

With Gaze Tracking Toward Noninvasive Eye Cancer Treatment

Stephan Wyder*, Fabian Hennings, Simon Pezold, Jan Hrbacek, and Philippe C. Cattin

Abstract—We present a new gaze tracking-based navigation scheme for proton beam radiation of intraocular tumors and we show the technical integration into the treatment facility. Currently, to treat a patient with such a tumor, a medical physicist positions the patient and the affected eye ball such that the radiation beam targets the tumor. This iterative eye positioning mechanism requires multiple X-rays, and radio-opaque clips previously sutured on the target eyeball. We investigate a possibility to replace this procedure with a noninvasive approach using a 3-D model-based gaze tracker. Previous work does not cover a comparably extensive integration of a gaze tracking device into a state-of-the-art proton beam facility without using additional hardware, such as a stereo optical tracking system. The integration is difficult because of limited available physical space, but only this enables to quantify the overall accuracy. We built a compact gaze tracker and integrated it into the proton beam radiation facility of the Paul Scherrer Institute in Villigen, Switzerland. Our results show that we can accurately estimate a healthy volunteer's point of gaze, which is the basis for the determination of the desired initial eye position. The proposed method is the first crucial step in order to make the proton therapy of the eye completely noninvasive.

Index Terms—Applied optics, camera calibration, gaze tracking, homography, infrared (IR) ray-tracing, intraocular tumors, navigation system, ocular oncology, proton beam therapy, tumor targeting.

I. INTRODUCTION

UVEAL Melanomas are the most frequent primary intraocular tumors. Untreated, they may lead to blindness and to death (caused by metastasis). Nowadays, such tumors can be treated successfully with proton radiation while conserving the eye (eye retention rate of approx. 90%), and frequently preserving the eye function [1]. One of the key success factors of the therapy is the fact that the radiation dose can be delivered very precisely, due to the Bragg peak. The maximum of the Bragg curve, the Bragg peak, corresponds to the highest energy loss of charged particles at a certain depth in the tissue. This is notably important for organs with a high density of critical structures like the human eye [2].

The current method of choice for therapy planning by Goitein and Miller [3] has been in use with small alternations for more than three decades. Despite its successful outcome, the main limitation of the current method is that it requires an invasive

patient preparation. The surgeon sutures radio-opaque clips to the outer scleral surface around the tumor periphery in order to define the tumor location within the eye and to be able to target the tumor during radiation therapy. Two orthogonal X-ray units and a motorized chair are integral parts of the treatment facility. X-rays are used to localize the clips allowing for an iterative alignment of the eye for the actual therapy.

Our objective is to bridge the gap between the fields of computer based navigation and clinical proton therapy with the goal to make the entire workflow noninvasive. A suitable gaze tracking integration is the key. The main purpose of a gaze tracker is to estimate where a human subject is looking. Additionally, some gaze trackers are able to estimate the exact position and orientation of a subject's eye in three-dimensional (3-D) space. We reuse this by-product of the gaze tracker to align the eye to the treatment beam. An integral part of those types of gaze trackers are infrared (IR) LEDs at a known position allowing to determine the eye location in a 3-D space.

By introducing a gaze tracking system into the treatment scheme, we aim to overcome the invasive preparation of patients for the treatment.

Gaze tracking systems have already been used in proton beam therapy to gate the proton beam in case of sudden eye motion [4]. To target tumors noninvasively, Rügsegger *et al.* introduced an OCT-based tumor targeting system [5]. Fassi *et al.* [6] as well as Via *et al.* [7] also introduced gaze tracking into proton therapy to treat eye tumors. Their work covers the general idea of gaze tracking-supported tumor targeting, and they propose a specific gaze tracking system, presented independently of the actual treatment. However, their integration of the gaze tracking system into the treatment facility requires additional space-consuming and expensive hardware.

In contrast to the previous work, we strive for a whole new navigation scheme and a gaze tracking system that is completely integrated into the treatment. Our system is designed such, that it supports a fluent transition to a completely noninvasive treatment.

This paper is structured as follows: In Section II, we cover all aspects of our treatment reformation, the hardware, the algorithm, the treatment scheme, and the coregistration between coordinate systems (CS). In Section III, we present the accuracy of the overall integration, and finally, we summarize the benefits of our development and we present further work in Section IV.

II. METHODS

We present our contribution in a top-down manner. We start with the proposed treatment scheme, then we continue with the

Manuscript received May 08, 2015; revised October 16, 2015 and November 26, 2015; accepted November 29, 2015. Date of publication December 04, 2015; date of current version August 18, 2016. This work was supported by Swiss National Science Foundation. *Asterisk indicates corresponding author.*

*S. Wyder is with the Department of Biomedical Engineering, University of Basel, Basel 4003, Switzerland (e-mail: stephan.wyder@unibas.ch).

F. Hennings and J. Hrbacek are with the Paul Scherrer Institute.

S. Pezold and P. C. Cattin are with the Department of Biomedical Engineering, University of Basel.

Digital Object Identifier 10.1109/TBME.2015.2505740

description of the gaze tracker integration into the treatment facility. Finally, we explain the developed gaze tracking hardware and software itself.

A. Treatment Scheme: Today and in the Future

Today, the workflow is basically as follows: After the patient has an intraocular tumor diagnosed, the ophthalmologist suggests a treatment with proton radiation at PSI and the patient has to undergo a planning phase. Amongst other ophthalmic examinations, different imaging modalities (mainly fundus and ultrasound imaging) are used to determine eye conditions, tumor location, and shape. The tumor is made visible for X-rays by suturing radio-opaque clips around the tumor periphery on the outer scleral surface (outer surface of eyeball). This is required to define the shape and to target the tumor during the actual treatment. After image acquisition and clip surgery, the medical physicist lays out a plan, based on an ellipsoidal eye model that includes the tumor shape and clips. With that plan, the ideal beam path through the eye is determined in a way, so that sensitive structures (e.g., anterior segment, macula, and optical nerve) are spared from radiation as much as possible.

The output of this planning stage is an estimated ideal gazing angle, assuming that the tumor within the eye is located at the isocenter, i.e., the reference point on the proton beam axis. For the actual radiation, the patient has to sit on the treatment chair and his or her head is fixated by a mask and a bite block. The planned gazing angle is controlled by asking the patient to fixate a shining diode that is appropriately positioned on the LED bar in front of his or her face. The patient has to be awake and be able to see and fixate the shining LED on the bar. Without the gaze tracker, the eye and tumor position of the patient cannot be measured directly. Therefore, the radio-opaque clips and two orthogonal X-ray systems are required to arrange the eye properly. This is done in an iterative way and usually requires several X-ray shots until the clips, visible on the X-ray projections, correspond to the planned clip positions. To preserve the adjusted gazing angle during radiation (gazing control), a medical physicist monitors the patient's eye motion over a camera. In case of sudden eye motion, a medical physicist manually interrupts the proton radiation.

Fig. 1 illustrates in its first column the scheme as it exists now and as it is described above. In the second column, we picture our proposed treatment scheme according to the steps below. The third column shows the scheme that we are working toward. The clip surgery would be kept for the transition period. In the future, it could be replaced by the production of a patient-specific eye model, which also contains the tumor [8], [9]. This eye model is based on a fusion of several image modalities, namely OCT, MRI, CT, and fundus imaging. For our purpose, the model could be registered with the 3-D model of the proposed gaze tracker. To do so, two distinct points and two angles would be used, which can be acquired in both models: center of pupil, center of corneal curvature, and two angles, describing the offset between optical and visual axis.

