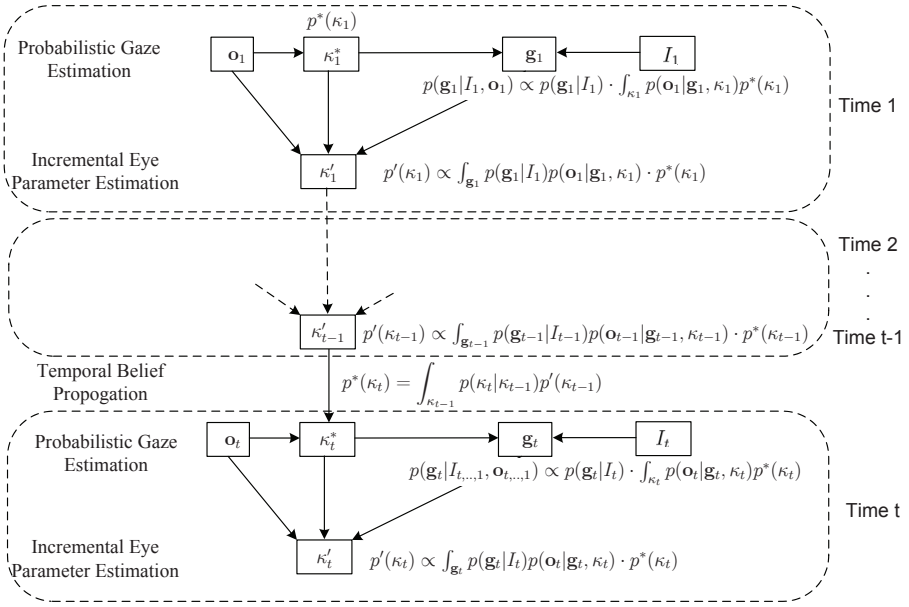


Fig. 14. Diagram of incremental probabilistic gaze estimation.

if letting $p^*(\kappa) = p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$, Eq. 14 is the same as Eq. 9; if letting $p(\kappa_t|I_{t-1,..,1}, \mathbf{o}_{t-1,..,1})$ be uniform distribution, Eq. 15 is the same as Eq. 7.

6 Experimental Results

We evaluate our system when the subject is looking at a standard 19-inch monitor (37.63cm×30.11cm). Our system allows free head movement, the range of the distance between the monitor and the subjects' eyes is about 45−70cm. To evaluate the traditional gaze estimation method, the subjects are often asked to look at some points on the screen. The gaze estimation error can be computed as the distance between these points and the estimated gaze points. In our method, the user does not need to look at any specific points. To evaluate our system, we implemented the traditional 3D gaze estimation system [14]. This system is first calibrated by asking the subject to look at nine points on the screen. The average accuracy of this system is one degree for different subjects. We compared our proposed method with this system.

To evaluate our method, we collected the optical axes of five subjects while they viewed the images on the screen. Each image displayed for about four seconds on the screen. Our gaze system collects eighty optical axes for each image (our gaze system captured the video of the eye and estimated the optical axes at 20 frames per second). Because the relationship between the saliency and gaze cannot be guaranteed during saccadic eye movements, we remove the saccadic movement in two steps. First, when the subject is looking at images, when he/she switches the image, we filter out the data in the beginning ($<$ 300ms) since we believe most of the gaze movements in the beginning are saccadic movements. Second, as the study continues, most of eye gaze movements are fixations and we filter out some very short gaze fixations (less than 100ms) and treat them as saccadic eye movements. Our study shows that less than 10% movements are saccadic eye movements after the initial stage.

To show the advantages of the incremental learning, we compared the incremental learning algorithm (Section 5) to the batch training method in Section.4.2.

