Towards an accurate 3D deformable eye model for gaze estimation

Chenyi Kuang[1], Jeffery O. Kephart[2], and Qiang Ji[1]

[1] Rensselaer Polytechnic Institute
{kuangc2,jiq}@rpi.edu
[2] IBM Thomas J. Watson Research Ctr.
kephart@us.ibm.com

Abstract. 3D eye gaze estimation has emerged as an interesting and challenging task in recent years. As an attractive alternative to appearance-based models, 3D model-based gaze estimation methods are powerful because a general prior of eye anatomy or geometry has been integrated into the 3D model hence they adapt well under various head poses and illumination conditions. We present a method for constructing an anatomically accurate 3D deformable eye model from the IR images of eyes and demonstrate its application to 3D gaze estimation. The 3D eye model consists of a deformable basis capable of representing individual real-world eyeballs, corneas, irises and kappa angles. To validate the model's accuracy, we combine it with a 3D face model (without eyeball) and perform image-based fitting to obtain eye basis coefficients The fitted eyeball is then used to compute 3D gaze direction. Evaluation results on multiple datasets show that the proposed method generalizes well across datasets and is robust under various head poses.

Keywords: Eyeball modeling · 3DMM · Gaze estimation.

1 Introduction

Eye gaze – an important cue for human behaviour and attention – has been widely explored in recent years by computer vision researchers. As interactive applications such as AR/VR, 3D avatar animation and driver behaviour monitoring [7–9,14] gain more popularity, various 3D gaze methods have been proposed (with much recent emphasis on deep-learning based models). Based on the devices and data they use, 3D gaze estimation methods can be divided into two categories: (1) appearance-based gaze estimation from images/videos; and (2) 3D eye model recovery and model-based gaze estimation. Appearance-based methods usually focus on extracting eye features from web cameras or IR cameras. Such methods can be sensitive to different head poses and illumination conditions; hence their generalization ability can be limited. 3D model-based methods takes a different strategy that entails recovering the anatomical structure of a person's eyeball. Based on the devices and data they require, 3D model-based methods can be further divided into two types: (a) personalized 3D eye model

recovery from IR camera systems and (b) 3D eye shape estimation from image features using a pre-constructed deformable eye basis.

The first type (2a) usually requires setting up specific devices and using a complex calculation process to handle light refraction, IR camera calibration, etc. These methods usually build a geometric eye model to represent the anatomical eyeball structure including pupil diameter, 3D pupil center and cornea curvature center. Based on such computation, some wearable devices are offered with a pre-installed and calibrated camera and illumination system for real-time 3D gaze estimation. However, the practicality and accessibility of such methods can be limited. More recently, as 3D morphable face models are successfully applied in accurate 3D face reconstruction and animation, similar experiments have been conducted for constructing a deformable eye model from 3D scans. Wood et al. [32] proposed a 3D deformable eye region model constructed from high-quality 3D facial scans, in which the eye region and the size of iris are parameterized using a PCA basis. Ploumpis et al. [23] constructed a large-scale statistic 3D deformable full head model, including face, ear, eye region and pupil size. Both [32] and [23] can be applied to eye 3DMM fitting using image feature points for recovering 3D gaze direction. Such statistic eye models provide a parameterized linear space for approximating the size of a new subject's eyeball and can be directly utilized in image-based fitting for gaze estimation.

This paper presents an accurate 3D deformable eye model constructed from recovered geometric parameters of multiple subjects. More specifically, we use the wearable device Tobii pro Glasses2 for data collection and compute individual geometric eyeball parameters through explicit IR camera calibration, pupil & iris detection and glint detection. The eyeball geometry is represented as two intersecting spheres: the eyeball and the cornea, with person-specific parameters for eyeball radius, cornea radius, iris radius and kappa angle. Based on the constructed model, we propose a two-phase framework of 3D gaze estimation for webcam images. In summary, the contributions of this paper include:

- Eye data collection with Tobii pro Glasses2 including 3D gaze direction, 3D gaze point and IR videos of eye region. Then personal eyeball parameters are recovered from data, including eyeball radius, cornea radius, iris radius and kappa angle. PoG (Point of Gaze) error is calculated for evaluating recovered parameters.
- An accurate parameterized 3D eye model with PCA eye basis that represents personal variations in 3D eye geometry.
- Integration of the constructed eye model with a sparse 3D face model, yielding a two-phase gaze estimation framework for monocular webcam images. Experimental results show that our model generalizes well to different benchmark datasets for 3D gaze estimation.

2 Related Works

Our proposed method takes advantage of techniques from 3D eye model recovery using infrared or RGBD cameras combined with model-based gaze estimation. We focus on reviewing related works in these two areas.

2.1 3D Eye Modeling

Infrared-camera based 3D eye model recovery systems are usually paired with pre-calibrated illuminators to generate detectable glints in the IR images [4, 13, 16]. Through glint tracking and solving for the light reflection equations on cornea surface, the 3D cornea center can be estimated. The 3D pupil center can be solved by ellipse calibration. The IR camera-light system can achieve good accuracy and precision for estimating eyeball geometry, but the setup process is complex. For the convenience of real time gaze estimation, multiple wearable devices have been developed like [12, 25–27]. Usually a one-time personal calibration is required by these devices before starting gaze tracking, which is used to recover personal 3D eye model information in advance.