Fig. 2 illustrates the treatment hardware arrangement for a better understanding of the following scheme:

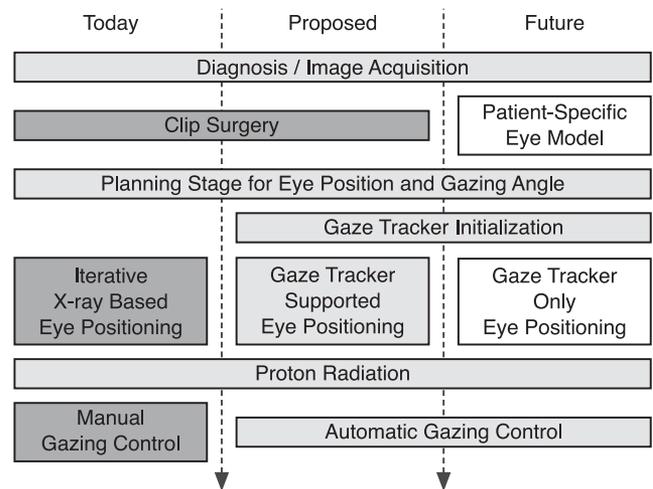


Fig. 1. Overview of treatment scheme (from top to bottom): Transition period from today (dark gray boxes) to the future (white boxes).

- 1) The medical physicist prepares the patient on the treatment seat and fixates its head ⑩ with an appropriate mask and a bite block. The seat can be adjusted with six degrees of freedom in order to have the tumor roughly at the isocenter. This is required because the treatment device has a fixed horizontal beam line ⑦.
- 2) The LED bar ② gets rotated and initialized with the appropriate LED turned ON, by means of the two polar coordinates coming from the treatment planning stage.
- 3) The technician rotates the entire gaze tracking hardware ③, ④ and ⑥ to one of 36 possible discrete positions so that the gaze tracker is roughly aligned parallel to the LED bar ②. The hot mirror ⑥ is then roughly positioned in front of the shining LED. This gaze tracker alignment is a prerequisite for getting the eye into the field of view of the camera ④.
- 4) Different eye positions along the beam axis might be needed for different patients, depending whether the tumor is more in the anterior or the posterior part of the eye. Accordingly, the optimal field of view and a good depth of focus of the camera ④ has to be ensured. Slight adaptations to a new patient might therefore be required. The focus of the lens ④ can be adjusted and the mirror ⑥ can be tilted in order to get a sharp and appropriate field of view for the camera ④. If no adaptations of the optical system (i.e., camera focus or mirror angle) are required, the camera calibration and homography can be reused from a previous session with the same gaze tracker position. If changes to the hardware were made, a gaze tracker calibration procedure is required to initialize the gaze tracking system (see below).
- 5) The exact position of our virtual camera ⑤ in the world CS is known at this stage, thanks to the gaze tracker calibration procedure.
- 6) A subject-specific parameter optimization is performed by having the patient ⑩ look at six calibration points given by the appropriate shining LED ② with known position

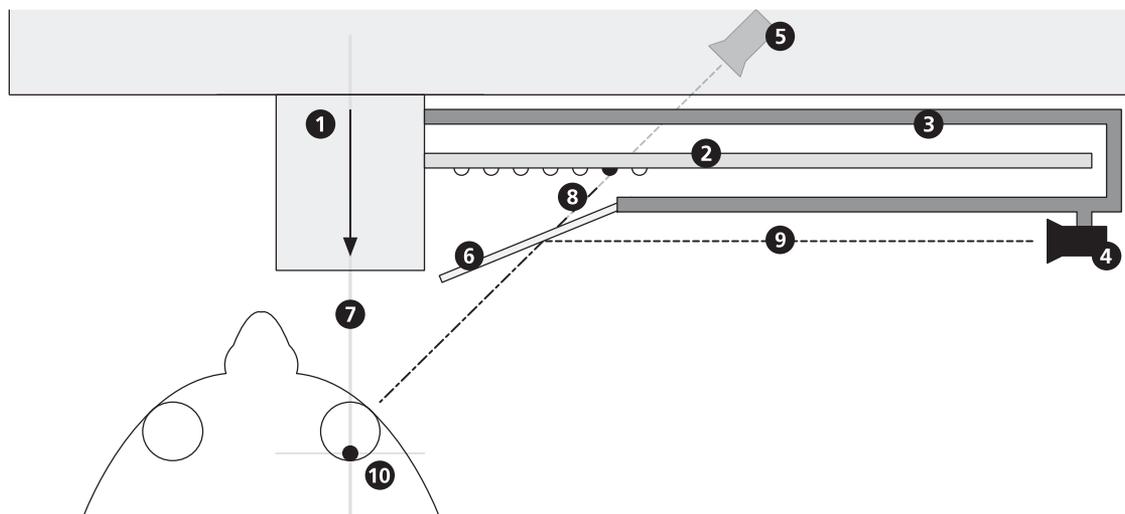


Fig. 2. Situation Plan (caudal view): ① Treatment device with proton beam; ② LED bar with activated fixation light, can be rotated around proton beam axis; ③ Holder for gaze tracker, can be rotated around proton beam axis; ④ IR camera and illumination (gaze tracker); ⑤ Virtual gaze tracker position; ⑥ Hot mirror reflecting IR wavelengths, while letting trough visible wavelength; ⑦ Proton beam and rotation axis; ⑧ Patient's line of sight, fixating activated LED; ⑨ IR rays, imaging the eye; ⑩ Patient with target eye tumor in focus at isocenter of treatment device.

in 3-D space. The algorithm has to be trained in order to fit the subject-specific parameters of the 3-D model to the appropriate patient getting treated.

- 7) Using the subject-specific parameters, the gaze tracker is able to estimate the point of gaze (PoG) and the position of certain anatomical landmarks of the eye (e.g., the pupil center) of the subject in the world CS. The PoG can additionally be used as fixation monitor to gate the proton beam in case of sudden eye motion. The determination of the pupil center and the nodal point of the eye (center of corneal curvature) enables us to initialize the planned position of the eye more precisely than by simply starting the iterative alignment process at an arbitrary position (simplified gaze tracker supported initialization, see Fig. 1). Since the initial estimate of the eye position is more accurate, fewer X-ray shots are required to align the tumor with the isocenter. In the future, the whole clip alignment with X-rays can be omitted, as soon as the 3-D eye model of the gaze tracker is coregistered with a patient-specific model [8], [9] showing the segmented tumor shape.

B. Gaze Tracker Calibration Procedure

Depending on the extent of adjustments made to the gaze tracker hardware since the last use, either a partial or a full calibration is required. If only the position of the gaze tracker was changed, a partial calibration is sufficient by following the instructions of steps 2 and 3. If either the focus or the aperture of the lens were changed, a full calibration is required by following the instructions of all of the following steps 1 to 3:

- 1) If the technicians made adjustments to the camera's lens ④, they have to perform a camera calibration [10] in order to get the linear and nonlinear intrinsic camera parameters, namely, the camera matrix and the radial and tangential distortion, respectively. This camera calibration makes it possible to map between pixel distances of images and

real-world distances in millimeter. Furthermore, it enables correcting for distortions in the acquired images. The calibration algorithm [10] requires a set of images showing a checkerboard pattern on a planar surface from different angles.

- 2) Once the camera ④ is initialized (calibrated), its position relative to the world CS has to be determined. Instead of the head fixation mask for the patient, a calibration cube is mounted on the motorized chair. The calibration cube contains radio-opaque markers and can be aligned exactly to the world CS (isocenter of the proton beam) with the help of the two orthogonal X-rays.
- 3) Additionally, attached to the same cube is a checkerboard pattern at a known position, which is visible for the camera ④ via the hot mirror ⑥. The technician acquires a homography, based on a camera image showing the checkerboard. This enables to get the appropriate translation vector and rotation matrix from the camera CS to the checkerboard CS (finding extrinsic parameters). The optical rays between the camera ④ and the checkerboard are deflected on the mirror ⑥. Therefore the whole homography actually represents the transformation from the calibration cube to a virtual camera position ⑤ behind the mirror ⑥. All coordinates of the gaze tracker ④ are transformed to the world CS by this transformation, thus the mirroring and the real camera position can be neglected.

