EUROGRAPHICS 2020 / U. Assarsson and D. Panozzo
(Guest Editors)

Volume 39 (2020), Number 2

Accurate Real-time 3D Gaze Tracking Using a Lightweight Eyeball Calibration

Q. Wen[1] D. Bradley[2] T. Beeler[2] S. Park[3] O. Hilliges[3] J. Yong[1] F. Xu[1]

[1] BNRist and School of Software, Tsinghua University [2] DisneyResearch|Studios [3] ETH Zurich

Figure 1: *We present a real-time 3D gaze tracking method using a single consumer-level RGB camera, without the use of any infra-red illuminators. The estimated 3D gaze target is visualized as a red point from two additional views, which are not used for actual tracking. Our novel calibration method is "ground-truth position free", allowing for deployment in a diverse array of application scenarios.*

Abstract

3D gaze tracking from a single RGB camera is very challenging due to the lack of information in determining the accurate gaze target from a monocular RGB sequence. The eyes tend to occupy only a small portion of the video, and even small errors in estimated eye orientations can lead to very large errors in the triangulated gaze target. We overcome these difficulties with a novel lightweight eyeball calibration scheme that determines the user-specific visual axis, eyeball size and position in the head. Unlike the previous calibration techniques, we do not need the ground truth positions of the gaze points. In the online stage, gaze is tracked by a new gaze fitting algorithm, and refined by a 3D gaze regression method to correct for bias errors. Our regression is pre-trained on several individuals and works well for novel users. After the lightweight one-time user calibration, our method operates in real time. Experiments show that our technique achieves state-of-the-art accuracy in gaze angle estimation, and we demonstrate applications of 3D gaze target tracking and gaze retargeting to an animated 3D character.

CCS Concepts

• ***Human-centered computing*** → ***Interaction techniques;*** • ***Computing methodologies*** → ***Motion capture;***

1. Introduction

Estimating the gaze direction of a user has important applications in human-computer interaction [FWT*17], virtual and augmented reality [PSK*16], performance capture [WXY16], and attention analysis [PSL*16]. While recent advances have been made towards estimating user gaze from monocular images alone [WBZ*15, KKK*16], many of these systems are focused on determining the point of regard on a 2D display. This limits the utility of such approaches to public display scenarios [ZCM*15] or studies of aggregated attention over multiple people [SZB16, PLH17]. However, estimating 3D gaze, that is the exact location of the user's attention in arbitrary scenes, is a very challenging problem if only 2D images are available. Being able to estimate 3D gaze in real-time would have important implications for robotics (e.g. human-

robot collaboration [PRSS16]), computer graphics (e.g. foveated rendering [PSK*16], video editing [WBM*18]), and HCI (e.g. gaze typing [MWWM17], target selection in AR [KEP*18]).

Estimating the position of a 3D gaze target is challenging because it requires an accurate estimation of the user's head-pose, eye-ball location, and diameter and requires to estimate unobservable, person-dependent features such as the interocular distance and the offset between the optical and visual axes. Clearly, inferring all of these parameters from a single RGB image is challenging and so far this task does require the use of near-eye head-mounted trackers [MBWK16, WLLA], or high-end eye trackers using machine-vision cameras [WKHA18].

Embracing these challenges we propose a novel model-fitting based approach that leverages a lightweight calibration scheme in order to track the 3D gaze of users in realtime and with relatively high-accuracy. To achieve this we make several contributions. First, to attain estimates of person-dependent parameters such as eye-ball size and the offset of visual and optical axis, we propose a lightweight calibration procedure. Our approach does not require knowledge of the exact 3D position of gaze targets and only makes the assumption that the user indeed fixates a target while moving the head around to various positions and orientations.

Second, given the calibrated user parameters, we employ an optimization-based model-fitting approach to estimate the gaze in the current frame, taking into account the current head pose computed from a sparse set of detected landmarks, and then optimizing the 3D gaze via a photometric energy and a vergence constraint. However, while this formulation allows for the estimation of 3D gaze targets, we note that the accuracy of this approach is sensitive to several factors including perspective foreshortening (an object appears shorter than it actually is when angled towards the viewer). These can systematically affect both the head pose and the gaze angle estimates.

To compensate for these errors, our third contribution is a gaze refinement technique that aims to remove systematic bias in the gaze estimates. This correction can be formulated as a linear regression problem that maps the current head pose and estimated gaze angles to a gaze correction vector. Importantly, this mapping can be computed offline from data captured via ground truth 3D gaze targets from a multi-view detection of fiducial markers or from calibrated monitors. The mapping is then applied at runtime in a user-independent fashion and we experimentally show that it significantly improves the gaze estimation accuracy.

We demonstrate the efficacy of our method via a prototypical implementation of a real-time 3D gaze estimation system that takes only 2D imagery as input and only relies on a simple calibration routine for which the true 3D target positions do not need to be known. We demonstrate our method on three different applications in computer graphics: visualizing the 3D gaze target in a mixed reality setting, applying the 3D gaze to a virtual CG character in a performance capture scenario, and using the estimated 3D gaze target point to drive the gaze of multiple characters in an augmented reality application.

In summary, in this paper we contribute:

- A lightweight calibration method which does not require knowledge of 3D positions of gaze targets, but only the groupings of them per fixation, with variations in head pose.
- A model-based eye gaze tracker optimized with a vergence constraint which allows for 3D target tracking.
- A person-independent regression to refine model-fitting outputs to yield 3.45° leave-one-out error on a self-collected dataset.
- A prototypical system running at interactive rates (28.6fps).

2. Related work

Traditionally much work on gaze estimation has often involved the use of reflections (glints) from infra-red light sources [YKLC02, YC05, GE08], coupled with a zoom lens [VC08, HF12], multiple cameras [BBGB19, AGT15], or depth sensors [LL14, XLCZ14, SLS15, WJ16, FMO16]. For a comprehensive overview of gaze estimation methods we refer the reader to [HJ09]. We consider eye tracking in natural light scenarios in the absence of infra-red glints, with just a single monocular RGB camera used as input.

Gaze Estimation with Calibration In contrast to feature-based methods which suffer from poor extrapolation and restriction to the screen-space [SVC12, HKN*14], 3D model-based methods promise to allow free head movements and better robustness, as an understanding of the underlying 3D geometry can be applied. Many approaches define a simplified eyeball model involving intersecting spheres, where the gaze direction is defined as a ray originating from the center of a perfectly spherical eyeball, and going through the pupil center (also known as the optical axis of the eye). Such models allow for reasonable calibration-free gaze tracking [YUYA08, WBZ*15, WBM*16a, WBM*16b, WSXC16, WXY16, WXLY17], and can be extended to consider a user-specific angular difference between optical axis and line-of-sight (visual axis) as shown in [ME10, WJ17, WWJ16].