6.1 Batch Training for Gaze Estimation

For batch training, we divided the eighty time frames when the subject viewed one image into training data (forty frames) and testing data (forty frames). Each subject viewed five images in this experiment. Please notice that we only use images with clear salient objects in our experiment. The saliency entropy is used as the criteria to select images from Google image search. Specifically, only images with entropy lower than 13 are selected in our experiment. These images usually have some clear salient objects as shown in Figure 17. For comparison, we also selected some poor images with high saliency entropy to study the impact of poor saliency map in section 6.3.

We used leave-one-image-out cross validation, i.e. when testing on forty frames of one image, we first learned the eye parameter probability from the training data of the other four images (Section.4.2). Saliency map is used to approximate the gaze prior distribution.

For a more effective system, we wanted to use less training data, because more training time may make the subject bored and easily distracted. We tested the

dependency of our method on the amount of the training data by using 160, 80, 40, and 20 frames of training data, i.e. 40, 20, 10, and 5 frames for each training image.

The average error and the standard deviation of error (over 200 test frames) are shown in Table 1. When the training frames were reduced, the average error did not increase much, but the standard deviation increased significantly because the eye parameter map did not converge yet. The average gaze estimation error of our proposed method achieved 2.40^o when there was enough training data (160 frames, about eight seconds), and remained low (2.5^o) with only 20 frames (less than one second).

6.2 Incremental Learning for Gaze Estimation

Based on our incremental learning algorithm, the system doesn't need to estimate the eye parameter probability beforehand using training frames. This system can automatically update the eye parameter probability and estimate the gaze when the subject starts using the system. Again, saliency map is used to approximate the gaze prior distribution.

The gaze estimation error and the standard deviation for the first 20, 40, 80, 120, 160, and 200 frames are shown in Table. 2. Although the error is large for the first few frames (<20 frames), it decreases quickly as the subject uses the systems. Compared with the batch training, the incremental learning achieves similar performance for the first twenty frames. However, when the subject is using the system, incremental learning continues improving the performance and can achieve an average accuracy of 17.07mm (1.77^o) for the first 200 frames. This process is done automatically, naturally, and without any user knowledge. This result outperforms the batch method in Section 6.1 because of the gaze temporal continuity. Furthermore, in the incremental learning experiment, training and testing frames are collectd when a subject is viewing the same set of images, while in the batch training experiment we use leave-one-image-out cross validation. The estimator's performance is always better than when testing the estimator using a data that is not part of the training data.

Some gaze estimation results (in both the original image and the saliency map) of subject 1 are shown in Figure 17. Without calibration, the results of our method are close to the results of the system with 9-point calibration. The subject may look at some region with low saliency, such as the white paper in the person's hand in Figure 17(A). In this case, by incrementally improving the eye parameter estimation and by combining gaze likelihood with the saliency map, our method can still follow the true gaze positions.

Please notice that both batch and incremental algorithm depends on the gaze prior probability (saliency map) $p(\mathbf{g}|I)$ in two aspects. First, eye parameter estimation depends on the integral of this prior probability as shown in Equation 15. Second, gaze posterior probability is estimated as the production of this prior and the gaze likelihood as shown in Equation 14. Prior probability is necessary in eye parameter estimation step, but it may not be necessary for gaze estimation step if the eye parameter is already accurately estimated so that the gaze likelihood itself is sufficient. Figure 18 shows the average error of posterior gaze estimation and the error of the gaze estimation with likelihood only, e.g., changing the prior in Eq.14 to uniform distribution. We can see that proposed gaze prior plays an important role in correcting the gaze estimation errors in the initial 150 frames but its role gradually diminishes as the frames go on. In fact, after 160 frames, the gaze likelihood achieves the same level of accuracy as the posterior estimation. This demonstrates that the posterior estimation is beneficial in the beginning when the eye parameter estimation has not converged (as shown in Fig. 15),