3D eye model recovery methods based on RGB-D-cameras have been proposed as well. Wang et al. [29] recovered subject-dependent 3D eye parameters including eyeball radius, cornea center to eyeball center offset, eyeball center to head center offset and kappa angle using a Kinect camera. Zhou et al. [34] recovered the 3D eyeball center, eyeball radius and iris center using the geometry relationship of two eye models with a Kinect camera. More recently, concurrent with the development of large scale 3D morphable face models [3, 11, 18, 22], researchers are using a similar process to construct a deformable eye region model. Woods et al. [32] constructed a 3D deformable eye model from large scale 3D facial scans. Their model contains a deformable shape basis for the iris, eye socket, eye lid and eye brow. Ploumpis et al. [23] presented a complete 3D deformable model for the whole human head that incorporates eye and eye region models. Compared with [32], their eye model uses finer-grained groups of eyeball, cornea and iris vertices and captures variations in pupil size. We are unaware of prior work that constructs a detailed eye mesh model that focuses simultaneously on modeling the variance of eyeball size, cornea size and iris size.

2.2 Model-based 3D Gaze Estimation

3D gaze estimation methods can be divided into two types: appearance-based methods (which take advantage of image features) and model-based methods, the former type is not discussed in detail and we focus on model-based methods in this paper. Model-based gaze estimation methods have two major advantages over appearance-based methods. First, 3D models are less vulnerable to variations in illumination because it contains a general geometry prior for the 3D eye anatomy that can be fit to different images. Second, 3D models can be rotated arbitrarily by assigning a rotation matrix, making them more robust to head and eye pose variations. Wang et al. [30] exploit a sparse 3D Face-Eye model that can

deform in eyeball center position, pupil position and eyeball radius. Based on the face model, 3D head poses can be solved in advance and then eyeball rotation and kappa angle are solved by minimizing eye-landmark error and gaze error. Woods et al. [32] and Ploumpis et al. [23] introduced an "analysis-by-synthesis" framework to fit their 3D eye model to image features. The 3D head pose and eye pose are optimized separately, and eye image texture is utilized to fit the eye pose. They achieved good accuracy in 3D gaze estimation without using gaze labels.

3 3D Deformable Eye Model

The main objective of data collection is to recover anatomically accurate parameters that capture individual eyeball structure. To fully utilize existing resources and tools, we choose a reliable eye tracking device, Tobii Pro Glasses2, which allows us to capture human gaze data in real-world environments in real time. We collected both infrared eye images and true gaze data from Tobii Pro Glasses2, where the former are utilized for computing eye model parameters for each participant and the latter are used as ground-truth to validate our calculation process. In all, we recruited 15 participants, each of whom were involved in multiple data collection experiments to ensure that the training and validation data are both valid. We detail our data collection and processing pipeline in section 3.1.

3.1 Data Collection and Personal Eye Parameter Recovery

Tobii Pro Glasses2 consists of a head unit, a recording unit and controller software. The head unit contains four eye tracking sensors (two for each eye) that take infrared eye region images from different angles to analyze gaze direction and one high-resolution scene camera capturing HD videos of what is in front of the person. Additionally, there are six IR illuminators on each side that generate glints in the eye images due to corneal reflection. For each participant, a pre-calibration before recording is required to ensure that the glass is properly worn and the sensors successfully capture the pupil center of both eyes. We invite each participant to wear the glass and sit in front of a $80cm \times 135cm$ screen at a distance of 1.5-1.9m. The participant is asked to track a moving dot in a 3×7 dot array displayed on the screen. During the recording, the participant is allowed to adjust their head orientation in case the gaze angle is too large to be well captured for some corner dots. Each participant is asked to repeat the recording experiment 2-3 times so that we can collect sufficient valid data for generating the model and performing validation. After each recording, four IR eye videos and one scene video as well as a trajectory file documenting 2D and 3D eye gaze information at each time sampling step is saved. We use the eye videos to recover personal 3D eyeball parameters. The scene video and trajectory file are utilized for the validation stage, which will be discussed in section 5.2.

Table 1: 3D Eyeball Parameters

3D parameters	**Notation**
Eyeball center	O_e
Eyeball radius	r_e
Cornea center	O_c
Cornea radius	r_c
iris& pupil center	O_i
iris radius	r_i
kappa angle	$\boldsymbol{\theta} = [\theta_1, \theta_2]$
optical axis	$\boldsymbol{n}_o$
visual axis	$\boldsymbol{n}_v$

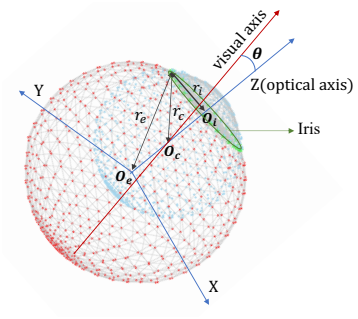

Fig. 1: 3D eye mesh

The personal 3D eye model is defined as a two-sphere system, where the larger sphere represents a 3D eyeball with center O_e and radius r_e and the smaller one represents the cornea with center O_c and radius r_c. The intersection of the two spheres results in a circular plane whose center is defined as iris center O_i. The pupil is assumed to be a concentric circle with the iris circle; hence the pupil center overlaps with the iris center. Geometrically, O_e, O_c, O_i are co-linear points and their connection forms the optical axis. The optical axis can be represented by horizontal and vertical angles (ϕ, γ):

$$\boldsymbol{n}_o(\phi, \gamma) = \begin{pmatrix} cos(\phi)sin(\gamma) \\ sin(\phi) \\ -cos(\phi)cos(\gamma) \end{pmatrix}$$

According to eyeball anatomy, the true gaze is defined by a visual axis that connects the fovea center, cornea center and the target object. We define a 2D vector $\boldsymbol{\theta} = [\theta_1, \theta_2]$ to be the Kappa angle representing the calibration term so that visual axis will be $\boldsymbol{n}_v = \boldsymbol{n}_o(\phi + \theta_1, \gamma + \theta_2)$. We summarize the geometric parameters to be recovered in Table. 1 and show defined 3D eyeball mesh in Fig. 1. We describe the process of recovering 3D eyeball geometry from Tobii data as below.