Required calibration samples for tuning a gaze estimator can be acquired explicitly or implicitly. An explicit procedure requires users to fixate on multiple points while measurements such as images of the eyes are collected. This arduous procedure can be simplified to require only one calibration point but only with the help of additional IR lights [GE08, VC08]. To reduce the burden on the end-user, much work has been done in the implicit personal calibration of 3D model-based tracking methods, by leveraging predicted visual saliency of stimuli [CJ11, CJ14, SMS12], analysing fixation maps [WWJ16], assuming binocular vergence on the used display device [ME10] and comparison against prior human gaze patterns on the same stimuli [AGVG13, LCS17]. These methods make strong assumptions of users' point of regards to certain stimuli (and thus may not generalize to new unseen stimuli or scenes), limit the gaze tracking to a 2D display, or use additional hardware (e.g. infra-red lights). In contrast, while our calibration scheme requires users to fixate on points in 3D space, there is no restriction on the target point position, and in particular we do not require knowledge of the exact fixation point location, and hence users can freely select easy-to-fixate stimuli or features. The ease of calibration does not come at significant cost to performance, with our calibrated eye tracking method being accurate and robust in particular to head movement.

While neural network based approaches have recently been

shown to perform particularly well on in-the-wild images [ZSFB15, KKK*16, ZSFB17, FCD18, PSH18, CLZ18, KRS*19], calibrating such methods with only few samples is challenging due to their reliance on neural networks with large number of parameters. Therefore, even personalized appearance based methods can suffer from relatively high errors in gaze [PZBH18, LYMO18], making them less suitable for computer graphics applications. Our method can outperform the current state-of-the-art in terms of angular accuracy, and is both effective and lightweight, running in real-time on a commodity CPU.

3D Point of Regard Estimation Given an accurate estimate of the line-of-sight of two eyes and their positions in 3D space, the two gaze rays are expected to intersect at a single point. The estimation of such 3D Point-of-Regard (PoR) is especially challenging, and thus far only performed with a high-end remote eye tracker [WKHA18] or head-mounted eye trackers [MBWK16, WLLA, MP08,TTTO14], in highly controlled settings. In addition, personal gaze of a crowd of people can be aggragated to estimate the unique target focused on by them [KKH*18]. To the best of our knowledge, we are the first to tackle the problem of 3D gaze target estimation of one person when using just a single commodity webcam as the input device. This is made possible by a binocular vergence constraint in our optimization formulation, which ensures that the gaze rays intersect in 3D space.

Eye Capture in Computer Graphics In the field of computer graphics, one area of recent interest has been high quality capture of eye shape [BBN*14, BBGB16], eyelids [BBK*15], and realistic data-driven eye rigging [BBGB19]. These methods aim for extreme realism in the application of performance capture, especially when coupled with robust face reconstruction methods [BBB*10, BHB*11, FHW*11, FNH*17]. However, the price of hyper-realism comes at the cost of complex offline reconstruction and tracking algorithms and dedicated capture setups. Looking more towards lightweight and real-time performance capture, several methods have coupled real-time facial tracking with online gaze estimation [WSXC16, WXY16], including real-time eyelid tracking [WXLY17]. Such applications are well suited for our advanced eye tracking approach. Another interesting application of eye capture is gaze editing in the scenario of video conferencing [KPB*12, CSBT03] or video editing [WBM*18]. Eye tracking is also important in the field of virtual reality [CKK19], where head-mounted displays typically occlude a large portion of the face and block important visual cues of where a user is looking. Many solutions for in-display gaze tracking have been proposed, and we refer to a recent survey of specialized devices [CAM18]. One example of tracking eyes in VR is gaze-aware facial re-enactment [TZS*18] (which has been subsequently also demonstrated on human portraits re-enactment [TZT*18]). Foveated rendering [PSK*16, WRK*16] is another field that relies on accurate gaze tracking, where computation cost is saved during image generation by synthesizing progressively less detail outside the eye fixation region. Furthermore, analyzing gaze patterns can provide insight into how users perceive 3D scenes [WKHA18] or television and film content [BH17]. Our technique for accurate real-time 3D gaze tracking could benefit all these methods including gaze editing, mixed reality, performance capture, foveated rendering and attention analysis.

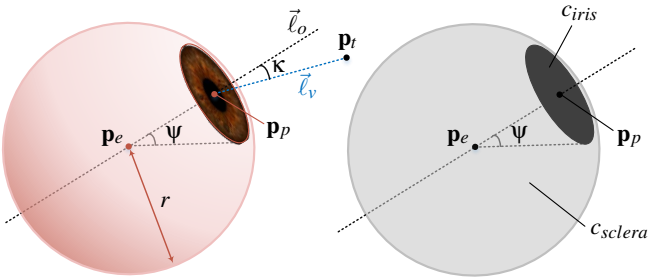

Figure 2: *The simplified 3D eyeball model in our tracking system. Left: geometry model. Right: binarized appearance model.*

3. Preliminaries

In this section, we introduce our proposed 3D eyeball model, which is comprised of a geometric component describing the shape and pose of the eyes as well as a photometric component describing its appearance. In addition we introduce the two constraints employed for fitting the model to image data, both during the proposed lightweight calibration to recover user specific eyeball parameters and during real-time 3D gaze tracking. The *vergence constraint* enforces the visual axes of the two eyes to intersect at a desired fixation point thus forming a valid gaze, while the *photometric consistency constraint* ensures that the appearance of the fit model matches the target images.

3.1. Eyeball Model

Human eyes exhibit very complex anatomical structure and motion, as described previously in work focusing on highly accurate modeling of eyes [BBN*14, BBGB16] and the oculomotor system [BBGB19]. To acquire such eye models they rely on a complex capture setup and involved processing, both not suited for our desired use-case of 3D gaze tracking from a single consumer RGB camera in real time. We hence employ a simplified eyeball model, comparable to previous model-based gaze tracking methods [WBM*16a, WBM*18, WSXC16, WXY16]. Then we add user-specific geometry parameters to the eyeball model and use the approximate appearance model proposed in [WXY16]. Further more, unlike previous work, we augment the model with a data-driven correction mechanism aiming to alleviate some of the approximation errors.

Eyeball Geometry. As shown in Fig. 2 (left), we approximate the shape of the eyeball with a sphere of radius r. The optical axis $\vec{\ell}_o$ of the eye is defined as the ray originating from the eyeball center $\mathbf{p}_e$ going through the center of the pupil $\mathbf{p}_p$, where $\mathbf{p}_e$ and $\mathbf{p}_p$ are defined with respect to the coordinate frame of the head. Without loss of generality, we define a local right-hand coordinate frame with origin at the center of the eyeball, its $y-$axis pointing upwards and the $z-$axis being collinear with the optical axis. Oftentimes, the optical axis is considered to correspond to the gaze direction, but it is well understood in ophthalmology that the gaze

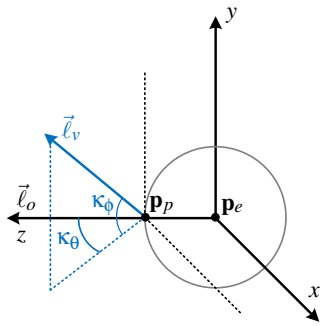

Figure 3: *Definitions of azimuthal angle κ_θ and polar angle κ_ϕ which are used to model κ (the angular offset between optical axis and visual axis).*

direction is collinear with the *visual axis* $\vec{\ell}_v$ instead, which deviates from the optical axis by κ degrees. While this angular offset typically is around 6 degrees in healthy eyes [AA11], it varies from subject to subject. Following [BBGB19] we exploit the symmetry of the eyes and model κ with a symmetric polar angle κ_ϕ and an antisymmetric azimuthal angle κ_θ (defined in Fig. 3), reducing the unknowns from four to two. The visual axis $\vec{\ell}_v = (\mathbf{p}_p, \mathbf{d}_v)$ is defined to originate at the center of the pupil $\mathbf{p}_p$ in the direction $\mathbf{d}_v = [\pm\cos\kappa_\phi \sin\kappa_\theta, \sin\kappa_\phi, \cos\kappa_\phi \cos\kappa_\theta]^T$, where the first component is positive for the right eye and negative for the left. The visual axes $(\vec{\ell}_v^L, \vec{\ell}_v^R)$ from the left and right eyes intersect at the fixation point $\mathbf{p}_t$, i.e. the point the person is looking at.