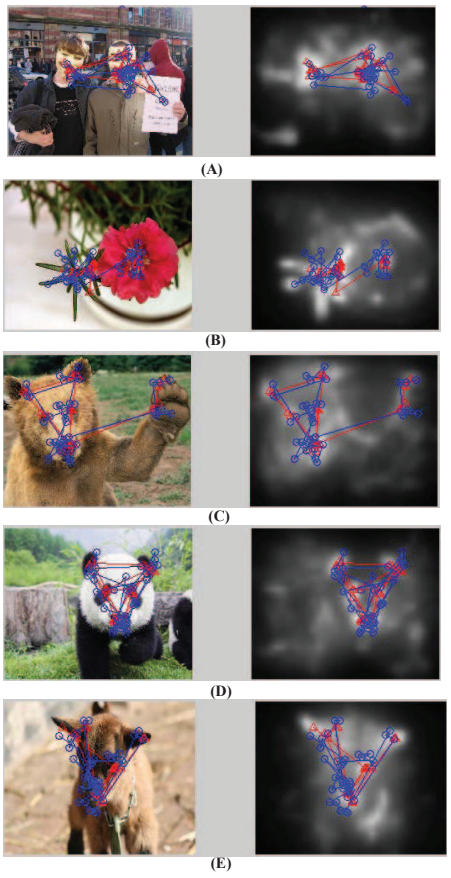


Fig. 17. Probabilistic Gaze Estimation Result. The red rectangles are the results of our proposed method. The blue circles are the results of the traditional method with 9-point calibration. Images in the left column include the gazes superimposed on the original image, while the right column includes the gazes superimposed on the saliency map.

TABLE 1
Gaze estimation error of five subjects with different training data size. Eye parameters are trained though batch training.

Training data size	20 frames		40 frames		80 frames		160 frames	
	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]	[mm]	[deg.]
Subject 1	23.6	2.45	23.4	2.43	23.1	2.40	21.0	2.18
Subject 2	19.0	1.98	17.6	1.83	18.5	1.92	18.6	1.93
Subject 3	16.4	1.70	16.0	1.66	15.8	1.64	15.5	1.61
Subject 4	28.4	2.95	28.1	2.93	26.7	2.77	27.8	2.89
Subject 5	33.0	3.43	32.6	3.39	32.2	3.34	32.7	3.40
Average	24.0	2.50	23.5	2.45	23.2	2.41	23.1	2.40
Std.	16.1	1.67	12.6	1.31	6.7	0.70	5.6	0.58

TABLE 2
Gaze estimation results of five subjects for the first N frames (N=10,20,40,80,120,160,200). Eye parameters are automatically updated after each frame.

Training data size	20 frames		40 frames		80 frames		120 frames		160 frames		200 frames	
	mm	deg.	mm	deg.	mm	deg.	mm	deg.	mm	deg.	mm	deg.
Subject 1	16.7	1.73	19.3	2.01	19.9	2.07	19.6	2.04	18.2	1.89	17.3	1.80
Subject 2	24.0	2.50	19.8	2.06	18.3	1.90	17.7	1.84	18.2	1.89	17.3	1.80
Subject 3	15.4	1.60	15.2	1.57	14.9	1.54	14.5	1.50	15.4	1.59	15.4	1.59
Subject 4	39.4	4.09	28.4	2.95	21.2	2.20	18.4	1.91	19.9	2.06	19.9	2.07
Subject 5	28.3	2.95	27.0	2.81	20.1	2.09	16.2	1.68	15.4	1.60	15.4	1.60
Average	**24.8**	**2.57**	**21.9**	**2.28**	**18.9**	**1.96**	**17.3**	**1.80**	**17.4**	**1.81**	**17.0**	**1.77**
Std.	16.7	1.72	10.1	1.04	6.6	0.69	5.5	0.57	4.5	0.46	4.3	0.44

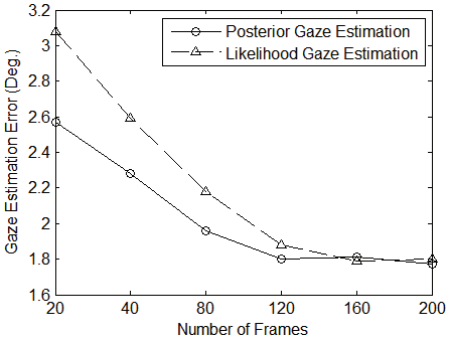


Fig. 18. Comparison of gaze estimation with posterior probability and gaze estimation with likelihood only.

but the two methods are asymptotically equivalent when there is enough training data.