3D pupil center We first process eye camera images for pupil ellipse detection, as depicted in Fig. 2(a). Then the 3D pupil center O_i(relative to the reference camera) is recovered through stereo rectification.

3D cornea center With pre-calibrated IR illuminators, We first detect glints $g_{1,1}, g_{1,2}$ caused by light I_1 in two images to calculate the 3D virtual glints v_1, similarly we can obtain another virtual glint v_2 caused by light I_2.The intersection of two light rays l_1, l_2 will be the 3D cornea center O_c. An illustration is shown in Fig. 2(b).

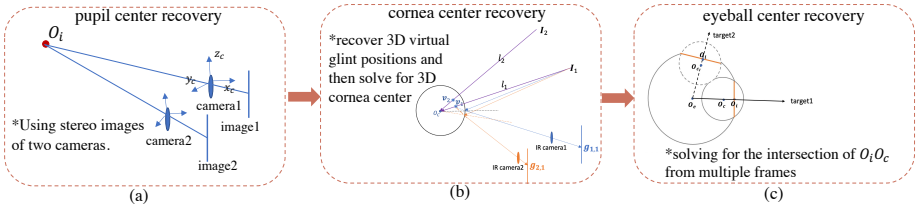

Fig. 2: 3D eyeball recovery process

3D eyeball center We assume that a user's head movement would not cause any position shift of the glass, hence O_e can be considered as a constant vector across one whole recording. As the rotating center of the eyeball, O_e can be estimated by solving for the intersection of O_iO_c (which is the connecting line between O_i and O_c) from multiple frames, as shown in Fig. 2(c).

Eyeball, cornea and iris radius By referencing the ellipse reconstruction method introduced by Kohlbecher et al. [15] and Chen et al. [5], we can recover the 3D circular function for the iris plane, i.e. the normal vector and iris radius r_i. As in 3D eye geometry we define the iris as the intersection plane of eyeball sphere and cornea sphere, r_e, r_c can be determined by:

$$\begin{aligned} r_c^2 &= r_i^2 + d_{ci}^2 \\ r_e^2 &= r_i^2 + (d_{ci} + d_{ec})^2 \end{aligned} \tag{1}$$

where d_{ci} and d_{ec} are the distance from O_c to O_i and O_e to O_c respectively and can be obtained from 3.1,3.1,3.1.

Kappa angle According to the accuracy report provided by Tobii Pro Glass2 [12], we extract 3D gaze direction for valid frames of a recording provided by Tobii as the ground truth $\hat{\boldsymbol{n}}_v$ and optimize for the kappa angle by

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \sum_{m=1}^{M} \arccos(\boldsymbol{n}_o(\phi^* + \theta_1, \gamma^* + \theta_2), \hat{\boldsymbol{n}}_v) \tag{2}$$

In Table 1, only camera-invariant parameters $\boldsymbol{p} = [r_e, r_c, r_i, \theta_1, \theta_2]$ are selected to construct a personal 3D eye mesh and we manually define O_e to be the origin and O_c, O_i are located on the Z-axis.

3.2 3D Deformable Eye Model Construction

We repeat the process in section 3.1 for each participant and we designate these users as "calibrated users" since their personal eye parameters $\boldsymbol{p}^{5\times1}$ are fully recovered by means of the Tobii device. The calibrated parameter set $\boldsymbol{P}$ =

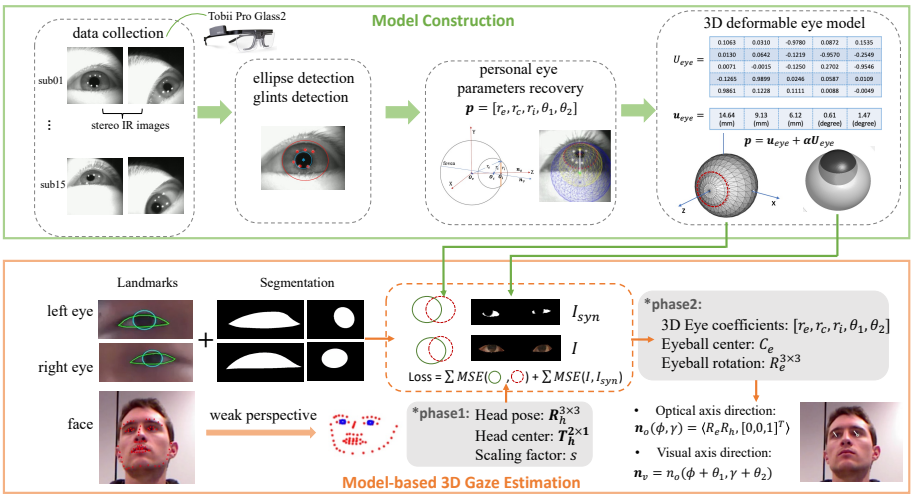

Fig. 3: Overview of our method: including a model construction module and a model-based 3D gaze estimation module.