We approximate eye motion as simple rotation around a fixed pivot, i.e. the center of the eye $\mathbf{p}_e$. Since we use a simplified appearance model as described below, we consider only left-right (θ) and up-down (ϕ) rotation of the eye, and ignore rotation around the optical axis known as torsion. Since the pivots of the left and right eyes are selected relative to the coordinate frame of the head, which is defined to exhibit the same orientation as the eyeball coordinate frames when the eyes look forward (i.e. $\theta = 0$ and $\phi = 0$), the pose of the eyeball relative to the head is fully described by a transformation matrix $\mathbf{T}_{eye} = (R(\theta,\phi), \mathbf{p}_e)$, and premultiplying $\mathbf{T}_{eye}$ with the transformation of the head $\mathbf{T}_{head}$ relative to the world coordinate frame yields the absolute pose of the eyeball. For convenience we select our world coordinate frame to coincide with the coordinate frame of the camera, since we employ only a single camera.

Eyeball Appearance. We employ an approximate appearance model proposed by Wen et al. [WXY16], which offers a parametric description for the two most salient parts of the eye appearance, namely the iris surrounded by the whitish sclera. As depicted in Fig. 2 (right), the iris is approximated by a single circle centered at $\mathbf{p}_p$ and also encompasses the dark pupil. ψ is the angle between a point on the edge of the circle and the optical axis. The intensities of the iris (c_{iris}) and sclera (c_{sclera}) are estimated from the image data using a Gaussian mixture model. We select the mean of the two largest Gaussians as c_{sclera} and c_{iris}, assuming $c_{sclera} > c_{iris}$, i.e. the intensity of the sclera is assumed to be always brighter than the intensity of the iris. Unlike the angle ψ, which is estimated during calibration only, the two intensity values are updated at every frame, and thus good generalization for eyeball appearance variations caused by different subjects, head poses and illumination conditions can be achieved. With this parametric model, the intensity c_j at any given vertex j of the eye may be queried by evaluating by

$$c_j = \begin{cases} c_{sclera} & \gamma_j > \psi \\ c_{iris} & \gamma_j \leq \psi \end{cases} \tag{1}$$

where γ_j is the angle of the vertex j with respect to the optical axis.

3.2. Constraints

Vergence Constraint. Assuming the person is looking at a specific point in space, the visual axes have to intersect at this point, which imposes a strong constraint on the eye gaze which we refer to as *vergence constraint*. The constraint is formulated as

$$\|\mathbf{T}_{eye}^L \vec{\ell}_v^L, \mathbf{T}_{eye}^R \vec{\ell}_v^R\|_v^2 = 0. \tag{2}$$

The norm $\|\vec{\ell}_1, \vec{\ell}_2\|_v$ measures the distance of the two rays $\vec{\ell}_1$ and $\vec{\ell}_2$ at their closest point and is defined as

$$\|\vec{\ell}_1, \vec{\ell}_2\|_v = \frac{|(\mathbf{o}_1 - \mathbf{o}_2) \cdot (\mathbf{d}_1 \times \mathbf{d}_2)|}{|\mathbf{d}_1 \times \mathbf{d}_2|}, \tag{3}$$

where $\vec{\ell} = (\mathbf{o}, \mathbf{d})$, $\mathbf{o}$ is a point on the ray $\vec{\ell}$ (in our case, the pupil center $\mathbf{p}_p$), and $\mathbf{d}$ is the direction away from the eyeball.

Photometric Consistency Constraint. When projecting the 3D eye into the 2D image, the intensity c_j of a vertex j should match the image intensity at the projected location, and hence we formulate a *photometric consistency constraint as*

$$\|c_j - \mathcal{I}(\pi(\mathbf{T}_{head}\mathbf{T}_{eye}\mathbf{x}_j))\|_2^2 = 0, \tag{4}$$

where $\mathbf{x}_j$ denotes the 3D position of vertex j, $\pi(\cdot)$ denotes the projection operator, and $\mathcal{I}(\cdot)$ samples the intensity of image $\mathcal{I}$ at the provided location.

4. 3D Model-based Gaze Tracking

Gaze tracking aims to recover suitable eyeball parameters (Section 3.1) from input imagery, leveraging the constraints defined in Section 3.2. While some parameters, such as the eyeball rotation defined by (θ, ϕ), might change over time and hence require to be estimated continuously (Section 4.3), others remain fixed for a given person and it is sufficient to estimate them once. We refer to this initial estimation of all user-specific parameters as calibration, which aggregates information over a short period of time as described in Section 4.2.

4.1. 3D Face Tracking

As a first step we need to recover the head pose $\mathbf{T}_{head}$ since our eyeball model is formulated relative to the head of the person. This is achieved using established model-based facial tracking algorithms. Specifically, we employ the system proposed by [CWZ*13]. The method employs a multilinear face model [CWZ*13] in combination with a 2D facial landmark detector [Kin09] to jointly fit identity, expression, and pose parameters by minimizing the difference between the detected 2D facial landmarks and the projected 3D facial landmarks. Since the 2D landmark detection takes place per

frame, they tend to exhibit temporal noise, and hence we smooth the head transforms temporally prior to eye gaze fitting. This is achieved via a simple, constant motion model prediction of the current head pose (rotation and translation), based on the two previous head poses. The average of the predicted current pose and the pose obtained by the tracker is used as the final head pose.

4.2. Lightweight Eyeball Calibration

In an initial calibration step, we ask the user to focus on a small number of calibration target points in order to estimate the eyeball radius r, eyeball centers $\mathbf{p}_e^{L,R}$, the angular shift between the optical axis and the visual axis $(\kappa_\theta, \kappa_\phi)$, and the iris angle ψ. These parameters are constant over all frames for a given person and will be referred to as μ. In addition to these time-invariant parameters, there is a set of time-varying parameters which we denote by τ, namely the per frame eyeball rotations $(\theta,\phi)_f^{L,R}$. Notice that the colors $(c_{sclera}, c_{iris})_f^{L,R}$ are also time varying parameters, but as they are estimated by the method introduced in [WXY16], we do not include them in τ.

To perform calibration, most previous methods rely on known calibration target points. This, however, is quite a strong requirement, since obtaining ground truth 3D points is a difficult task in itself, especially in a monocular setting, and hence not suited for unconstrained and user-friendly calibration. Instead, we propose a much more convenient and less constrained calibration procedure, where the user is asked to focus on a *specific yet unknown* point in the scene and move his head around (rotate and translate) while maintaining focus.