At this stage the gaze tracker CS and the world CS have been successfully coregistered. The reason for coupling the gaze tracker CS with the world CS of the treatment device is twofold: First, to save physical space, we use the existing LED bar of the treatment device to calibrate and validate the gaze tracker. The LED bar of the treatment device itself is precalibrated and can be controlled by means of two polar angles. Thus, the treatment device itself becomes part of the gaze tracking system and vice versa. Second, we reuse the output of the gaze tracker, namely two coordinates of the eyeball, within the treatment device. For

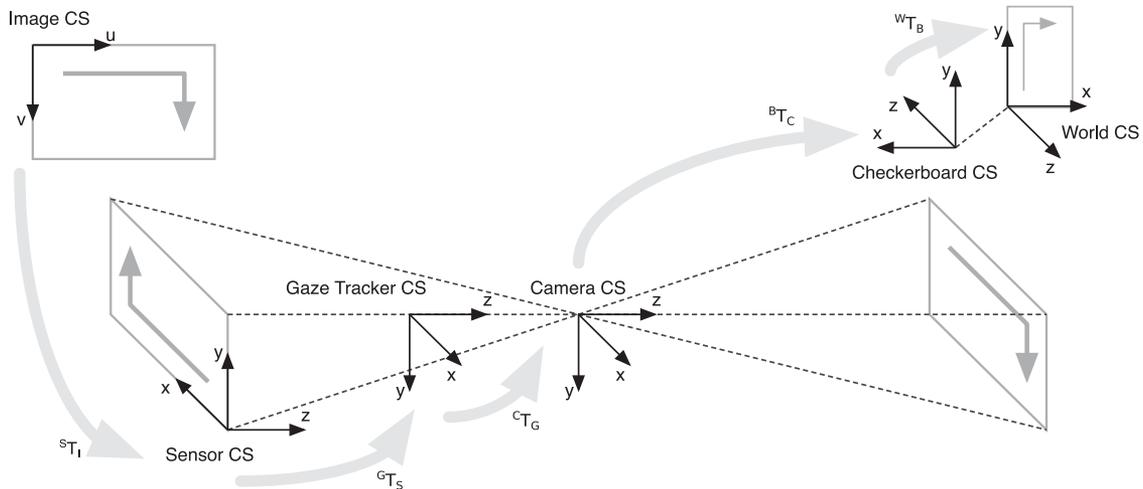


Fig. 3. CSs and its transformations (${}^N T_M$ = transformation from M CS to N CS, where M and N refer to the following CS abbreviations: I = image, S = sensor, G = gaze tracker, C = camera, B = checkerboard, W = world)—top left: image, top right: part of the treatment device, bottom: part of the camera, and the lens.

TABLE I
COREGISTRATION PIPELINE

${}^I \mathbf{p}$	Input: 2-D point in image CS.
${}^S \mathbf{p} = {}^S T_I \cdot {}^I \mathbf{p}$	${}^S T_I$ converting $[x, y]_{\text{px}}$ into $[x, y]_{\text{mm}}$ and adding third dimension to get $[x, y, z = 0]_{\text{mm}}$.
${}^G \mathbf{p} = {}^G T_S \cdot {}^S \mathbf{p}$	${}^G T_S$ translating to the principal point (center of sensor), rotating with 180° around the z -axis and translating along the z -axis with 17.52 mm (flange focal distance of c-mount cameras).
${}^C \mathbf{p} = {}^C T_G \cdot {}^G \mathbf{p}$	${}^C T_G$ depending on intrinsic camera parameter f (focal length). The translation along the z -axis consists of a removal of the flange focal distance and an addition of f .
${}^B \mathbf{p} = {}^B T_C \cdot {}^C \mathbf{p}$	${}^B T_C$ being an affine transformation coming from the extrinsic parameters of the camera calibration (homography).
${}^W \mathbf{p} = {}^W T_B \cdot {}^B \mathbf{p}$	${}^W T_B$ constant and given by dimensionality and location of used calibration cube.
${}^W \mathbf{p}$	Output: 3-D point in world CS.

those purposes, points in space must be transformed back and forth between the gaze tracker and the treatment device.

We propose a registration chain consisting of several CSs used as intermediate steps between the image CS on one side of the chain and the world CS (isocenter of the proton beam) on the other side.

Fig. 3 illustrates all involved CSs and shows how they are defined in space.

Table I illustrates the mentioned transformation chain, where ${}^M \mathbf{p}$ corresponds to a point vector in the M CS and ${}^N T_M$ represents a transformation matrix from M to N CS, where M and N refer to the following CS abbreviations: I = image, S = sensor, G = gaze tracker, C = camera, B = checkerboard, W = world.

The image, sensor, camera, and checkerboard CSs have to be used explicitly by the gaze tracking algorithm and the camera calibration. The gaze tracker CS origin is defined such, that we can measure (calibrate) the IR LEDs relative to it. We need at least one such “visible” CS, since all the others, close to the gaze tracker, are either hidden in the case (sensor CS) or virtual and not fixed relative to the hardware (image CS and camera CS). A good general overview of camera models and

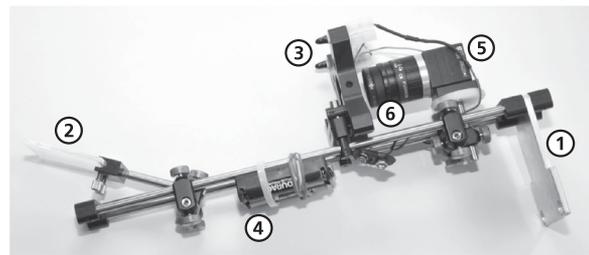


Fig. 4. Gaze tracking hardware for treatment facility: Consists of a frame ①, a hot mirror ②, two IR LEDs ③ with appropriate power supply ④, a camera ⑤ and a lens with an IR-pass filter ⑥.

calibration, coordinate transformations and its state-of-the-art nomenclature can be found in [11]. We do camera calibration and homography estimation with the OpenCV C++ library [12] by using the state-of-the-art algorithm described in [10].

C. Gaze Tracking Hardware

The present gaze tracker hardware consists of a frame, a hot mirror, an industry camera, two IR LEDs, and a battery pack. Fig. 2 illustrates the concept and Fig. 4 shows the prototype with all its components.

Especially the construction of the frame, which holds the remaining components, is challenging. The frame stability drastically impacts the reliability of the whole system. Slight displacements of hardware components may lead to drastic accuracy and repeatability loss. Position and angle changes of either the camera, the LEDs or the mirror could potentially occur during treatment because of vibrations of the treatment apparatus. Consequently, a very rigid frame construction is required to reach the desired mathematical stability in the 3-D model. However, optical engineering usually involves a lot of trial and adjustment with optical components for aspects which cannot be precalculated. That is why we finally ended up with optomechanical components to build the prototype for the treatment facility.

Those components enable us to build a rigid construction, while slight adjustments are still possible.

The hardware components are introduced briefly below: The industry camera (XIMEA MQ013MG-E2) has a C-mount for lenses, runs with 60 Hz at a resolution of 1280×1024 px.

The lens has a focal length of 35 mm and additionally an IR-pass filter, which blocks visible wavelengths below 650 nm.