$[\boldsymbol{p_1}, \cdots, \boldsymbol{p_{15}}]^{5\times15}$ can be used to construct a linear model that describes variation in eye parameters:

$$\boldsymbol{\mathcal{M}}_{eye} = (\boldsymbol{\mu}_{eye}, \boldsymbol{U}_{eye}) \tag{3}$$

where $\boldsymbol{\mu}_{eye}^{5\times1}$ and $\boldsymbol{U}_{eye}^{5\times5}$ are the average 3D eye parameter and an orthogonal PCA basis computed from $\boldsymbol{P}$. For model-based 3D gaze estimation, we reconstruct $\boldsymbol{p}$ for an "uncalibrated user" by letting $\boldsymbol{p} = \boldsymbol{\mu}_{eye} + \boldsymbol{\alpha}\ \boldsymbol{U}_{eye}$ with the personal coefficient $\boldsymbol{\alpha}$, rather than repeating the complex data collection and processing procedure using the Tobii device.

We define a fixed eye mesh topology for the two-sphere 3D eye model, where the mesh vertices are divided into different groups, including $\{\Omega_1$:"eyeball", Ω_2"cornea",Ω_3:"iris boundary",Ω_4:"pupil center"$\}$. An average 3D eye mesh can be scaled corresponding to personal eye parameters $[r_e, r_c, r_i]$, resulting in the personal mesh generation process described as: $\boldsymbol{V}^{N\times3} = f(r_e, r_c, r_i)$.

4 3D Gaze Estimation

We propose a two-phase framework for single-frame based 3D gaze estimation using the constructed 3D eye model, as shown in Fig. 3. This framework can be implemented by an optimization-based fitting or deep-model-based regression scheme. In this section we will discuss the algorithm for 3D gaze estimation through model fitting.

In **phase1**, we solve for a weak-perspective camera viewpoint for the 3D head by fitting a 3D head model [17] to 68 2D facial landmarks of the face image. We implement the SNLS algorithm proposed by Bas te al. [1] and obtain

the optimal 3D head pose $\boldsymbol{R}_h^{3\times3}$, head center position $\boldsymbol{T}_h^{2\times1}$ and a scaling factor s. Then, in **phase2**, we use a combination of iris landmark loss, rendering loss and geometrical constraints for left and right eyeball to optimize the eyeball center location $[\boldsymbol{C}_{el}, \boldsymbol{C}_{er}]$ in the head coordinate system (HCS), the personal eye coefficient $\boldsymbol{\alpha}$ and the eyeball rotation $\boldsymbol{R}_e^{3\times3}$. Loss terms are as follows, taking the left eye as an example.

Iris landmark loss: similar to **phase1**, we obtain 2D iris landmarks $\boldsymbol{x}_{iris}^{2d}$ from an iris detector and formulate the projection loss as:

$$L_{iris,l} = \|\boldsymbol{x}_{iris,l}^{2d} - s^* P[\boldsymbol{R}_e \boldsymbol{R}_h^* \{\boldsymbol{V}_i, i \in \Omega_3\} + \boldsymbol{R}_h^* \boldsymbol{C}_{el}] + s^* \boldsymbol{T}_h^* \|^2 \tag{4}$$

where $\boldsymbol{V} = f(\boldsymbol{p}(1), \boldsymbol{p}(2), \boldsymbol{p}(3))$, and $\boldsymbol{p}(\boldsymbol{\alpha}) = \boldsymbol{\mu}_{eye} + \boldsymbol{\alpha}\ \boldsymbol{U}_{eye}$.

Rendering loss: Since optimizing all parameters merely using iris landmarks is an ill-posed problem, we add a texture loss for the sclera and iris regions by projecting the eye mesh to the image frame, which is defined as:

$$L_{img,l} = \frac{\sum_m A_{m,l} \|I - I_{syn}(s^*, \boldsymbol{\alpha}, \boldsymbol{C}_{el}, \boldsymbol{R_e})\|_2}{\sum_m A_{m,l}} \tag{5}$$

where $A_{m,l}$ is the binary mask for the left eye region generated by 2D eye landmarks.

Geometrical constraints: from **phase1** we can recover 3D eye landmarks $\boldsymbol{x}_{eye}^{3d}$ among the facial landmarks. Since they ought to be very close to the surface of eyeball sphere, we define a regularization term for the eyeball center and radius:

$$L_{geo,l} = \sum_i \|r_e - |\boldsymbol{x}_{eye,l,i}^{3d} - \boldsymbol{C}_{el}|_1\|^2 \tag{6}$$

where $r_e = \boldsymbol{p}(1)$ and $\boldsymbol{p} = \boldsymbol{\mu}_{eye} + \boldsymbol{\alpha}\ \boldsymbol{U}_{eye}$. Eq. 4, 5, 6 apply to both eyes. We further constrain the left and right eyeball centers to be symmetric in HCS, i.e., $\boldsymbol{C}_e = [\boldsymbol{C}_{el}(1), \boldsymbol{C}_{el}(2), \boldsymbol{C}_{el}(3)] = [-\boldsymbol{C}_{er}(1), \boldsymbol{C}_{er}(2), \boldsymbol{C}_{er}(3)]$. The resulting overall cost function is:

$$\begin{aligned} \arg \min_{\boldsymbol{Ce}, \boldsymbol{\alpha}, \boldsymbol{R_e}} & \lambda_1 (L_{iris,l} + L_{iris,r}) + \lambda_2 (L_{img,l} + L_{img,r}) \\ & + \lambda_3 (L_{geo,l} + L_{geo,r}) + \lambda_4 \|\boldsymbol{\alpha}\|_2 \end{aligned} \tag{7}$$

the last term is used to avoid unreasonable personal eye shapes.