If the focus point remains constant over time, the visual axes of both eyes for all frames have to intersect at the target location:

$$\sum_{f\in\mathcal{F}}\sum_{f'\in\mathcal{F}} \|\mathbf{T}_{head,f}\mathbf{T}^L_{eye,f}\vec{\ell}^L_v, \mathbf{T}_{head,f'}\mathbf{T}^R_{eye,f'}\vec{\ell}^R_v\|^2_v = 0, \quad (5)$$

where $\mathcal{F}$ is the set of $|\mathcal{F}|$ frames. This equation can be considered as an extension of the single frame vergence constraint in Eq. (2).

In practice, we are not limited to a single target point, but can ask the user to sequentially focus onto a number of different scene points Ω, yielding a set of frames $\mathcal{F}_\omega$ per target point $\omega\in\Omega$. Combined with the photometric consistency term defined in (4), the overall energy to be minimized is defined as

$$\begin{aligned}\arg\min_{\mu,\tau} &\sum_{\omega\in\Omega}\sum_{f\in\mathcal{F}_\omega}\sum_{j\in\mathcal{V}_f} \|c_j - \mathcal{I}_f(\pi(\mathbf{T}_{head,f}\mathbf{T}_{eye,f}\mathbf{x}_j))\|^2_2 \\ &+\lambda\sum_{\omega\in\Omega}\sum_{f\in\mathcal{F}_\omega}\sum_{f'\in\mathcal{F}_\omega} \|\mathbf{T}_{head,f}\mathbf{T}^L_{eye,f}\vec{\ell}^L_v, \mathbf{T}_{head,f'}\mathbf{T}^R_{eye,f'}\vec{\ell}^R_v\|^2_v,\end{aligned} \quad (6)$$

where we jointly minimize over both the set of time-invariant μ and time-variant parameters τ. $\mathcal{V}_f$ denotes the set of visible vertices on the eyeball model at frame f (defined by the area enclosed by 2D eye landmarks detected in Section 4.1), and λ is a balancing parameter set to 10^7 in our implementation.

4.3. Online 3D Gaze Tracking

Once calibrated, the time-invariant parameters μ are fixed and we estimate the time-varying parameters τ for every frame f by minimizing

$$\arg\min_{\tau} \sum_{j\in\mathcal{V}_f} \|c_j - \mathcal{I}_f(\pi(\mathbf{T}_{head,f}\mathbf{T}_{eye,f}\mathbf{x}_j))\|^2_2 + \lambda\|\mathbf{T}^L_{eye,f}\vec{\ell}^L_v, \mathbf{T}^R_{eye,f}\vec{\ell}^R_v\|^2_v. \quad (7)$$

Once more, the energy is a composition of the vergence and photometric consistency constraints introduced in Section 3.2.

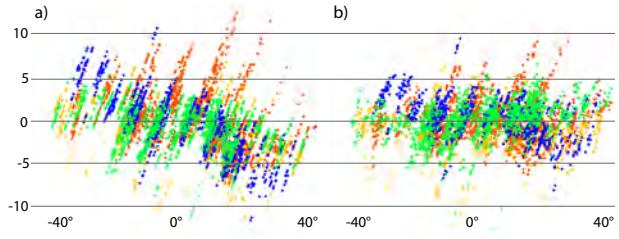

Figure 4: *Here we plot the errors relative to the left-right rotation of the eye (θ) for four different subjects. a) Due to modeling and approximation errors, the residuals exhibit a systematic error, leading to a slanted distribution. b) The proposed data-driven gaze correction attenuates the systematic error leading to a mostly zero-mean distribution of the residuals.*

4.4. Data-Driven Gaze Correction

Due to the approximations in our geometry and appearance model, such as assuming a spherical eyeball without corneal bulge, the gaze estimation is bound to exhibit systematic errors. This can be observed in Fig. 4 (a). We propose to address this systematic error via a data-driven correction scheme. To this end we employ a linear regression since it is stable, generalizes well and can be evaluated with almost no overhead, which is important for the envisioned real-time usecase. As depicted visually in Fig. 4 (b) and assessed quantitatively in the next section, the proposed regression succeeds at correcting for the systematic errors and as a consequence substantially improves the estimated gaze.

Our 3D gaze tracking technique relies heavily on the appearance of the eyes in the images, especially the region around the limbus [WXY16]. Due to the perspective projection model of a pinhole camera, the projected 2D shape of the limbus is mainly determined by the 3D pose of the eyeball relative to the camera. This in turn, is essentially defined by the translation $\mathbf{t}_{head}$ of the head in conjunction with the eyeball rotation defined by the gaze itself $\mathbf{g} = (\theta^L, \phi^L, \theta^R, \phi^R)$. We hence learn a linear mapping from these seven-dimensional input features to a four-dimensional gaze correction vector $\delta = (\delta_{\theta^L}, \delta_{\phi^L}, \delta_{\theta^R}, \delta_{\phi^R})$. This mapping is given by:

$$\delta = \mathbf{A}(\mathbf{t}_{head}, \mathbf{g})^{\mathrm{T}} + \mathbf{b}, \quad (8)$$

with $\mathbf{A}\in\mathbb{R}^{4\times 7}$ and $\mathbf{b}\in\mathbb{R}^4$.

We found that the data-driven gaze correction is largely user-independent. This confirms our assumption that the observed, uncorrected errors are due to modeling errors. This observation allows us to train the regression offline, and to use it for entirely unseen users. Furthermore, given the simplicity of the model there

is practically no overhead to the per user calibration introduced in Section 4.2. In the next section, we provide a thorough analysis of the proposed 3D gaze tracking technique and the data-driven gaze correction strategy.

5. Results

In this section, we provide a thorough assessment of the two key novelties presented in this paper, the lightweight calibration and the data-driven gaze correction. We further assess the quality of the overall system by comparing to a state-of-the-art CNN-based method, trained and tested on the same data. Unlike prior art, which typically reports errors on the estimated gaze angles only, we also analyse the effectiveness of the proposed method for tracking points in 3D space, both quantitatively by reporting metric accuracy and qualitatively by visualizing the tracked focus point on the input videos. Lastly, we demonstrate retargeting of the captured gaze onto 3D avatars hinting at usecases in the realm of VR or AR.

5.1. Setup

We use a consumer webcamera (Logitech C920) to capture 1080p RGB images for both the eyeball calibration and the gaze tracking. 9 subjects (6 males and 3 females) participate in the experiments. To calibrate the eyeball, each subject is asked to sequentially fixate 4 target points displayed on a monitor while translating and rotating their head. For these points we do not know the ground truth positions. To evaluate the system and to train the regression, we displayed a grid of 5×3 target points on a monitor for all the 9 subjects, and additionally positioned a floating target (a 2cm cube with AR markers) at various locations inside a 600mm×400mm×300mm volume for 3 of the subjects. For the 3D scene targets we measure ground truth positions using a calibrated multi-camera setup, and to measure the ground truth positions of the 2D screen targets we capture a mirror with marker tags in multiple poses, again providing multi-view geometry in order to triangulate the 2D scene targets in 3D. The head pose statistics in our dataset are listed in Tab.1.