Compared to the most recent calibration-free gaze estimation method [25], which asks the subject to watch a ten-minute video for training and achieves an accuracy of six degrees, our proposed method doesn't need training data beforehand and can adapt to the user very quickly (in 80 frames or less than four seconds), and continues to improve the accuracy as person uses it. The average accuracy can achieve 1.77 degrees. Furthermore, in our 3D gaze estimation framework, the subject can have natural head movement without fixing his/her head in a chin-rest.

6.3 Gaze Estimation with Low-quality Saliency Map

Above saliency-based gaze estimation depends on the quality of saliency map. Here, we consider two cases of low-quality saliency map.

First, we tested our incremental learning algorithm when the subjects are looking at images without salient objects. As discussed in section 6.1, we selected 100 bad images with high entropy from Google image search. Some example images are shown in Figure 19. In this case, the high-salient regions are evenly distributed in the images. We randomly select 5 images out of 100 bad images in our experiment. The gaze estimation error is summarized in Table 3. Compared to Table 2, the error increases significantly because the salient map cannot provide good prior information for these images.

Second, we consider the case when the image includes salient objects but the saliency estimation algorithm cannot predict an accurate saliency map. To simulate the saliency estimation error, we add noise to the saliency map. For each pixel in the saliency map, we add a uniform noise $\varepsilon = \mathcal{U}(0, \sigma)$. The saliency map with noise level $\sigma = 0.8, 1.6, 2.4$ are shown in Figure 20. The average gaze estimation error are shown in Table 4. When the noise is large, the saliency region is ambiguous in the map and the gaze error increase significantly. The above experiments show that the success of saliency-based gaze estimation highly depends on the quality of saliency map.

6.4 Probabilistic Gaze Estimation with Gaussian Gaze Distribution

In this section, we study the performance of gaze estimation without using saliency map but instead assuming gazes are normally distributed with the mean in the center of the screen as discussed in section 4.1.2. Specifically, each user was unconscious of the gaze tracking system. He/she was naturally watching a movie in full screen mode. The experiment lasted

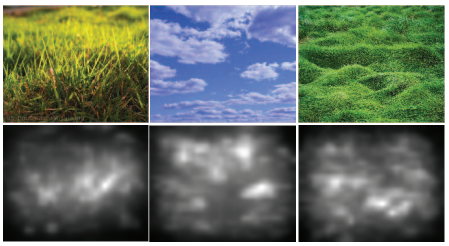


Fig. 19. Saliency maps (bottom row) of images without salient objects (top row).

TABLE 3
Gaze estimation error (in degree) of five subjects looking at images without salient objects.

Training data size	20 frames	40 frames	80 frames	120 frames	160 frames	200 frames
Subject 1	3.47	1.97	1.85	2.06	2.35	2.52
Subject 2	3.34	3.38	2.88	2.61	2.29	2.14
Subject 3	3.96	3.82	3.34	2.96	2.76	2.76
Subject 4	8.15	6.16	4.59	3.75	3.35	3.03
Subject 5	6.67	5.85	4.98	4.58	3.92	3.52
Average	5.12	4.24	3.53	3.19	2.94	2.80

about five minutes. The gaze estimation algorithm is similar to Algorithm 1. The difference is that we assume gaze follows a Gaussian distribution for each gaze point $p(\mathbf{g}_t)$. Thus, the probability distribution $p(\mathbf{g}_t|I_t)$ derived from the saliency map is replaced by Gaussian probability distribution $p(\mathbf{g}_t)$.