Kappa angle refinement: analysis of the 3D eye basis $\boldsymbol{U}_{eye}$ shows that the kappa angles have very weak correlation with other parameters or, we can consider $\boldsymbol{\theta}$ as independent variable from $[r_e, r_i, r_c]$. Hence, we add an optional loss term to refine the 3D eye model parameters when gaze labels are available. The gaze loss is defined as:

$$L_{gaze} = \arccos \langle \boldsymbol{n}_o(\phi_o + \theta_1, \gamma_o + \theta_2\ ), \hat{\boldsymbol{n}}_v \rangle \tag{8}$$

where ϕ_o, γ_o are function of $\boldsymbol{R}_e$ expressed by $\boldsymbol{n}_o = \begin{pmatrix} cos(\phi)sin(\gamma) \\ sin(\phi) \\ -cos(\phi)cos(\gamma) \end{pmatrix} = \boldsymbol{R}_e \boldsymbol{R}_h \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$.

Our 3D gaze estimation framework is summarized in Algorithm. 1.

Algorithm 1 3D Gaze Estimation Algorithm

phase1: head pose estimation

Input: $\begin{cases}\text{landmarks} : \boldsymbol{x}_{face}^{2d} \\ \text{Deformable face model} : \boldsymbol{B}, \bar{\boldsymbol{B}}\end{cases}$

Fitting: weak perspective, SNLS algorithm [1].

Output: $[\boldsymbol{R}_h^*, \boldsymbol{T}_h^*, s^*]$.

phase2: 3D Gaze estimation

Input: $\begin{cases}\text{image, landmarks \& eye mask} : I, \boldsymbol{x}_{eye}^{2d}, \boldsymbol{x}_{iris}^{2d}, A_m \\ \text{head pose and scaling factor} : [\boldsymbol{R}_h^*, \boldsymbol{T}_h^*, s^*] \\ \text{3D deformable eye model} : \boldsymbol{U}, \bar{\boldsymbol{u}}\end{cases}$

Initialization: $\boldsymbol{R_e^0} = \text{Identity}(3), \boldsymbol{\alpha}^0 = \boldsymbol{0}, \boldsymbol{C}_e^0$ initialized by face model.

Fitting:

if gaze label $\hat{\boldsymbol{n}}_v$ available **then**

$$\begin{aligned}\boldsymbol{C}_e^*, \boldsymbol{R_e^*}, \boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{C}e,\boldsymbol{\alpha},\boldsymbol{R_e}} \; & \lambda_1(L_{iris,l} + L_{iris,r}) \\ & + \lambda_2(L_{img,l} + L_{img,r}) + \lambda_3(L_{geo,l} + L_{geo,r}) \\ & + \lambda_4\|\boldsymbol{\alpha}\|_2 + \lambda_5 L_g\end{aligned}$$

else

$$\begin{aligned}\boldsymbol{C}_e^*, \boldsymbol{R_e^*}, \boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{C}e,\boldsymbol{\alpha},\boldsymbol{R_e}} \; & \lambda_1(L_{iris,l} + L_{iris,r}) + \lambda_4\|\boldsymbol{\alpha}\|_2 \\ & + \lambda_2(L_{img,l} + L_{img,r}) + \lambda_3(L_{geo,l} + L_{geo,r})\end{aligned}$$

Output: $[\boldsymbol{C}_e^*, \boldsymbol{R_e^*}, \boldsymbol{\alpha}^*]$

Gaze: optical axis: $\boldsymbol{n}_o(\phi_o^*, \gamma_o^*) = \boldsymbol{R}_e^* \boldsymbol{R}_h^* \begin{bmatrix}0 & 0 & 1\end{bmatrix}^T$
visual axis: $\boldsymbol{n}_v(\phi_o^* + \boldsymbol{\alpha}^*(4), \gamma_o^* + \boldsymbol{\alpha^*}(5))$

5 Experiments

5.1 Experiment settings

Datasets: We conduct two types of experiments to evaluate the constructed model:

- **Model validation on Tobii recordings**. As mentioned in section 3.1, we collect multiple recordings for each participant and leave one recording out for validation;
- **Benchmark datasets**. We select two datasets with full face images available. Columbia Gaze dataset [24] contains 56 subjects with 21 gaze angles under 5 head poses. EyeDiap [10] contains 16 subjects with different sessions.

On benchmark datasets, we first perform 2D facial landmark detection using [2] and 2D iris detection using [20]. We use FaceScape [35] as the 3D face shape model to perform head pose estimation in **phase1**.

Table 2: Average 3D gaze error and PoG error on Tobii recordings of 15 participants.

azimuth angle error (in degree)	elevation angle error (in degree)	PoG error/screen size(in cm)
3.32	3.56	7.64/(80*135)