Table 1: *Statistics of head translation (mm) and rotation (degrees) in our dataset and [WJ17]. The head poses in our dataset are with respect to the world space (i.e. camera space) and the mean and standard deviation values are computed across all the 9 subjects.*

	Ours			[WJ17]		
	Mean		Std. Dev.	Mean		Std. Dev.
x	5.4	±	94.9	-30.0	±	32.0
y	-45.1	±	35.1	-11.0	±	27.0
z	643.6	±	52.4	525.0	±	39.0
yaw	-1.5	±	16.5	3.0	±	19.5
pitch	5.0	±	5.8	2.8	±	12.7
roll	1.9	±	3.0	-82.0	±	7.1

Performance We deployed our system on a computer with an Intel Core i7-4790 CPU (3.6 GHz) and 32 GB memory. For each input frame in the tracking stage, the optimizations for both fitting the face and tracking the gaze are performed by the Ceres solver [AMO]. The facial landmark detection takes about 6ms, the face fitting 12ms and the gaze tracking 17ms. The gaze regression adds almost no overhead to the tracking pipeline (less than 1ms), as it is a low-dimensional linear regression model. In general, our system takes about 35ms for the gaze tracking in each frame.

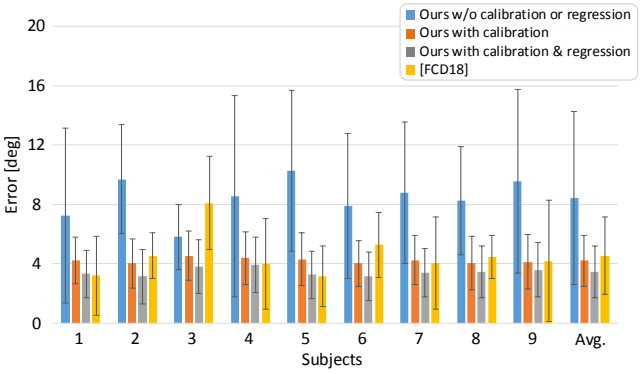

Figure 5: *Tracking errors of all 9 subjects with and without calibration or regression modules. Blue: calibration-free tracking (average error:* 8.45°*). Orange: tracking with user-specific eyeball parameters but without the gaze regression (average error:* 4.21°*). Grey: our full method (average error:* 3.45°*). Yellow: a state-of-the-art person-independent CNN baseline [FCD18] (average error:* 4.55°*).*

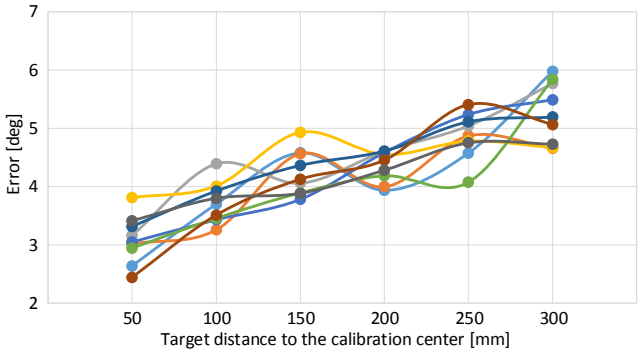

Figure 6: *Tracking errors of the evaluation targets relative to their distances to the calibration center (the average position of the 4 calibration targets). Each color represents one subject.*

5.2. Evaluation

Eyeball Calibration. First we evaluate the effectiveness of the proposed lightweight eyeball calibration. To show the importance of calibrating for user-specific eyeball parameters, we compare our calibrated parameters with the average eyeball parameters of all the 9 subjects in gaze tracking. As shown in Fig. 5, calibration substantially increases the quality of gaze tracking and reduces the overall error by a factor of 2 from 8.45° to 4.21°s on average.

In our calibration procedure, we use four calibration targets on a screen arranged in a square with a diagonal of 200mm. It is to be expected that accuracy will be highest within this square and

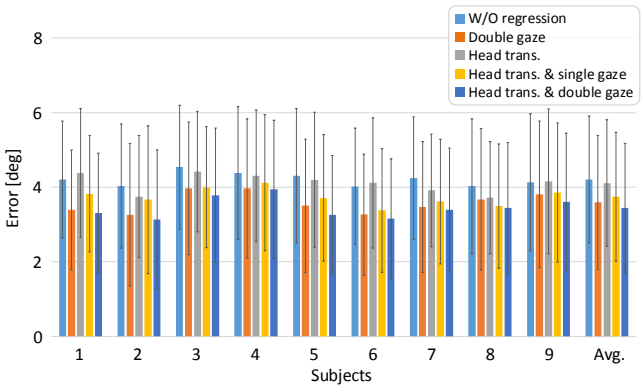

Figure 7: *Tracking errors of different regression inputs. Light blue: without regression. Orange: the gaze of both eyes. Gray: head translation only. Yellow: head translation and the gaze of one eye, i.e each eye use its gaze separately. Dark blue: our full method, head translation and the gaze of both eyes*

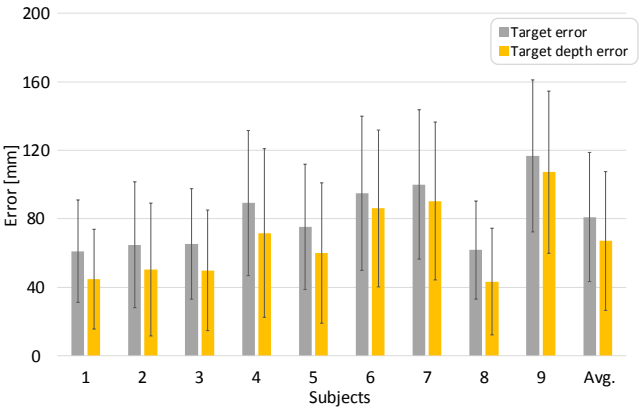

Figure 8: *Tracking errors of the 9 subjects using our full method in terms of target errors. Gray: target errors. Yellow: target depth errors along the visual axes.*

degrade as the method starts to extrapolate. Fig. 6 validates this assumption and demonstrates that accuracy is best within the square ($\sim 3.4^\circ$) and degrades gracefully the further the targets are away from the center of the calibration square. To improve the calibration and the tracking for 3D, we can use calibration targets away from the screen, for example by placing objects in front of the screen without the need to know the exact 3D locations of them.

Data-driven Gaze Correction To assess the effectiveness and the generality of the proposed linear gaze regression we train and test several regressions in a leave-one-out manner. We train a regressor using the evaluation data of 8 subjects and test its performance on the evaluation data of the remaining subject. From Fig. 5, we see that for all the subjects, the regression reduces the tracking errors noticeably (an average reduction of 0.76°). Furthermore, since the data of the left-out subject is not used in the training, the experiments indicate good generalization behaviour and hence the regression can be pre-trained once in an offline process and applied to novel users.

We further evaluate the selection of the input features to the regression in Fig. 7. As was to be expected, the gaze of the two eyes plays the main role in the regression. Head translation cannot be used in isolation to refine the tracking, but it helps significantly to improve the performance in conjunction with the gaze. This supports our intuition that the pose of the eye relative to the camera is critical since it defines the appearance of the eye in the imagery. Interestingly, regressing the eye gaze of the left and right eyes jointly is very beneficial, presumably since gaze estimation is coupled via the vergence constraint during tracking.