Our camera system captured totally 6000 optical axes in this five-minute experiment. The gaze estimation error for five subjects are summarized in Table 5. Please note that the results in Table 5 and the results of saliency map in Table 2 are based on different testing data sets, so that they are not directly comparable. However, we can at least draw the following conclusion: The gaze estimation with Gaussian prior needs longer time to converge (from a few seconds to five minutes) and if given enough training time (after 3000 frames), this method can achieve the same level of accuracy as the saliency map based method.

6.5 Eye Parameter Estimation

In our incremental learning, we continue updating the κ probability map in the experiment. After the experiment, we can extract the maximum point in the probability map as our estimation of eye parameter κ. This is our best estimation of κ after improving it using all the training frames. We extract κ^S using 200 frames in saliency-based incremental method (Sec.6.2), κ^B using saliency-based batch method (Sec.6.1), and κ^G using 6000 frames in Gaussian-based method (Sec. 6.4). We compare our eye parameter estimation with κ^* which is obtained from nine-point calibration in Table 6. We can see that our eye parameter estimate is close to the eye parameters from calibration. We also notice that, given enough training data, the eye parameters using batch and incremental methods are very similar.

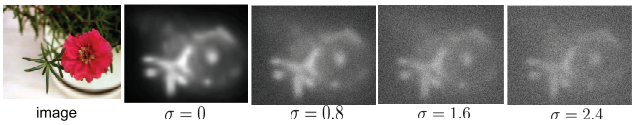


Fig. 20. Saliency maps with different levels of noise.

TABLE 4
Average gaze estimation error (in degree) of five subjects for the first 200 frames with noisy saliency map.

Noise Level σ	0	0.8	1.6	2.4
Gaze Error	1.77	1.79	2.06	2.75

In order to further evaluate our eye parameter estimation, we ask the person to look at nine fixed points on the screen. We estimate gazes using κ^*, κ^B, κ^S and κ^G respectively and take the fixed points as ground-truth to compute the gaze estimation errors. The gaze estimation errors and the horizontal and vertical angle bias of the five subjects are summarized in Table 7. Here, the estimation bias is accessed using the mean signed difference (MSD): $\mathrm{MSD}(x) = \sum_{i=1}^{n} \frac{(\hat{x}_i - x_i)}{n}$, where $\hat{x}_i$ is the estimate and x_i is the ground truth. We also compute the error and bias when we directly use the optical axis to estimate gaze without any calibration, i.e., set $\kappa = (0, 0)$. As expected, the estimation error with optical axis is very large. The error of our method with saliency map or Gaussian prior is a little higher than the calibration-based method, but our method doesn't need the cumbersome calibration procedure, and it can keep improving the gaze estimation when the user continues using the computer naturally.Notice that, compared to the average error of the first N frames in Table 2 and 5, this is the error using the eye parameter after N frames training procedure. Thus this gaze estimation error is smaller.

7 Conclusion

In this paper, we proposed a new probabilistic gaze estimation framework by combining the saliency map with the 3D eye gaze model. Compared to the traditional method, our proposed approach doesn't need the cumbersome and unnatural personal calibration procedure. Compared with the most recent calibration-free method [25], our system allows natural head movement. In addition, by considering the uncertainties of eye parameter and gaze in our probabilistic framework, our system significantly improves the accuracy from six degrees to less than three degrees. By using a novel incremental learning framework, our system doesn't need any training data from the subject beforehand. It can adapt to the user quickly and improves its performance as the subject naturally uses the system. Finally, we further extend our system without computing the saliency map by

TABLE 7
Comparison of gaze estimation error using the estimated eye parameters (κ^S, κ^G) and the error using the calibrated eye parameters κ^*.