5.2 Evaluation on Tobii recordings

For each "calibrated" participant with fully recovered eye parameters $\boldsymbol{p}$, we perform validation experiments on one of the unused recordings. On the validation IR video, frame-based 3D model fitting is conducted by minimizing the MSE between projected 3D iris vertices and detected 2D iris landmarks. After optimizing for eyeball center $\boldsymbol{O_e}$ and eyeball rotation $\boldsymbol{R_e}$, the optical axis can be represented as $\boldsymbol{n}_o(\phi_o, \gamma_o) = \boldsymbol{R}_e[0,\ 0,\ 1]^T$ then visual axis can be calculated by $\boldsymbol{n}_v = \boldsymbol{n}_o(\phi_o + \theta_1, \gamma_o + \theta_2)$. Since the visual axis is the unit direction vector connecting cornea center and target object, the point of gaze (PoG) will be the intersection of the left- and right-eye gaze vectors $\boldsymbol{O}_{c,l} + l_1\boldsymbol{n}_{v,l}$ and $\boldsymbol{O}_{c,r} + l_2\boldsymbol{n}_{v,r}$, which is computed by solving for l_1 and l_2. In the trajectory file provided by Tobii Pro Glass2, we are able to extract the 3D true gaze vector and the detected 3D target. We validate our model construction procedure in section 3.1 by evaluating angular gaze error and PoG error. Since eyeball is a sphere structure and can be only rotated along two direction, we decompose the 3D gaze vector provided by Tobii into two free rotation angles: horizontal angle $\hat{\phi}_v$ and vertical angle $\hat{\gamma}_v$, then write the gaze vector as $\hat{\boldsymbol{n}}_v(\hat{\phi}_v, \hat{\gamma}_v)$. Comparing the estimated visual axis $\hat{\boldsymbol{n}}_v(\phi_o + \theta_1, \gamma_0 + \theta_2)$ with the ground truth $\hat{\boldsymbol{n}}_v(\hat{\phi}_v, \hat{\gamma}_v)$, the angular gaze error can be reflected by a horizontal angle error $\Delta\phi = |\hat{\phi}_v - (\phi_o + \theta_1)|$ and a vertical angle error $\Delta(\gamma) = |\hat{\gamma}_v - (\gamma_o + \theta_2)|$. Results are summarized in Table 2.

5.3 Evaluation on benchmark datasets

Evaluations on IR eye videos mentioned in section 5.2 validate that the recovered parameters $\boldsymbol{p} = [r_e, r_c, r_i, \theta_1, \theta_2]$ fit each participant well and can be taken as valid data for constructing our deformable 3D eye model. In addition to that, we evaluated how well the proposed 3D eye model predicts gaze directions for webcam datasets: Columbia Gaze and EyeDiap. For both datasets, we estimate 3D eyeball parameters $[\boldsymbol{C_e}, \boldsymbol{R_e}, \boldsymbol{\alpha}]$for each subject, under the condition of using 3D gaze labels or not. When no gaze label is involved, i.e., we use step.10 and step.12 in Algorithm.1 to estimate gaze direction for each image. We can also use gaze labels and do step.9 in Algorithm.1 to estimate $\boldsymbol{C}_e^*$ and $\boldsymbol{\alpha}^*$ for a subset of images of one subject and then use the average result as initialization for the remained images of this subject. For the second case, we can get more accurate estimated gaze since $\boldsymbol{C}_e^*$ and $\boldsymbol{\alpha}^*$ are more consistent in terms of subject identity. In all, we designed three experiments with no gaze label(0% column), 5% labels and 10% labels for each subject. We compared our model with SOTA

Table 3: Average angular error in under 5 different head angles in Columbia Gaze dataset, **H**: horizontal gaze angle error, **V**: vertical gaze angle error.

Gaze error (H,V)	Percentage of gaze label used		
Head pose	0%	5%	10%
-30°	(8.18,6.80)	(6.80,6.20)	(6.54,6.31)
-15°	(8.20,6.54)	(6.54,6.06)	(6.42,5.89)
0°	(7.80,6.50)	(6.00,5.87)	(6.05,5.64)
15°	(7.90,6.54)	(6.12,5.54)	(6.18,5.32)
30°	(8.24,6.66)	(6.28,5.96)	(6.17,5.88)
Avg.	(8.06,6.61)	(6.35,5.93)	(6.28,5.81)

Table 4: Comparing with state-of-art models on Columbia Gaze, EyeDiap-VGA video and EyeDiap-HD video using different percentage of gaze labels.

Datasets	[33]	[28]	[32]	[30]	**Ours**(with (·)% labels)		
					0%	5%	10%
Columbia Gaze	9.7	10.2	8.9	7.1	9.0	6.5	**6.1**
EyeDiap-VGA	21.2	22.2	**9.44**/21.5	17.3	11.4/19.6	10.2/16.7	9.6/**16.0**
EyeDiap-HD	25.2	28.3	11.0/22.2	16.5	10.5/18.1	9.8/15.4	**9.6**/**14.7**

3D eye modeling methods,including [28, 30, 32, 33], for evaluating our 3D eyeball geometry and fitting algorithm. Most appearance-based gaze estimation models like [6, 19, 21, 31] which usually need full gaze labels and a complex training process to extract deep features from eye images and map to human gaze, rather than estimating 3D geometry and perform 3DMM-fitting. Therefore, we do not compare our model with these methods in this paper. It's worth mentioning that our 3D deformable eye model can be integrated into a deep model framework and combined with appearance-based methods. We'll continue with this part in future research.

The results for Columbia Gaze are shown in Table 3. We use different percentages of gaze labels to get refined kappa angles. Comparing results in Table 3 vertically, our model fits well to different head pose angles, although for larger head angles the gaze estimation accuracy is slightly reduced. From Table 3, it can be seen that our fitting algorithm's estimates of personal 3D eye model parameters and 3D gaze directions improve substantially when the percentage of gaze labels is raised from 0% to 5%, with the angular error decreasing from 9.0° to 6.5°. Increasing the percentage of gaze labels to 10% has only a marginal benefit (angular error 6.1°). Our model outperforms [32] even when we use no gaze labels, and it outperforms [30] (which uses around 14% gaze labels when fitting for their 3D eye model) even when we just use 5% or 10% gaze labels.