Two IR LEDs (Osram Components SFH 4555), with a wavelength $\lambda = 860$ nm, 100 mA and 1.5 V each, are responsible for the ambient lighting of the eye as well as for the glints (specular reflections) on the subject's cornea used by the gaze tracker for spatial referencing. The two LEDs are connected in series with three 1.5-V batteries and a series resistor of 15 Ω .

The image quality is mainly given by the following parameters: aperture of the lens, exposure time of the camera, and amount of available light given by the LEDs.

Unfortunately, the human eye has no protection mechanism to avoid damage from IR light, since it is not visible. Therefore, it has to be ensured, that the radiation intensity limits, given by the safety norms, are not exceeded. In our case, the minimal distance between the eye and the LEDs has to be 250 mm, which is guaranteed by design.

A hot mirror is a coated piece of glass, which reflects IR light but is permeable for visible light. We use a hot mirror because we can save physical space by reflecting the optical rays of our 3-D gaze tracking model. Likewise, we do not want to hide a lit red LED on the LED bar, which could potentially be behind the mirror. Fig. 2 illustrates the principle and the physical space available.

D. Gaze Tracking Algorithm

The purpose of the gaze tracking algorithm is to estimate the PoG of the subject under observation. A good overview of existing gaze tracking algorithms is presented in [13]. We work with an extended version of a popular 3-D model described in [14]. The elegance of this model lies in its flexibility, since several different hardware configurations are possible. The configurations vary from one camera with one light source to multiple cameras with multiple light sources. The more cameras and lights used, the fewer the parameters that have to be optimized within the 3-D model. However, all the component positions have to be determined and registered with the gaze tracker CS. Our configuration with one camera and two light sources is a proper compromise between size of the apparatus and complexity of the model's solution space.

The algorithm can mainly be subdivided into the following parts: *Feature Detection*, *Model Fitting*, and *Subject-Specific Parameter Optimization*.

1) *Feature Detection*: As feature points on the eye, we use the coordinates of the pupil center and of both IR reflexions (glints). These features are detected in a semisupervised manner. By performing it this way, compared to a fully automatic manner, we reduce the complexity of the feature detection, and with that, the calculation time, whereas we increase the stability for different light conditions. Since we work in the IR range, the light conditions change mainly when sunlight is present or when camera gain, aperture, or exposure time is changed.

Let $\Omega = \{1 \dots 1280\} \times \{1 \dots 1024\}$ denote the discrete domain of the images provided by the camera, let $\Omega_R \subseteq \Omega$ denote a rectangular region of interest, and let $I : \Omega_R \rightarrow \{0 \dots 255\}$ with $I = I(u, v)$ denote the image function that maps from pixel coordinates (u, v) within the region of interest to grayscale values. The region Ω_R is set manually to extract the eye. To detect the pupil and the glints, the supervisor also sets seed points $\{(u_i^p, v_i^p)\}_{i=1}^{N_p}$ for the pupil, $\{(u_i^i, v_i^i)\}_{i=1}^{N_i}$ for the iris, and $\{(u_i^g, v_i^g)\}_{i=1}^{N_g}$ for the glints. For the respective pixel intensities, we calculate the average values a .

$$a_p = \frac{1}{N_p} \sum_{i=1}^{N_p} I(u_i^p, v_i^p) \quad (1a)$$

$$a_i = \frac{1}{N_i} \sum_{i=1}^{N_i} I(u_i^i, v_i^i) \quad (1b)$$

$$a_g = \frac{1}{N_g} \sum_{i=1}^{N_g} I(u_i^g, v_i^g) \quad (1c)$$

then, we determine thresholds t_p and t_g for the pupil and glint, respectively, as

$$t_p = \frac{1}{2}(a_p + a_i) - s_p \quad \text{and} \quad t_g = a_g - s_g \quad (2)$$

where s_p and s_g are predefined tolerance values. Finally, we use the thresholds to calculate two binary images I_p and I_g as follows:

$$I_p(u, v) = \begin{cases} 1, & I(u, v) < t_p \\ 0, & I(u, v) \geq t_p \end{cases} \quad \text{and} \quad (3a)$$

$$I_g(u, v) = \begin{cases} 1, & I(u, v) > t_g \\ 0, & I(u, v) \leq t_g. \end{cases} \quad (3b)$$

Having the binary images, we label the individual blobs of I_p and I_g by applying a standard eight-connected component labeling algorithm. This results in \hat{I}_p and \hat{I}_g , where the individual, separated blobs are labeled by an identifier number. The glint segmentation does not need any further processing, since the IR reflexions are by far the brightest regions in Ω_R . Since the pupil segmentation does not yet result in such a homogeneous result, we apply a closing morphological operator \bullet to the labeled pupil image \hat{I}_p

$$\tilde{I}_p = \hat{I}_p \bullet S \quad (4)$$

where S is a structuring element with a circular shape and a radius of 9 px. This operator smoothes in particular the border region of the pupil.

In the case where several extracted pupil regions k occur, we have to distinguish the correct one from the wrong ones. Potentially wrong, dark regions, for example, come from mascara of women's eyes. For that, we calculate two different region properties, the equivalent diameter d_k and the eccentricity e_k for all k regions (pupil candidates). We first calculate the equivalent diameter d_k , where n_k is equal to the number of pixels

corresponding to the region k

$$d_k = \sqrt{\frac{4 \cdot n_k}{\pi}}. \quad (5)$$

To get the eccentricity of the regions, we have to get the second central moment to calculate the covariance matrices C_k . The two major eigenvalues λ_{a_k} and λ_{b_k} of each covariance matrix indicate the length of the major axis \mathbf{a} and the minor axis \mathbf{b} . The eccentricity e_k can then be calculated for all regions k by

$$e_k = \sqrt{1 - \frac{\lambda_{\mathbf{b}_k}^2}{\lambda_{\mathbf{a}_k}^2}}. \quad (6)$$

Finally, the pupil candidate \check{I}_p is selected by taking the label k of the biggest blob $\max(d_k)$ for all k , where the eccentricity $e_k < 0.7$

$$\check{I}_p : \Omega_R \rightarrow \{0, 1\}, \check{I}_p := \begin{cases} 1 & \check{I}_p = \arg \max_k(d_k) \\ 0 & \text{else.} \end{cases} \quad (7)$$

Having the pupil candidate \check{I}_p , we extract the edges \check{I}_p of the region by applying the Sobel operator. To finally get the coordinates of the pupil center, we fit a circle into the edge image \check{I}_p with the Hough transform. The parametric form of a circle has three degrees of freedom (x and y coordinates and radius). In order to save calculation time, we limit the radius parameter to $\max(d_k) \pm 5$ px. Having the Hough accumulator H , we take the average of the four strongest peaks $\{\mathbf{h}_1, \dots, \mathbf{h}_4\}$ of it

$$\mathbf{p} = \frac{1}{4} \cdot \sum_{n=1}^4 \mathbf{h}_n. \quad (8)$$

The first vector component, $p_1 = p_x$ corresponds to the x coordinate, $p_2 = p_y$ to the y coordinate, and $p_3 = r$ to the radius of the detected pupil. Taking four peaks, rather than just one, enables us to take ellipsoidal shapes into account as well. This is in particular important, when the pupil, assumed to be a circle, is viewed from the side.

There may be cases where more than two glints are visible, two deflected by the mirror and two reflections coming directly from the LEDs. We calculate the centroids for all glint candidates in \hat{I}_g and we compare them to the centroid of the pupil (p_x, p_y). The glint pair $(g_{x_1}, g_{y_1}, g_{x_2}, g_{y_2})$ closest to the pupil center is the pair we are looking for. Finally, we transform all three feature points to the world CS and pass them to the gaze estimation algorithm, i.e., the model fitting. Fig. 5 shows a typical eye image with its features.