5.3. Comparison to Others

Our calibration method requires very little effort from the user, and more importantly does not require any knowledge of the exact calibration point position. This allows for arbitrary targets in either 2D (screen) or 3D (space) to be selected by the user. Despite this flexibility, our approach yields better performance than a state-of-the-art person-independent gaze estimator, which has significantly higher model capacity. We demonstrate this by training the current state-of-the-art architecture, the RT-GENE convolutional neural network (CNN) [FCD18] on our data, with pre-processing steps following [ZSB18] [†] . We modify the network to share model parameters across the two eye image encoders, as we find that this improves performance. The CNN is trained using the Adam optimizer [KB14] with a learning rate of 5×10^{-4} and batch size of 64 for 50 epochs, with an exponential learning rate decay schedule [‡] . The RT-GENE method evaluated is a single model (not ensembled), where the learning parameters were tuned carefully to make the implementation competitive. We show our results in Fig. 5 where our method not only performs 24.2% better than the CNN method on average, but also more consistently across subjects with a standard deviation of 0.27° (vs 1.48° for the CNN method). We can see that when evaluated on our evaluation dataset, our method is more accurate and robust compared to the best performing learning-based approach.

The most comparable and state-of-the-art model-based gaze-estimation method to ours is that of [WJ17], which runs in real-time and requires only a commodity RGB camera. They report 3.5° of error over their 10 subjects, and we achieve 3.45° of error over our 9 subjects. Please note that we evaluate our method using a completely different dataset from [WJ17], although we try our best to collect a dataset comparable with theirs (Tab. 1). In addition, their approach is calibrated using known ground-truth calibration target positions in 3D, while our light-weight calibration does not have this requirement.

[†] We use [HR17] for face detection and [DZCZ18] for facial landmark detection as this produces more consistent eye image patches.

[‡] Decay with a multiplier of 0.1 every 15 epochs.

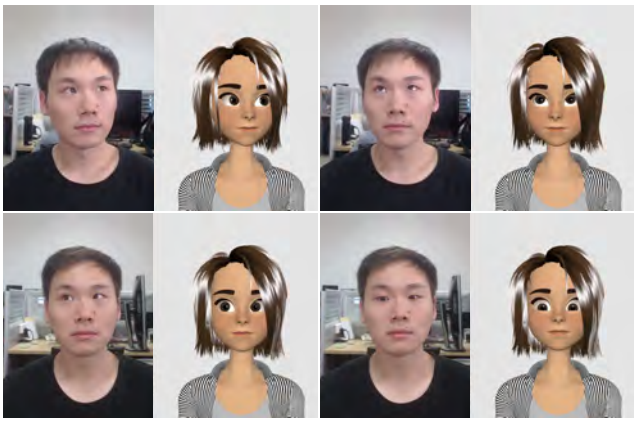

Figure 9: *Some frames selected from the gaze retargeting results.*

Figure 10: *Two frames selected from our augmented-reality application.*

5.4. Discussions and Applications

We now unpack the 3D gaze tracking capabilities of our method and hint at potential applications via some qualitative results. Results in terms of 3D target accuracies are shown in Fig. 8, where we achieve an average Euclidean error of 80.96mm. While this error is not neglible, please note that without our proposed vergence constraint, the two eyes' gaze rays are not guaranteed to converge to a single 3D point at all, making this task impossible for existing learning-based approaches. Even small errors in estimating the visual axis can lead to very large errors in the estimated depth or z-value of the triangulated gaze target. Therefore, we further estimate the depth error of the estimated gaze target along the gaze direction. This z-error is shown in Fig. 8, where we can see that about 80% of the target errors come from inaccurate depth estimation (67.03mm on average in our experiments).

In addition, we show three sequence results in the supplementary video: (a) visualizing the 3D gaze target, (b) applying the 3D gaze to a virtual CG character in a performance capture scenario, and (c) using the 3D target to drive the gaze of multiple 3D characters in the same scene as the actor, demonstrating an augmented-reality application. Some selected frames are shown in Fig. 1, Fig. 9 , and Fig. 10 for (a), (b), and (c) respectively. In our 3D gaze target tracking sequences, we can see that despite the difficulty of the task, our tracking works well albeit with some jitter. This is certainly exaggerated by small fluctuations in the estimated gaze directions of the two eyes, and the distance between eyeball and gaze target. For the gaze retargeting result, the current automatic solution gives visually pleasing result which can serve as a good initialization for animators to generate vivid facial animations. And the result can be further refined following [DJL*15, DJA*16, DJS*19] to avoid exaggerated eye movements.

6. Future Work and Conclusions

In this paper we propose a 3D gaze estimation method that takes only monocular images as input. However, the technique is not without limitations and the method could be further improved. For example, we noticed that the gaze error increases in frames with poor eyelid detection accuracy or in frames where the iris is less exposed than usually. Detecting and handling such cases explicitly would significantly improve overall accuracy and robustness. A further interesting direction for future work would be a more comprehensive dataset both for calibration (and potential training of data-driven methods) and a set evaluation procedure for the novel task of monocular 3D gaze estimation. However, collecting and labeling of such a dataset is beyond the scope of this paper.

In conclusion, in this paper we have proposed a real-time 3D gaze tracking technique using a single RGB imagery only. First, we propose a novel lightweight calibration method which accurately estimates user-specific parameters. Compared to the traditional techniques, our method does not require known 3D positions of the gaze targets for calibration, and thus can be performed by end users. Furthermore, we propose an online gaze tracking method and a linear regression to reduce systematic errors caused by modeling assumptions. The system runs in real time and the regression correction can be trained once and applied to new users.

We show experimentally that our technique yields high-accuracy in gaze estimation, comparable or better than state-of-the-art appearance based methods (which can not easily be modified to the 3D task). Finally, we demonstrate visual results on 3D gaze target tracking and apply our technique to gaze retargeting with visually pleasing results.

Acknowledgement

This work is supported by the National Key R&D Program of China (2018YFA0704000), NSFC (No.61822111, 61727808, 61671268, 61672307) and Beijing Natural Science Foundation (JQ19015, L182052). J. Yong and F. Xu are the corresponding authors.

References

[AA11] AVUDAINAYAGAM K. V., AVUDAINAYAGAM C. S.: Simple method to measure the visual axis of the human eye. *Optics letters 36*, 10 (2011), 1803–1805. 4

[AGT15] ARAR N. M., GAO H., THIRAN J.-P.: Robust gaze estimation based on adaptive fusion of multiple cameras. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (2015), vol. 1, pp. 1–7. 2

[AGVG13] ALNAJAR F., GEVERS T., VALENTI R., GHEBREAB S.: Calibration-free gaze estimation using human gaze patterns. In *Proceedings of the IEEE international conference on computer vision* (2013), pp. 137–144. 2

[AMO] AGARWAL S., MIERLE K., OTHERS: Ceres solver. http://ceres-solver.org. 6

[BBB*10] BEELER T., BICKEL B., BEARDSLEY P., SUMNER B., GROSS M.: High-quality single-shot capture of facial geometry. In *ACM Transactions on Graphics (ToG)* (2010), vol. 29, p. 40. 3

[BBGB16] BÉRARD P., BRADLEY D., GROSS M., BEELER T.: Lightweight eye capture using a parametric model. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 117. 3