On EyeDiap we performed the fitting on VGA images and HD images. We present our results on EyeDiap in the last three columns of Table 4. For fair comparison, we divide the testing data into ("screen target") / ("floating target") similar to [32] and show the 3D gaze error separately. For [28, 30, 33] we list

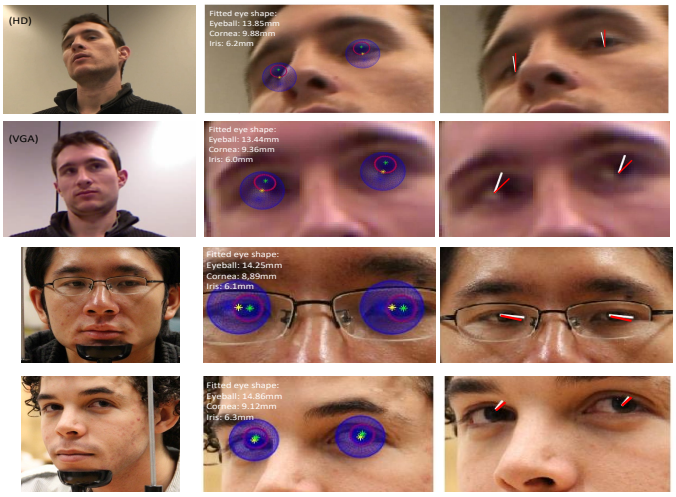

Fig. 4: Example fitting results on EyeDiap [10] and Columbia Gaze [24] datasets. **column 1**: input image. **column2**: projected 3D eyeball vertices(blue), iris vertices (red), eyeball center (yellow) and pupil center (green). Estimated eyeball & cornea & iris radius are displayed upper left. **column3** estimated gaze (white) and ground-truth gaze direction (red).

the average error on EyeDiap since they do not explicitly split the data. On EyeDiap-VGA videos, event with no gaze labels our model achieves the best gaze estimation results (19.6°) on “floating target” data that exhibit large head pose angles compared to [32] (21.5°). Our gaze estimation performances are further improved when utilizing 10% gaze labels, achieving (9.6°/16.0°). On EyeDiap-HD videos, which have higher resolution than VGA videos, our model outperforms all of the four methods for both "screen target" video and "floating target" video. Fitting examples on Columbia dataset and EyeDiap are visualized in Fig. 4.

To summarize, our 3D model and fitting algorithm achieves state-of-the-art gaze estimation accuracy even when using a small percentage of subject gaze labels. Furthermore, the proposed two-phase fitting algorithm is more robust against large head poses and can still perform well under illumination and image resolution variations.

6 Conclusion

We propose the first 3D eye model with a deformable basis for eyeball radius, cornea radius, iris radius and kappa angle. The 3D eye geometry contains a sphere for eyeball, a smaller sphere for the cornea and the iris plane. The 3D eye geometry is fully parameterized by the eye model coefficients and can be used to approximate the variance in 3D eye shape for different person. We use a

wearable device Tobii Pro Glass2 for data collection and preliminary model validation. We present a two-phase fitting algorithm for single-image based 3D gaze estimation using the constructed eye basis. With our 3D eye model and fitting method, personal eye shape parameters and eyeball rotations can be recovered from image pixel feature. Evaluations on benchmark datasets show that our model generalizes well to web-camera images with various head poses, illumination and resolution. The fitting process introduced in our paper can be further transplanted into a deep-model based framework. In the future, we pursue to integrate the 3D eye model into appearance-based deep models for accurate and generalizable 3D gaze estimation.

Acknowledgment The work described in this paper is supported in part by the U.S. National Science Foundation award CNS 1629856.

References

1. Bas, A., Smith, W.A.: What does 2d geometric information really tell us about 3d face shape? International Journal of Computer Vision **127**(10), 1455–1473 (2019) 4, 3
2. Bulat, A., Tzimiropoulos, G.: How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In: International Conference on Computer Vision (2017) 5.1
3. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: Facewarehouse: A 3d facial expression database for visual computing. IEEE Transactions on Visualization and Computer Graphics **20**(3), 413–425 (2013) 2.1
4. Chen, J., Tong, Y., Gray, W., Ji, Q.: A robust 3d eye gaze tracking system using noise reduction. In: Proceedings of the 2008 symposium on Eye tracking research & applications. pp. 189–196 (2008) 2.1
5. Chen, Q., Wu, H., Wada, T.: Camera calibration with two arbitrary coplanar circles. In: European Conference on Computer Vision. pp. 521–532. Springer (2004) 3.1
6. Cheng, Y., Huang, S., Wang, F., Qian, C., Lu, F.: A coarse-to-fine adaptive network for appearance-based gaze estimation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 10623–10630 (2020) 5.3
7. Fuhl, W.: From perception to action using observed actions to learn gestures. User Modeling and User-Adapted Interaction **31**(1), 105–120 (2021) 1
8. Fuhl, W., Santini, T., Kasneci, E.: Fast camera focus estimation for gaze-based focus control. arXiv preprint arXiv:1711.03306 (2017) 1
9. Fuhl, W., Santini, T., Reichert, C., Claus, D., Herkommer, A., Bahmani, H., Rifai, K., Wahl, S., Kasneci, E.: Non-intrusive practitioner pupil detection for unmodified microscope oculars. Computers in biology and medicine **79**, 36–44 (2016) 1
10. Funes Mora, K.A., Monay, F., Odobez, J.M.: Eyediap: A database for the development and evaluation of gaze estimation algorithms from rgb and rgb-d cameras. In: Proceedings of the symposium on eye tracking research and applications. pp. 255–258 (2014) 5.1, 4
11. Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schönborn, S., Vetter, T.: Morphable face models-an open framework. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 75–82. IEEE (2018) 2.1
12. Glass2, T.P.: tobii pro eye tracker data quality report (2017), `https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-glasses-2-accuracy-and-precision-test-report.pdf` 2.1, 3.1
13. Hennessey, C., Noureddin, B., Lawrence, P.: A single camera eye-gaze tracking system with free head motion. In: Proceedings of the 2006 symposium on Eye tracking research & applications. pp. 87–94 (2006) 2.1
14. Hutchinson, T.E., White, K.P., Martin, W.N., Reichert, K.C., Frey, L.A.: Human-computer interaction using eye-gaze input. IEEE Transactions on systems, man, and cybernetics **19**(6), 1527–1534 (1989) 1
15. Kohlbecher, S., Bardinst, S., Bartl, K., Schneider, E., Poitschke, T., Ablassmeier, M.: Calibration-free eye tracking by reconstruction of the pupil ellipse in 3d space. In: Proceedings of the 2008 symposium on Eye tracking research & applications. pp. 135–138 (2008) 3.1
16. Lai, C.C., Shih, S.W., Hung, Y.P.: Hybrid method for 3-d gaze tracking using glint and contour features. IEEE Transactions on Circuits and Systems for Video Technology **25**(1), 24–37 (2014) 2.1