Subject	κ^*		κ^G		κ^B		κ^S		Optical Axis	
	$error$	$bias$	$error$	$bias$	$error$	$bias$	$error$	$bias$		
1	1.0	(-0.3,0.1)	1.0	(-0.2, 0.3)	1.0	(-0.2,0.3)	1.0	(-0.2,0.2)	2.7	(2.6,-0.5)
2	1.0	(-0.4,-0.3)	1.3	(-0.5,-0.9)	1.0	(-0.3,0.3)	1.0	(-0.4,0.2)	4.3	(-3.2, 2.7)
3	0.9	(0.7,0.1)	1.2	(1.0,0.5)	1.0	(0.7,0.3)	1.0	(0.7,0.5)	3.4	(-2.7,1.9)
4	1.1	(0.1,-0.4)	1.2	(0.6,-0.4)	1.3	(0.8,0.5)	1.3	(0.8,0.5)	3.1	(-2.5, -1.5)
5	1.3	(0.3, 0.6)	1.9	(-0.1,1.3)	1.4	(0.2, -0.7)	1.5	(0.2, -0.8)	1.4	(0.7, -0.6)

TABLE 5
Gaze estimation error (in degree) of five subjects for the first N frames. The gaze prior is assumed as Gaussian distribution.

Training data size	100 frames	1000 frames	2000 frames	3000 frames	4000 frames	5000 frames	6000 frames
Subject 1	5.46	4.60	3.54	2.61	2.22	1.84	1.62
Subject 2	2.94	1.59	1.03	1.42	1.45	1.51	1.51
Subject 3	3.83	3.60	2.24	1.81	1.54	1.30	1.16
Subject 4	3.95	1.75	1.88	1.51	1.33	1.38	1.33
Subject 5	3.56	2.11	2.27	1.64	1.28	1.10	1.02
Average	3.94	2.73	2.19	1.80	1.57	1.43	1.32

TABLE 6
Comparison of the estimated eye parameters κ^S and κ^G against the eye parameters κ^* from calibration.

Subject	κ^*		κ^G		κ^B		κ^S	
	α	β	α	β	α	β	α	β
1	-3.11	0.02	-3.0	0.2	-3.0	0.2	-3.0	0.0
2	3.07	-2.43	3.0	-3.0	3.2	-2.4	3.0	-2.4
3	3.51	-1.33	3.8	-1.0	3.4	-1.2	3.4	-1.0
4	2.51	1.01	3.0	1.0	3.2	1.8	3.2	1.8
5	-0.51	1.02	-1.0	1.6	-0.6	-0.2	-0.4	-0.2

assuming the prior gaze distribution follows a Gaussian distribution, with a mean located in the center of the screen. This not only improves the speed of our method (without the need of computing saliency map), but also extend its application scope.

REFERENCES

[1] R. J. Jacob, "The use of eye movements in human computer interaction technqiues: What you look at is what you get," *ACM Tracnsations on Information Systems*, vol. 9, pp. 152–169, 1991.

[2] S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 246–253, 1999.

[3] Z. Zhu and Q. Ji, "Eye and gaze tracking for interactive graphic display," *Machine Vision and Applications*, vol. 15, pp. 139–148, 2004.

[4] S. Liversedge and J. Findlay, "Saccadic eye movements and cognition," *Trends in Cognitive Science*, vol. 4, pp. 6–14, 2000.

[5] M. Mason, B.Hood, and C. Macrae, "Look into my eyes : Gaze direction and person memory," *Memory*, vol. 12, pp. 637–643, 2004.

[6] K. Hyoki, M. Shigeta, N. Tsuno, Y. Kawamuro, and T. Kinoshita, "Quantitative electro-oculography and electroencephalography as indexes of alertness," *Electroencephalogr. Clinical Neurophysiology*, vol. 106, pp. 213–219, 1998.

[7] K. Tan, D. Kriegman, and H. Ahuja, "Appearance based eye gaze estimation," *Proceedings of IEEE Workshop Applications of Computer Vision*, pp. 131–136, 2002.

[8] C. H. Morimoto and M. R. Mimica, "Eye gaze tracking techniques for interactive applications," *Computer Vision and Image Understanding, Special Issue on Eye Detection and Tracking*, vol. 98, pp. 4–24, 2005.