2) *Model Fitting*: The basis of the 3-D gaze estimation model [14] is a set of equations describing relations in a 3-D vector space. The whole system of equations has to be solved repeatedly, for each re-estimation of a PoG, respectively, the estimation of the position of the eye. Solving the system is basically equivalent to estimating the unknown model parameters. The equations and the corresponding diagram can be found in the Appendix.

The system of equations can be subdivided into three subsystems, which get solved sequentially (see Fig. 7). The first system consists of 13 scalar equations with 12 scalar unknowns. The

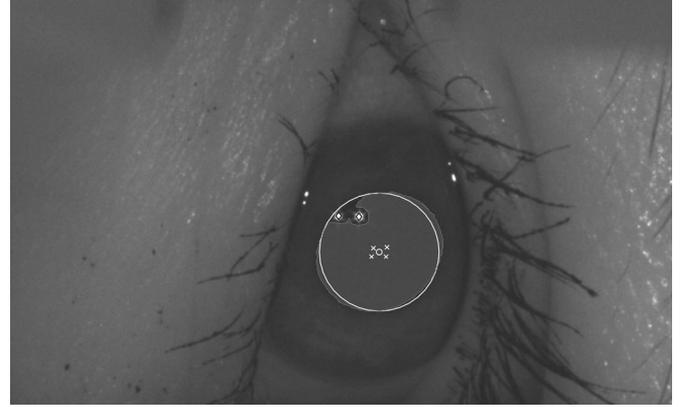


Fig. 5. Typical region of interest of the eye with features: Diamonds correspond to glints, the crosses correspond to Hough peaks resulting in the circle fit for the pupil.

individual equations bring certain points of the model in relation by using equations of lines, Euclidean distances between points and the law of reflection. The system is overdetermined and nonlinear because its unknowns are not separable (mixed unknown parameters). The respective output of the previous system is used to solve the next set of equations. The second system consists of a line equation and of a distance between points. Here, we finally end up with four scalar equations and four scalar unknowns, and the system is of quadratic order and can, therefore, be solved analytically. The third system, finally, consists of a Euclidean distance between two points, a triple product showing the coplanarity of three points and the law of refraction describing the situation on the cornea surface. This system contains three scalar equations with three scalar unknowns. It is nonlinear, again, because of mixed terms consisting of unknown parameters.

The model mainly consists of four different kinds of parameters.

- 1) Hardware related parameters have to be set once according to the gaze tracking hardware. This parameter set comprises of the position of the light sources, the nodal point of the camera, and the details about the camera sensor.
- 2) Subject-specific parameters are constant for a certain subject and cannot be measured directly: angles and lengths within the eye and the mean index of refraction of the aqueous humor and the cornea. These parameters get calibrated for each subject. As starting values for the optimization, tabulated average values are used.
- 3) Input parameters are the coordinates of the detected eye features (glints and pupil).
- 4) The rest of the parameters are only time dependent and describe the position of certain points within the subject's eye and the subject's PoG.

For more information on parameters, refer to [14].

Precise initial values are required for the last group of parameters, since the systems of equations are nonlinear and we are not using global optimization. The theoretical eye model of Gullstrand and Le Grand [15], and the hypothetical tumor position at the isocenter, enables us to estimate the required initial guess for the points in the eye.

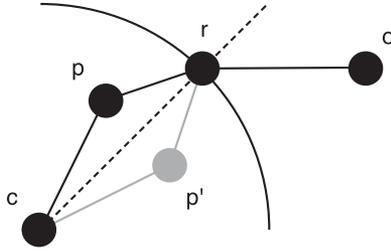


Fig. 6. Points within the eye: c is the center of corneal curvature, p is the pupil, r is the point of refraction, o is the nodal point of the camera.

The parameter fit for the gaze estimation may find a numerically correct solution which is, however, not correct in a physical sense. This results in a wrong PoG estimation. Parameter p of the 3-D model [14] is especially affected. The 3-D model ensures the coplanarity of the corneal curvature c , the pupil center p , the point of refraction r , and the nodal point o of the camera, but it cannot distinguish between p and p' . Fig. 6 illustrates the situation.

In addition to the correct point p , there is a numerically correct result p' , which, however, is physically implausible because of the law of refraction. To overcome this, we introduce an additional equation, with which we compare the direction of two normals \mathbf{n}_a and \mathbf{n}_b . The direction of the normal, given by the cross product of two vectors, is defined by the right-hand rule. This allows us to distinguish the two cases. The normal \mathbf{n}_a on the plane defined by c , p , and r points in the same direction as the normal \mathbf{n}_b on the plane defined by p , r , and o

$$\mathbf{n}_a = [(\vec{c} - \vec{p}) \times (\vec{r} - \vec{p})] \quad (9)$$

$$\mathbf{n}_b = [(\vec{p} - \vec{r}) \times (\vec{o} - \vec{r})] \quad (10)$$

$$\cos^{-1} \left(\frac{\mathbf{n}_a \cdot \mathbf{n}_b}{\|\mathbf{n}_a\| \cdot \|\mathbf{n}_b\|} \right) = 0. \quad (11)$$

Additionally, we introduce a stochastic optimization by solving the third system again with randomly chosen initial values, when the system finds a wrong local minimum. This wrong local minimum can be detected when the residual of (11) is far from zero.

3) *Subject-Specific Parameter Optimization:* The previously mentioned parameter set consists of

- 1) R = Radius of corneal curvature.
- 2) K = Distance between the center of the pupil and the center of corneal curvature.
- 3) n_1 = Effective index of refraction of the cornea and the aqueous humor combined.
- 4) α_{eye} = Horizontal angle between visual and optical axes of the eye.
- 5) β_{eye} = Vertical angle between visual and optical axes of the eye.

All subjects have to go through a calibration procedure, where they have to look at six predefined points. The exact location of those points is known. Therefore, the average error between the PoG estimate and the true position of those six points is used to optimize the five subject-specific parameters. In the first step, we use tabulated values to get a first estimate of the PoG.

```

1: procedure OPTIMIZEPARAMETERS
2:   repeat
3:      $R, K, n_1, \alpha, \beta \leftarrow \text{GETNEWPARAMETERGUESS}()$ 
4:     procedure GAZETRACKING
5:        $f \leftarrow \text{DETECTFEATURES}(i)$   $\triangleright f = \text{features}$ 
6:                                      $\triangleright i = \text{image}$ 
7:     OPTIMIZE SYSTEM ONE()
8:     SOLVE SYSTEM TWO()
9:     while  $r \neq 0$  do  $\triangleright r = \text{residual}$ 
10:       $ig \leftarrow \text{GETNEWINITIALGUESS}()$ 
11:       $r \leftarrow \text{OPTIMIZE SYSTEM THREE}(ig)$ 
12:    end while
13:    return  $p, c, \alpha, \beta, \text{PoG}$ 
14:                                      $\triangleright \text{optical \& visual axis}$ 
15:                                      $\triangleright \text{PoG} = \text{point of gaze}$ 
16:  end procedure
17: until  $\text{PoG} \approx \text{tCP}$   $\triangleright \text{tCP} = \text{true calibration point}$ 
18: end procedure

```

Fig. 7. Gaze tracking algorithm.

Fig. 7 illustrates the workflow of the algorithm.