17. Li, R., Bladin, K., Zhao, Y., Chinara, C., Ingraham, O., Xiang, P., Ren, X., Prasad, P., Kishore, B., Xing, J., Li, H.: Learning formation of physically-based face attributes (2020) 4
18. Li, T., Bolkart, T., Black, M.J., Li, H., Romero, J.: Learning a model of facial shape and expression from 4d scans. ACM Trans. Graph. **36**(6), 194–1 (2017) 2.1
19. Liu, Y., Liu, R., Wang, H., Lu, F.: Generalizing gaze estimation with outlier-guided collaborative adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3835–3844 (2021) 5.3
20. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C., Yong, M., Lee, J., et al.: Mediapipe: A framework for building perception pipelines. arxiv 2019. arXiv preprint arXiv:1906.08172 5.1
21. Palmero, C., Selva, J., Bagheri, M.A., Escalera, S.: Recurrent cnn for 3d gaze estimation using appearance and shape cues. arXiv preprint arXiv:1805.03064 (2018) 5.3
22. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3d face model for pose and illumination invariant face recognition. In: 2009 sixth IEEE international conference on advanced video and signal based surveillance. pp. 296–301. Ieee (2009) 2.1
23. Ploumpis, S., Ververas, E., O'Sullivan, E., Moschoglou, S., Wang, H., Pears, N., Smith, W.A., Gecer, B., Zafeiriou, S.: Towards a complete 3d morphable model of the human head. IEEE transactions on pattern analysis and machine intelligence **43**(11), 4142–4160 (2020) 1, 2.1, 2.2
24. Smith, B.A., Yin, Q., Feiner, S.K., Nayar, S.K.: Gaze locking: passive eye contact detection for human-object interaction. In: Proceedings of the 26th annual ACM symposium on User interface software and technology. pp. 271–280 (2013) 5.1, 4
25. Song, G., Cai, J., Cham, T.J., Zheng, J., Zhang, J., Fuchs, H.: Real-time 3d face-eye performance capture of a person wearing vr headset. In: Proceedings of the 26th ACM international conference on Multimedia. pp. 923–931 (2018) 2.1
26. Swirski, L., Dodgson, N.: A fully-automatic, temporal approach to single camera, glint-free 3d eye model fitting. Proc. PETMEI pp. 1–11 (2013) 2.1
27. Tsukada, A., Kanade, T.: Automatic acquisition of a 3d eye model for a wearable first-person vision device. In: Proceedings of the Symposium on Eye Tracking Research and Applications. pp. 213–216 (2012) 2.1
28. Vicente, F., Huang, Z., Xiong, X., De la Torre, F., Zhang, W., Levi, D.: Driver gaze tracking and eyes off the road detection system. IEEE Transactions on Intelligent Transportation Systems **16**(4), 2014–2027 (2015) 4, 5.3, 5.3
29. Wang, K., Ji, Q.: Real time eye gaze tracking with kinect. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 2752–2757. IEEE (2016) 2.1
30. Wang, K., Ji, Q.: Real time eye gaze tracking with 3d deformable eye-face model. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1003–1011 (2017) 2.2, 4, 5.3, 5.3
31. Wang, Y., Jiang, Y., Li, J., Ni, B., Dai, W., Li, C., Xiong, H., Li, T.: Contrastive regression for domain adaptation on gaze estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19376–19385 (2022) 5.3
32. Wood, E., Baltrušaitis, T., Morency, L.P., Robinson, P., Bulling, A.: A 3d morphable eye region model for gaze estimation. In: European Conference on Computer Vision. pp. 297–313. Springer (2016) 1, 2.1, 2.2, 4, 5.3, 5.3

33. Xiong, X., Liu, Z., Cai, Q., Zhang, Z.: Eye gaze tracking using an rgbd camera: A comparison with a rgb solution. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. pp. 1113–1121 (2014) 4, 5.3, 5.3
34. Zhou, X., Cai, H., Li, Y., Liu, H.: Two-eye model-based gaze estimation from a kinect sensor. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp. 1646–1653. IEEE (2017) 2.1
35. Zhu, H., Yang, H., Guo, L., Zhang, Y., Wang, Y., Huang, M., Shen, Q., Yang, R., Cao, X.: Facescape: 3d facial dataset and benchmark for single-view 3d face reconstruction. arXiv preprint arXiv:2111.01082 (2021) 5.1