[9] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics, PartB*, vol. 34, pp. 234–245, 2004.

[10] J.-G. Wang and E. Sung, "Study on eye gaze estimation," *IEEE Transactions on Systems, Man and Cybernetics, PartB*, vol. 32, no. 3, pp. 332–350, 2002.

[11] C. H. Morimoto, A. Amir, and M. Flickner, "Detecting eye position and gaze from a single camera and 2 light sources," *Proceedings of the International Conference on Parttern Recognition*, 2002.

[12] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," *Computer Vision and Pattern Recognition. IEEE Computer Society Conference on*, 2003.

[13] J. Chen, Y. Tong, W. Gray, and Q. Ji, "A robust 3d eye gaze tracking system using noise reduction," *Proceedings of the symposium on Eye tracking research & applications*, 2008.

[14] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Transcations on Biomedical Engineering*, 2006.

[15] ——, "Remote point-of-gaze estimation requiring a single-point calibration for applications with infants," *Proceedings of the 2008 symposium on Eye tracking research & applications*, 2008.

[16] J. Chen and Q. Ji, "3d gaze estimation with a single camera without ir illumination," *Pattern Recognition, 2008. 19th International Conference on*, 2008.

[17] Z. Zhu, Q. Ji, and K. P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," *International Conference on Pattern Recognition*, 2006.

[18] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, 2010.

[19] T. Huchinson, K. P. W. Jr., and K. Reichert, "Human computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, pp. 1527–1533, 1989.

[20] LC Technologies Inc., http://www.eyegaze.com, date Last Accessed, 12/15/2010.

[21] Applied Science Laboratories, http://asleyetracking.com, date Last Accessed, 12/15/2010.

[22] SensoMotoric Instruments, http://www.smivision.com, date Last Accessed, 12/15/2010.

[23] D. Model and M. Eizenman, "An automatic personal calibration procedure for advanced gaze estimation systems," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, 2010.

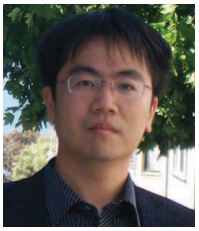
[24] W. Miao, J. Chen, and Q. Ji, "Constraint-based gaze estimation without active calibration," *9th International Conference on Face and Gesture Recognition*, 2011.

[25] Y. Sugano, Y. Matsushita, and Y. Sato, "Calibration-free gaze sensing using saliency maps," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[26] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems (NIPS)*, 2006.

[27] C. W. Oyster, *The Human Eye: Structure and Function*. Sinauer Associate, Inc., 1999.

[28] D. Koller and N. Friedman, *Probabilistic Graphical Models : Principles and Techniques*. The MIT Press, 2009.



Jixu Chen received his Ph.D degree in Electrical Engineering from Rensselaer Polytechnic Institute (RPI), Troy, NY in 2011. He is currently a researcher in Computer Vision Lab at GE Global Research. His research interests include computer vision, machine learning, and human-computer interaction. He is a member of the IEEE and the IEEE Computer Society.

Qiang Ji received his Ph.D 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 recently served as a program director at the National Science Foundation (NSF), where he managed NSF's computer vision and machine learning programs. He also held teaching and research positions with the Beckman Institute at University of Illinois at Urbana-Champaign, the Robotics Institute at Carnegie Mellon University, the Dept. of Computer Science at University of Nevada at Reno, and the US Air Force Research Laboratory. Prof. Ji currently serves as the director of the Intelligent Systems Laboratory (ISL) at RPI.

Prof. Ji's research interests are in computer vision, probabilistic graphical models, information fusion, and their applications in various fields. He has published over 150 papers in peer-reviewed journals and conferences. His research has been supported by major governmental agencies including NSF, NIH, DARPA, ONR, ARO, and AFOSR as well as by major companies including Honda and Boeing. Prof. Ji is an editor on several related IEEE and international journals and he has served as a general chair, program chair, technical area chair, and program committee member in numerous international conferences/workshops.