[BBGB19] BÉRARD P., BRADLEY D., GROSS M., BEELER T.: Practical person-specific eye rigging. In *Computer Graphics Forum* (2019), vol. 38, pp. 441–454. 2, 3, 4

[BBK*15] BERMANO A., BEELER T., KOZLOV Y., BRADLEY D., BICKEL B., GROSS M.: Detailed spatio-temporal reconstruction of eyelids. *ACM Transactions on Graphics (TOG) 34*, 4 (2015), 44. 3

[BBN*14] BÉRARD P., BRADLEY D., NITTI M., BEELER T., GROSS M.: High-quality capture of eyes. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 33*, 6 (2014), 223:1–223:12. 3

[BH17] BREEDEN K., HANRAHAN P.: Gaze data for the analysis of attention in feature films. *ACM Trans. Appl. Percept. 14*, 4 (2017), 23:1–23:14. 3

[BHB*11] BEELER T., HAHN F., BRADLEY D., BICKEL B., BEARDSLEY P., GOTSMAN C., SUMNER R. W., GROSS M.: High-quality passive facial performance capture using anchor frames. In *ACM Transactions on Graphics (TOG)* (2011), vol. 30, p. 75. 3

[CAM18] COGNOLATO M., ATZORI M., MÜLLER H.: Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. *Journal of Rehabilitation and Assistive Technologies Engineering 5* (2018), 1–13. 3

[CJ11] CHEN J., JI Q.: Probabilistic gaze estimation without active personal calibration. In *CVPR 2011* (2011), pp. 609–616. 2

[CJ14] CHEN J., JI Q.: A probabilistic approach to online eye gaze tracking without explicit personal calibration. *IEEE Transactions on Image Processing 24*, 3 (2014), 1076–1086. 2

[CKK19] CLAY V., KÖNIG P., KÖNIG S.: Eye tracking in virtual reality. *Journal of Eye Movement Research 12*, 1 (2019). 3

[CLZ18] CHENG Y., LU F., ZHANG X.: Appearance-based gaze estimation via evaluation-guided asymmetric regression. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 100–115. 3

[CSBT03] CRIMINISI A., SHOTTON J., BLAKE A., TORR P. H. S.: Gaze manipulation for one-to-one teleconferencing. In *Proc. ICCV* (2003). 3

[CWZ*13] 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 (2013), 413–425. 4

[DJA*16] DUCHOWSKI A. T., JÖRG S., ALLEN T. N., GIANNOPOULOS I., KREJTZ K.: Eye movement synthesis. In *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications* (2016), ACM, pp. 147–154. 8

[DJL*15] DUCHOWSKI A., JÖRG S., LAWSON A., BOLTE T., ŚWIRSKI L., KREJTZ K.: Eye movement synthesis with 1/f pink noise. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games* (2015), ACM, pp. 47–56. 8

[DJS*19] DUCHOWSKI A. T., JÖRG S., SCREWS J., GEHRER N. A., SCHÖNENBERG M., KREJTZ K.: Guiding gaze: expressive models of reading and face scanning. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (2019), ACM, p. 25. 8

[DZCZ18] DENG J., ZHOU Y., CHENG S., ZAFERIOU S.: Cascade multi-view hourglass model for robust 3d face alignment. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (2018), pp. 399–403. 7

[FCD18] FISCHER T., CHANG H. J., DEMIRIS Y.: RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments. In *ECCV* (2018). 3, 6, 7

[FHW*11] FYFFE G., HAWKINS T., WATTS C., MA W.-C., DEBEVEC P.: Comprehensive Facial Performance Capture. In *Eurographics* (2011). 3

[FMO16] FUNES-MORA K. A., ODOBEZ J.-M.: Gaze estimation in the 3d space using rgb-d sensors. *IJCV 118*, 2 (Jun 2016), 194–216. 2

[FNH*17] FYFFE G., NAGANO K., HUYNH L., SAITO S., BUSCH J., JONES A., LI H., DEBEVEC P.: Multi-View Stereo on Consistent Face Topology. *Comput. Graph. Forum 36*, 2 (2017), 295–309. 3

[FWT*17] FEIT A. M., WILLIAMS S., TOLEDO A., PARADISO A., KULKARNI H., KANE S. K., MORRIS M. R.: Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *CHI* (2017), pp. 1118–1130. 1

[GE08] GUESTRIN E. D., EIZENMAN M.: Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (2008), pp. 267–274. 2

[HF12] HENNESSEY C., FISET J.: Long range eye tracking: bringing eye tracking into the living room. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (2012), pp. 249–252. 2

[HJ09] HANSEN D. W., JI Q.: In the eye of the beholder: A survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence 32*, 3 (2009), 478–500. 2

[HKN*14] HUANG M. X., KWOK T. C., NGAI G., LEONG H. V., CHAN S. C.: Building a self-learning eye gaze model from user interaction data. In *Proceedings of the 22Nd ACM International Conference on Multimedia* (New York, NY, USA, 2014), MM '14, ACM, pp. 1017–1020. 2

[HR17] HU P., RAMANAN D.: Finding tiny faces. In *CVPR* (2017). 7

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014). 7

[KEP*18] KYTÖ M., ENS B., PIUMSOMBOON T., LEE G. A., BILLINGHURST M.: Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 81. 2

[Kin09] KING D. E.: Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research 10* (2009), 1755–1758. 4

[KKH*18] KODAMA Y., KAWANISHI Y., HIRAYAMA T., DEGUCHI D., IDE I., MURASE H., NAGANO H., KASHINO K.: Localizing the gaze target of a crowd of people. In *Asian Conference on Computer Vision* (2018), Springer, pp. 15–30. 3

[KKK*16] KRAFKA K., KHOSLA A., KELLNHOFER P., KANNAN H., BHANDARKAR S., MATUSIK W., TORRALBA A.: Eye Tracking for Everyone. In *CVPR* (2016). 1, 3

[KPB*12] KUSTER C., POPA T., BAZIN J.-C., GOTSMAN C., GROSS M.: Gaze correction for home video conferencing. *ACM Trans. Graph. (Proc. of ACM SIGGRAPH ASIA) 31*, 6 (2012). 3

[KRS*19] KELLNHOFER P., RECASENS A., STENT S., MATUSIK W., TORRALBA A.: Gaze360: Physically unconstrained gaze estimation in the wild. In *Proceedings of the IEEE International Conference on Computer Vision* (2019), pp. 6912–6921. 3

[LCS17] LU F., CHEN X., SATO Y.: Appearance-based gaze estimation via uncalibrated gaze pattern recovery. *IEEE Transactions on Image Processing 26*, 4 (2017), 1543–1553. 2

[LL14] LI J., LI S.: Eye-model-based gaze estimation by rgb-d camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2014), pp. 592–596. 2

[LYMO18] LIU G., YU Y., MORA K. A. F., ODOBEZ J.: A differential approach for gaze estimation with calibration. In *BMVC* (2018). 3

[MBWK16] MLOT E. G., BAHMANI H., WAHL S., KASNECI E.: 3d gaze estimation using eye vergence. In *HEALTHINF* (2016), pp. 125–131. 2, 3

[ME10] MODEL D., EIZENMAN M.: An automatic personal calibration procedure for advanced gaze estimation systems. *IEEE Transactions on Biomedical Engineering 57*, 5 (2010), 1031–1039. 2