III. RESULTS

A. Experimental Setup

We performed our experiments with ten healthy volunteers to show the accuracy of the gaze tracking system integrated into the treatment facility. Each volunteer was placed on the treatment chair of the proton radiation system at PSI. Usually, a custom-made head mask is manufactured for a patient, which fixates the head and allows only eye rotation. Since the volunteers did not have their own custom-made head masks, we used the same demo mask for all of them. Because of that, all the volunteers were able to still slightly move their head, and with that, the position of the eye (roughly ± 3 mm in all directions). This, however, should not have had a negative influence on the results, as the model is capable of handling head movement. The motorized chair was moved to the initial position, which is normally used to treat the right eye. The cornea is thus positioned quite close to the collimator, ≈ 2 cm. The distance between cornea and LED bar is roughly 11 cm. We recorded 12 different predefined calibration points per volunteer. To do so, we reused the already existing LED bar. The LEDs on the bar have a diameter of 2 mm. Since they are so close to the eye, it is very hard for the volunteer on the chair to focus well on the LEDs. Hence, a halo effect appears around the lit LED. Usually, the calibration points get distributed over the whole field of view of the subject in order to realize the best result for the calibration of the subject-specific parameters. In our case, however, we were limited to a small area to place the calibration points. This is because of the possibilities to position the fixation light with the LED bar. Per volunteer, the 12 calibration points were recorded twice, one round after the other. For every calibration point, we recorded the coordinates of the calibration point together with an image of the eye, which was gazing at the corresponding calibration point. The data were divided into four

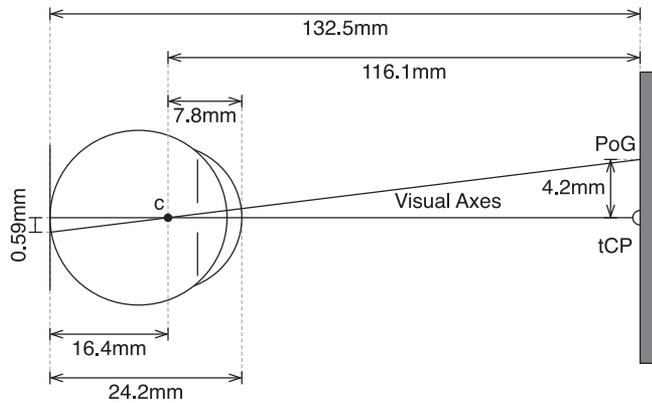


Fig. 8. Measured error at LED bar and estimated error on retina.

uniformly distributed subsets with six points each. On the first subset, we performed the calibration to get the subject-specific parameters. On the other three sets, we tested the accuracy of the PoG estimation with the previously optimized parameters (testing point sets).

For our experiments, we first calibrated the gaze tracker and acquired the images of the volunteers on-site at the treatment facility. Afterward, we performed the actual gaze tracking and evaluated the results retrospectively. Our algorithm is so far implemented in MATLAB and is not optimized for speed, yet. The duration for extracting the features and estimating one PoG is about 300 ms on a state-of-the-art laptop (2.8 GHz Intel Core i7). The duration for the optimization of the subject-specific parameters of one subject is about 80 s.

B. Experimental Results

The goal is to reach a sub-millimeter accuracy when determining the three dimensions of points in the subject's eye. Those points, coming from the 3-D model, are the point c , referring to the center of corneal curvature and the point p , the pupil center. Together with the angles α_{eye} and β_{eye} , both the optical and the visual axis can be determined exactly. However, since we cannot directly measure the accuracy at points c and p , we first analyze the accuracy of our system by comparing the PoG estimation with the true calibration point (tCP) position. With that result, we calculate an error estimate on the isocenter (tumor position) by means of the intercept theorem.

The mean distance between the estimated points of gaze and the tCP is 4.20 mm. This is the overall average of ten volunteers with three point sets each and six calibration points per set ($6 \cdot 3 \cdot 10 = 180$ distances between estimated and true PoG). Those deviations can be measured directly. To show the accuracy independent of the distance to the calibration plane, the error is normalized and expressed in degrees measured at the nodal point of the eye c . We use the theoretical eye model of Gullstrand and Le Grand [15] to get averaged eye length data (see Fig. 8).

The calibration bar with the LEDs is 132.5 mm away from the isocenter, and we assume that the tumor, and with this the retina of the eye, is at the isocenter. From the tabulated eye length values [15], we estimate the hypothetical distance from

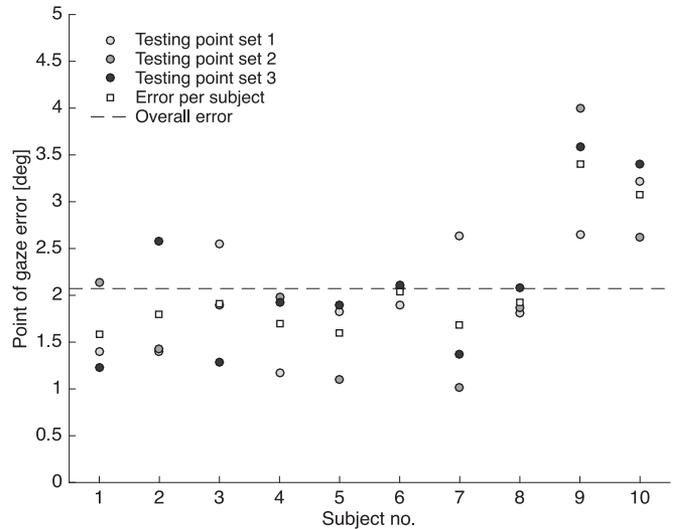


Fig. 9. Average PoG errors.

the calibration plane to the nodal point of the eye c as

$$132.5 \text{ mm} - (24.2 \text{ mm} - 7.8 \text{ mm}) = 116.1 \text{ mm}. \quad (12)$$

To get the millimeter-to-degree conversion we calculate

$$\tan^{-1}(4.20/116.1) = 2.07^\circ. \quad (13)$$

All measured data in millimeters are distance normalized and converted to degrees with the formula shown above and shown in Fig. 9. There is one column per subject and three circles per subject showing the averaged error per testing point set. The squares show the average error for one subject and the dashed line shows the overall error $2.07^\circ \hat{=} 4.20 \text{ mm}$.

To get an error estimate on the retina's depth (and thus tumor position), we use the intercept theorem and calculate

$$\frac{4.2 \text{ mm} \cdot 16.4 \text{ mm}}{116.1 \text{ mm}} = 0.59 \text{ mm}. \quad (14)$$

Fig. 8 illustrates this error estimate. Fig. 10 shows the optimized subject-specific parameters (one parameter set per volunteer).

IV. DISCUSSION

We want to improve eye cancer treatment. For that reason, we built and presented a compact gaze tracker integrated into the OPTIS gantry of the Paul Scherrer Institute in Villigen, Switzerland. The development and the integration of our gaze tracking hardware and software enabled us to estimate the points of gaze of ten volunteers, with a mean accuracy of $2.07^\circ \hat{=} 4.20 \text{ mm}$. This PoG error results in an error estimate of 0.59 mm on the retina. The retina is our region of interest, namely the potential tumor position, which we truly want to keep track of.

The result is encouraging, taking the complexity of the hardware and of the gaze tracking algorithm into account. The sub-millimeter accuracy is required for eye cancer treatment, and with our method we reach this requirement. The accuracy of our gaze tracking integration is expected in that range.

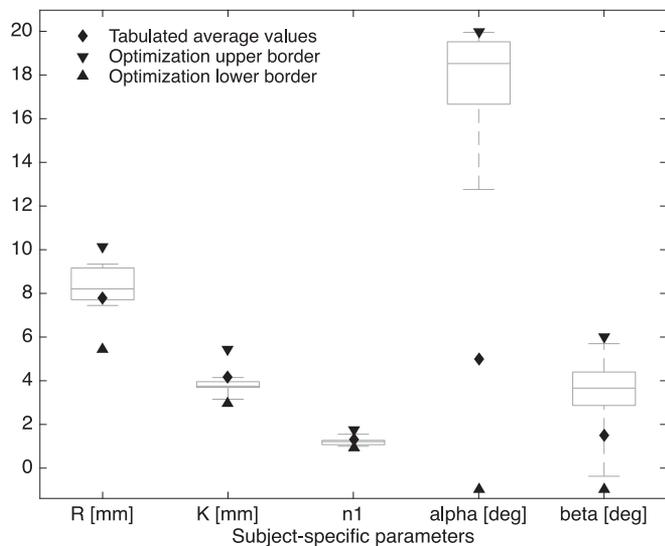


Fig. 10. Optimized subject-specific eye parameters. For an explanation of the parameters, see Section II-D3.

When we compare our error in degrees with the error of the results shown in [14], then their result is more accurate roughly by a factor of two. This is not surprising, since the pipeline of the proposed system additionally contains the camera pose estimation (homography), which makes the system more complicated than those described in [5], [7], and [14]. Furthermore, the type and location of the calibration points (LED's), given by the treatment device, makes the subject-specific parameter optimization difficult. An accurate parameter optimization under those circumstances might, therefore, be more challenging than in a more general setup with less boundary conditions, as described in the mentioned papers. However, our results are promising when we look at the calibration points we used, which have, with the halo effect, roughly a diameter of 5 mm and we look at the rather small distribution of the calibration points.

The estimated accuracy of 0.59 mm on the depth of the retina is promising. However, optimization is necessary to fulfill the requirement of an overall accuracy below 1 mm.

The existing overall error is made up of several error sources. Two possible sources of error contribution are the resolution of the sensor (physical limitation) and the accuracy of the feature detection (algorithmic inaccuracies). A third error source is the calibration of the subject-specific parameters. The big calibration-fixation points, and the difficulty to focus on them, might be suboptimal. Additionally, we are dependent on the compliance of the subjects, how accurately they actually look at a calibration point. A fourth source of error is the homography estimation, which has an influence onto the transformation of points in space. And last but not least, we rely on the 3-D coordinates of the LEDs on the LED bar. These absolute coordinates are maybe inaccurate. One of the two outliers (subject 10) may be explained by the fact, that the corresponding volunteer was wearing contact lenses during the experiments.

When we look at the distribution of the individual subject-specific parameters (see Fig. 10) and compare them to the

tabulated average values, we can see that the parameters α_{eye} and β_{eye} are comparatively far away from anatomically meaningful values. Those parameters may compensate for certain systematic errors mentioned earlier.

To get true accuracies rather than estimates at the depth where it is interesting for us, namely at the depth of the eye, we have to be able to measure the true points within the eye, which are estimated by the gaze tracker. To do so, a simplified eye phantom would be suitable, where the points in the eye (pupil, center of corneal curvature) can be calculated and compared to the estimated values. This will be subject of further research.

The integration of the proposed gaze tracking system enables us to cross check the proposed navigation scheme, which will replace the currently used initialization and navigation scheme of the treatment step-by-step.

The positive results allow us to continue working with this solution. Further hardware and software improvements are planned, which will especially lead to an even more user-friendly gaze tracker. However, accuracy improvements are also expected, since we plan, amongst other things, to use more precise and suitable optical components. We also plan to integrate an eye torsion monitoring mechanism, since the proposed gaze tracker cannot handle eye torsion so far.

Finally, our gaze tracking-based navigation scheme, together with a patient-specific eye model [8], [9], could potentially lead to a completely noninvasive proton therapy.

A major contribution of ours is the integration of the system into an already existing treatment facility. This makes the presented work quite application- and hardware-specific. However, the individual principles and the overall concept might be interesting for other facilities as well.

V. CONCLUSION

We developed and integrated a compact gaze tracker into the eye tumor treatment facility of the Paul Scherrer Institute in Villigen, Switzerland. With this gaze tracker, we want to accurately localize the eye in space, which supports the proper alignment of the head and the eye to the proton beam. With the proposed new treatment scheme, we aim at replacing the current tumor targeting system, which is accurate but also invasive. Our approach is completely noninvasive and we showed that the error estimate for the eye localization fulfills the required sub-millimeter accuracy. Further work, however, is required until we can treat the first patient completely noninvasively. The current results look promising and we are confident that, once the system is completed, the patients will benefit from the new treatment scheme.

APPENDIX 3-D MODEL EQUATIONS

Index i corresponds to the number of light sources, in our case, two: $i = \{1, 2\}$. Index j corresponds to the number of cameras. Since we have just one camera, this index can basically be neglected. Fig. 11 illustrates the ray-tracing diagram of the used gaze tracking model [14].

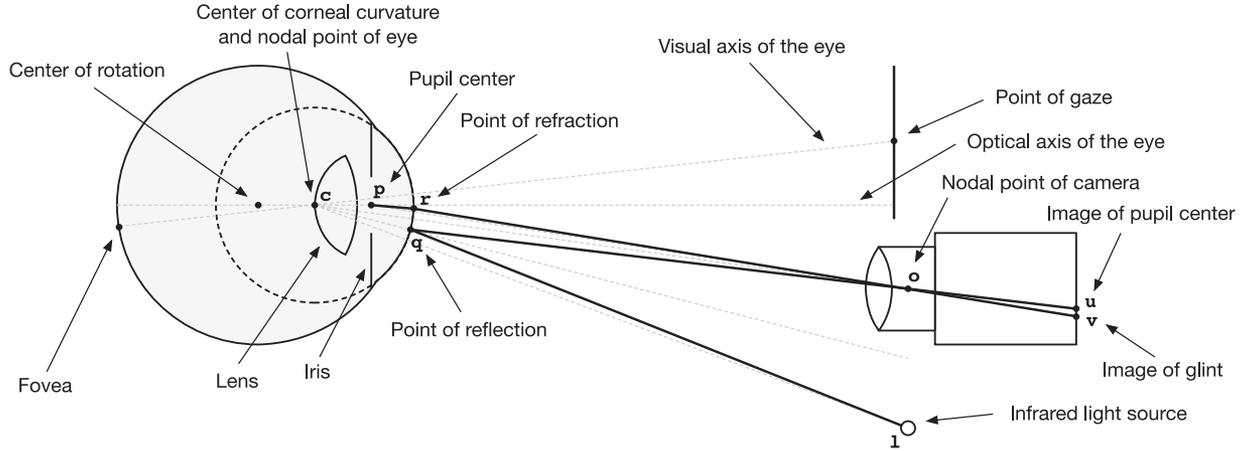


Fig. 11. Ray-tracing diagram of the 3-D gaze tracking model based on to the scheme in [14].