[MP08] MUNN S. M., PELZ J. B.: 3d point-of-regard, position and head orientation from a portable monocular video-based eye tracker. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (2008), pp. 181–188. 3

[MWWM17] MOTT M. E., WILLIAMS S., WOBBROCK J. O., MORRIS M. R.: Improving dwell-based gaze typing with dynamic, cascading dwell times. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), ACM, pp. 2558–2570. 2

[PLH17] PAPOUTSAKI A., LASKEY J., HUANG J.: Searchgazer: Webcam eye tracking for remote studies of web search. In *CHIIR* (2017). 1

[PRSS16] PALINKO O., REA F., SANDINI G., SCIUTTI A.: Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2016), IEEE, pp. 5048–5054. 2

[PSH18] PARK S., SPURR A., HILLIGES O.: Deep pictorial gaze estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 721–738. 3

[PSK*16] PATNEY A., SALVI M., KIM J., KAPLANYAN A., WYMAN C., BENTY N., LUEBKE D., LEFOHN A.: Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph. 35*, 6 (2016), 179:1–179:12. 1, 2, 3

[PSL*16] PAPOUTSAKI A., SANGKLOY P., LASKEY J., DASKALOVA N., HUANG J., HAYS J.: Webgazer: Scalable webcam eye tracking using user interactions. In *IJCAI* (2016), pp. 3839–3845. 1

[PZBH18] PARK S., ZHANG X., BULLING A., HILLIGES O.: Learning to find eye region landmarks for remote gaze estimation in unconstrained settings. In *ACM ETRA* (2018). 3

[SLS15] SUN L., LIU Z., SUN M.-T.: Real time gaze estimation with a consumer depth camera. *Information Sciences 320* (2015), 346–360. 2

[SMS12] SUGANO Y., MATSUSHITA Y., SATO Y.: Appearance-based gaze estimation using visual saliency. *IEEE transactions on pattern analysis and machine intelligence 35*, 2 (2012), 329–341. 2

[SVC12] SESMA L., VILLANUEVA A., CABEZA R.: Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (New York, NY, USA, 2012), ETRA '12, ACM, pp. 217–220. 2

[SZB16] SUGANO Y., ZHANG X., BULLING A.: Aggregaze: Collective estimation of audience attention on public displays. In *29th Annual Symposium on User Interface Software and Technology* (2016), UIST '16, pp. 821–831. 1

[TTTO14] TAKEMURA K., TAKAHASHI K., TAKAMATSU J., OGASAWARA T.: Estimating 3-d point-of-regard in a real environment using a head-mounted eye-tracking system. *IEEE Transactions on Human-Machine Systems 44*, 4 (2014), 531–536. 3

[TZS*18] THIES J., ZOLLHÖFER M., STAMMINGER M., THEOBALT C., NIESSNER M.: Facevr: Real-time gaze-aware facial reenactment in virtual reality. *ACM Transactions on Graphics (TOG) 37*, 2 (2018), 25. 3

[TZT*18] THIES J., ZOLLHÖFER M., THEOBALT C., STAMMINGER M., NIESSNER M.: Headon: Real-time reenactment of human portrait videos. *ACM Trans. Graph. 37*, 4 (2018), 164:1–164:13. 3

[VC08] VILLANUEVA A., CABEZA R.: A novel gaze estimation system with one calibration point. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 38*, 4 (2008), 1123–1138. 2

[WBM*16a] 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* (2016), pp. 297–313. 2, 3

[WBM*16b] WOOD E., BALTRUŠAITIS T., MORENCY L.-P., ROBINSON P., BULLING A.: Learning an appearance-based gaze estimator from one million synthesised images. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (2016), pp. 131–138. 2

[WBM*18] WOOD E., BALTRUŠAITIS T., MORENCY L.-P., ROBINSON P., BULLING A.: Gazedirector: Fully articulated eye gaze redirection in video. In *Computer Graphics Forum* (2018), vol. 37, pp. 217–225. 2, 3

[WBZ*15] WOOD E., BALTRUAITIS T., ZHANG X., SUGANO Y., ROBINSON P., BULLING A.: Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)* (Washington, DC, USA, 2015), ICCV '15, IEEE Computer Society, pp. 3756–3764. 1, 2

[WJ16] WANG K., JI Q.: Real time eye gaze tracking with kinect. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (2016), pp. 2752–2757. 2

[WJ17] 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* (2017), pp. 1003–1011. 2, 6, 7

[WKHA18] WANG X., KOCH S., HOLMQVIST K., ALEXA M.: Tracking the gaze on objects in 3d: How do people really look at the bunny? *ACM Trans. Graph. 37*, 6 (2018), 188:1–188:18. 2, 3

[WLLA] WANG X., LINDLBAUER D., LESSIG C., ALEXA M.: Accuracy of monocular gaze tracking on 3d geometry. In *Workshop on Eye Tracking and Visualization (ETVIS) co-located with IEEE VIS*. 2, 3

[WRK*16] WEIER M., ROTH T., KRUIJFF E., HINKENJANN A., PÉRARD-GAYOT A., SLUSALLEK P., LI Y.: Foveated real-time ray tracing for head-mounted displays. *Computer Graphics Forum 35* (2016), 289–298. 3

[WSXC16] WANG C., SHI F., XIA S., CHAI J.: Realtime 3d eye gaze animation using a single rgb camera. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 118. 2, 3

[WWJ16] WANG K., WANG S., JI Q.: Deep eye fixation map learning for calibration-free eye gaze tracking. In *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications* (2016), pp. 47–55. 2

[WXLY17] WEN Q., XU F., LU M., YONG J.-H.: Real-time 3d eyelids tracking from semantic edges. *ACM Transactions on Graphics (TOG) 36*, 6 (2017), 193. 2, 3

[WXY16] WEN Q., XU F., YONG J.-H.: Real-time 3d eye performance reconstruction for rgbd cameras. *IEEE transactions on visualization and computer graphics 23*, 12 (2016), 2586–2598. 1, 2, 3, 4, 5

[XLCZ14] 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* (2014), pp. 1113–1121. 2

[YC05] YOO D. H., CHUNG M. J.: A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding 98*, 1 (2005), 25–51. 2

[YKLC02] YOO D. H., KIM J. H., LEE B. R., CHUNG M. J.: Non-contact eye gaze tracking system by mapping of corneal reflections. In *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition* (2002), pp. 101–106. 2

[YUYA08] YAMAZOE H., UTSUMI A., YONEZAWA T., ABE S.: Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (2008), pp. 245–250. 2

[ZCM*15] ZHANG Y., CHONG M. K., MÜLLER J., BULLING A., GELLERSEN H.: Eye tracking for public displays in the wild. *Personal and Ubiquitous Computing 19*, 5 (2015), 967–981. 1

[ZSB18] ZHANG X., SUGANO Y., BULLING A.: Revisiting data normalization for appearance-based gaze estimation. In *ETRA* (2018). 7

[ZSFB15] ZHANG X., SUGANO Y., FRITZ M., BULLING A.: Appearance-based gaze estimation in the wild. In *CVPR* (2015). 3

[ZSFB17] ZHANG X., SUGANO Y., FRITZ M., BULLING A.: Mpiigaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 41*, 1 (2017), 162–175. 3