Reconstruction of optical axis (System 1):

$$\mathbf{q}_{ij} = \mathbf{o}_j + k_{q,ij} \cdot (\mathbf{o}_j - \mathbf{u}_{ij}) \quad (15)$$

$$\|\mathbf{q}_{ij} - \mathbf{c}\| = R \quad (16)$$

$$\begin{aligned} & (\mathbf{l}_i - \mathbf{q}_{ij}) \bullet (\mathbf{q}_{ij} - \mathbf{c}) \cdot \|\mathbf{o}_j - \mathbf{q}_{ij}\| \\ &= (\mathbf{o}_j - \mathbf{q}_{ij}) \bullet (\mathbf{q}_{ij} - \mathbf{c}) \cdot \|\mathbf{l}_i - \mathbf{q}_{ij}\| \end{aligned} \quad (17)$$

$$\mathbf{c} - \mathbf{o} = k_{c,b} \mathbf{b}_{\text{norm}} \quad (18)$$

$$\mathbf{b}_{\text{norm}} = \frac{\mathbf{b}}{\|\mathbf{b}\|} \quad (19)$$

$$\mathbf{b} = [(\mathbf{l}_1 - \mathbf{o}) \times (\mathbf{u}_1 - \mathbf{o})] \times [(\mathbf{l}_2 - \mathbf{o}) \times (\mathbf{u}_2 - \mathbf{o})]. \quad (20)$$

Reconstruction of optical axis (System 2):

$$\mathbf{r}_j = \mathbf{o}_j + k_{r,j} \cdot (\mathbf{o}_j - \mathbf{v}_j) \quad (21)$$

$$\|\mathbf{r}_j - \mathbf{c}\| = R. \quad (22)$$

Reconstruction of optical axis (System 3):

$$(\mathbf{r}_j - \mathbf{o}_j) \times (\mathbf{c} - \mathbf{o}_j) \bullet (\mathbf{p} - \mathbf{o}_j) = 0 \quad (23)$$

$$\begin{aligned} & n_1 \cdot \|(\mathbf{r}_j - \mathbf{c}) \times (\mathbf{p} - \mathbf{r}_j)\| \cdot \|\mathbf{o}_j - \mathbf{r}_j\| \\ &= n_2 \cdot \|(\mathbf{r}_j - \mathbf{c}) \times (\mathbf{o}_j - \mathbf{r}_j)\| \cdot \|\mathbf{p} - \mathbf{r}_j\| \end{aligned} \quad (24)$$

$$\|\mathbf{p} - \mathbf{c}\| = K. \quad (25)$$

Reconstruction of visual axis

$$\frac{\mathbf{p} - \mathbf{c}}{\|\mathbf{p} - \mathbf{c}\|} = \begin{bmatrix} \cos(\varphi_{\text{eye}}) \cdot \sin(\theta_{\text{eye}}) \\ \sin(\varphi_{\text{eye}}) \\ -\cos(\varphi_{\text{eye}}) \cdot \cos(\theta_{\text{eye}}) \end{bmatrix} \quad (26)$$

$$\mathbf{g} = \mathbf{c} + k_g \cdot \begin{bmatrix} \cos(\varphi_{\text{eye}} + \beta_{\text{eye}}) \cdot \sin(\theta_{\text{eye}} + \alpha_{\text{eye}}) \\ \sin(\varphi_{\text{eye}} + \beta_{\text{eye}}) \\ -\cos(\varphi_{\text{eye}} + \beta_{\text{eye}}) \cdot \cos(\theta_{\text{eye}} + \alpha_{\text{eye}}) \end{bmatrix} \quad (27)$$

$$k_g = \frac{c_z}{\cos(\varphi_{\text{eye}} + \beta_{\text{eye}}) \cdot \cos(\theta_{\text{eye}} + \alpha_{\text{eye}})}. \quad (28)$$

ACKNOWLEDGMENT

The authors would like to thank the mechanical engineering team at PSI who offered fast and creative solutions. In addition, we thank all the volunteers who participated patiently. Further thanks go to the members of medical image analysis center of the University of Basel. They supported this work with their deep and broad knowledge and with helpful feedback.

REFERENCES

- [1] E. Egger *et al.*, "Eye retention after proton beam radiotherapy for uveal melanoma," *Int. J. Radiation Oncol. Biol. Phys.*, vol. 55, no. 4, pp. 867–880, 2003.
- [2] A. Ramasubramanian and C. Shields, *Retinoblastoma*. New Delhi, India: Jaypee Brothers, Medical Publishers, 2012.
- [3] M. Goitein and T. Miller, "Planning proton therapy of the eye," *Med. Phys.*, vol. 10, no. 3, pp. 275–283, 1983.
- [4] D. Shin *et al.*, "Eye tracking and gating system for proton therapy of orbital tumors," *Med. Phys.*, vol. 39, no. 7, pp. 4265–4273, 2012.
- [5] M. B. Rügsegger *et al.*, "Noninvasive referencing of intraocular tumors for external beam radiation therapy using optical coherence tomography: A proof of concept," *Med. Phys.*, vol. 41, no. 8, pp. 081704-1–081704-9, 2014.
- [6] A. Fassi *et al.*, "Optical eye tracking system for noninvasive and automatic monitoring of eye position and movements in radiotherapy treatments of ocular tumors," *Appl. Opt.*, vol. 51, no. 13, pp. 2441–2450, May 2012.
- [7] R. Via *et al.*, "Optical eye tracking system for real-time noninvasive tumor localization in external beam radiotherapy," *Med. Phys.*, vol. 42, no. 5, pp. 2194–2202, 2015.
- [8] M. B. Rügsegger *et al.*, "Statistical modeling of the eye for multimodal treatment planning for external beam radiation therapy of intraocular tumors," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 84, no. 4, pp. e541–e547, 2012.
- [9] S. De Zanet *et al.*, "Landmark detection for fusion of fundus and MRI toward a patient-specific multimodal eye model," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 532–540, Feb. 2015.
- [10] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [11] J. Salvi *et al.*, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognit.*, vol. 35, no. 7, pp. 1617–1635, 2002.
- [12] G. Bradski, "The opencv library," *Doctor Dobbs J.*, vol. 25, no. 11, pp. 120–126, 2000.
- [13] D. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.
- [14] E. Guestrin and E. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, Jun. 2006.
- [15] W. Lotmar, "Theoretical eye model with aspherics," *J. Opt. Soc. Amer.*, vol. 61, no. 11, pp. 1522–1529, Nov. 1971.



Stephan Wyder was born in Bern, Switzerland, in 1982. He received the B.Sc. degree in computer science from the Bern University of Applied Sciences, Bern, Switzerland, in 2008, and the M.Sc. degree in biomedical engineering from the University of Bern, Bern, Switzerland, in 2012.

Furthermore, he has more than five years of work experience in the ophthalmic industry, where he worked on the development and integration of gaze tracking systems. He joined the Faculty of Medicine, Basel University in 2013 to work on the research

project *Computer Assisted Proton Beam Radiotherapy for Intraocular Tumors*, funded by Swiss National Science Foundation, where he is responsible for the subproject *Gaze Tracking Based Intraocular Tumor Targeting for Proton Beam Radiotherapy*.

Fabian Hennings photograph and biography not available at the time of publication.



Simon Pezold was born in Ludwigsburg, Germany, in 1985. He received the *Diplom* degree (corresponding to a combined BS/MS) in medical informatics from the University of Heidelberg, Heidelberg, Germany, in 2010. He is currently pursuing the Ph.D. degree in biomedical engineering at the University of Basel, Basel, Switzerland.

His research interests include medical image segmentation and medical applications of computer-assisted navigation.

Jan Hrbacek photograph and biography not available at the time of publication.



Philippe C. Cattin was born in Switzerland in 1967. He received the B.Sc. degree from the University of Applied Science in Brugg/Windisch in 1991. In 1995 he received the M.Sc. degree in computer science and in 2003 the Ph.D. degree in robotics from ETH Zürich, Switzerland. From 2003 to 2007 he was a Postdoctoral Fellow with the Computer Vision Laboratory at ETH Zürich. In 2007 he became an Assistant Professor at the University of Basel and was promoted to Associate Professor in 2015. He is the founder of the Medical Image Analysis Center at the

Medical Faculty of the University of Basel. He is currently the head of the recently founded Department of Biomedical Engineering at the University of Basel.

His research interests include medical image analysis, image-guided therapy and robotics-guided laser osteotomy. As a Principal Investigator he has finished many projects in these areas and published over 100 papers, patents and book chapter.