Stereo Eye Tracking with a Single Camera for Ocular Tumor Therapy

Stephan Wyder and Philippe C. Cattin

University of Basel, Department of Biomedical Engineering, Basel, Switzerland
{stephan.wyder,philippe.cattin}@unibas.ch

Abstract. We present a compact and accurate stereo eye tracking system using only one physical camera. The proposed eye tracking system is intended as a navigation system for ocular tumor therapy. There, the available physical space to mount an eye tracker is limited. Furthermore, high system accuracy is demanded. However, high eye tracker accuracy and system compactness often disagree. Current established eye trackers can live with that compromise, desktop devices focus more on accuracy whereas mobile devices focus on compactness. We combine a stereo eye tracking algorithm with a clever arrangement of two planar mirrors and a single camera to get high accuracy, precision and a compact design altogether. We developed an eye tracking prototype and tested the system with ten healthy volunteers. We show that the proposed eye tracker is more accurate and robust, while at the same time equally compact as a comparable eye tracking system containing one instead of two mirrors.

1 Introduction

Ocular tumors are a severe disease that may lead to blindness or even death if left untreated. Nowadays, specialists successfully treat the disease by radiating the patient's primary tumor with charged particles [3]. There is, however, a drawback: Although the tumor radiation itself is noninvasive, an invasive patient preparation is required. A surgeon thereby sutures radio-opaque clips on the outer scleral surface of the diseased eye. These clips are used to target the tumor during radiation therapy.

By introducing an eye tracker, also referred to as gaze trackers, into the current treatment workflow, clip surgery could be avoided [9, 10]. Eye trackers are devices able to estimate where a person is looking, i.e. the point of gaze [5], [7]. Certain eye trackers are based on a 3D model and therefore even have the ability to estimate the location of the eye in 3D space. This property enables us to use an eye tracker as a navigation system, namely to localize the eye and target the tumor during radiation therapy [9, 10].

The following eye tracker properties are important for this medical application: High accuracy and precision, computational speed and stability, and compact hardware design, enabling its integration into the radiation facility. However, eye tracker accuracy and compactness often disagree.

X. Chen, M. K. Garvin, J. Liu, E. Trucco, Y. Xu (Eds.): OMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, Iowa Research Online, pp. 81–88, 2016. Available from: http://ir.uiowa.edu/omia/2016_Proceedings/2016/

We present a new type of eye tracking system, extending an existing solution [10], where we are able to drastically increase accuracy and stability without having a significantly bigger device. The proposed eye tracker consists of one physical camera and we complement it with two planar mirrors. The integration into the treatment workflow remains unaffected and is done as described in [10].

By observing a scene (eye) over two mirrors, stereo images can be captured with a single camera (catadioptric stereo [2], [8]). This construction enables us to make the device compact and accurate at the same time. This is because we can optimally deflect different optical paths between the eye and the camera with the introduced mirrors. Furthermore, our point of gaze estimation is very accurate and the eye position estimation is very precise due to the virtual stereo camera frame (triangulation).

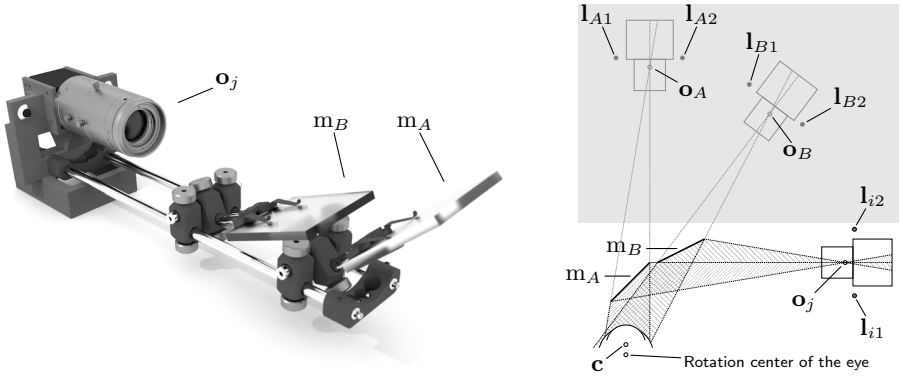
We tested our eye tracker with ten healthy volunteers and we show that our system is more accurate and reliable than a comparable eye tracking device [10], containing one instead of two planar mirrors. The proposed eye tracker is primarily designed and developed for ocular tumor therapy. However, we believe that this idea can easily be adapted for any other application, where compactness, high accuracy, and flexibility for the integration are demanded.

2 Methods

Hardware Setup and Calibration. The eye tracker (Fig. 1a) consists of an industry camera (XIMEA MQ022RG-CM), a 50 mm lens with an infrared-pass filter ($\lambda = 650\,\text{nm}$), two infrared LEDs ($\lambda = 860\,\text{nm}$), and two hot-mirrors (infrared reflection from 750 nm to 1125 nm). All components are mounted on rails with optomechanical holders. Using a mirror in general enables to optimally deflect the optical path between the eye and the camera. Therefore, we can place the camera where physical space is available (Fig. 1b). A hot-mirror, in our special case, is coated to reflect infrared waves and to transmit the visible part of the light-spectrum. It has the advantage that it does not obscure the view of the subject's eye, which might be only a couple of centimeters behind one of the hot-mirrors. Two infrared LEDs are used to illuminate the scene (eye) and to produce reflections on the cornea of the subject (glints). The LED positions are calibrated in advance and given by design.

To enable estimating the point of gaze and the eye position, we first need to know the absolute positions of the virtual cameras (nodal points / optical centers $\mathbf{o}_A$, $\mathbf{o}_B$) and the virtual LEDs ($\mathbf{l}_{A1}$, $\mathbf{l}_{A2}$, $\mathbf{l}_{B1}$, $\mathbf{l}_{B2}$). Therefore, the camera-mirror setup needs to be calibrated to get the intrinsic camera parameters.

The following steps describe the calibration procedure: (**i**) First, the operator arranges the hot-mirror positions and the tilting angles, in order to have a good camera view onto the target eye. (**ii**) The camera focus and aperture have to be adjusted properly. (**iii**) The operator then calibrates the camera once for a certain focus/aperture adjustment to get the intrinsic camera parameters [6], [11]. (**iv**) At this stage, all images coming from the physical camera $\mathbf{o}_j$ get undistorted to correct for lens errors. (**v**) Additionally, all images get flipped



(a) Eye tracker consisting of an industry camera, a lens with an infrared-pass filter, two infrared LEDs, and two hot-mirrors (m_A,m_B). All parts are mounted with either a plastic holder or optomechanical components on rigid rails. The rails also serve as fixation point, such that the tracker can be attached to the treatment apparatus.

(b) Eye nodal point ($\mathbf{c}$), camera ($\mathbf{o}_j$), LEDs ($\mathbf{l}_{i1}$,$\mathbf{l}_{i2}$), hot-mirrors, virtual cameras ($\mathbf{o}_A$,$\mathbf{o}_B$), and virtual LEDs ($\mathbf{l}_{A1}$,$\mathbf{l}_{A2}$,$\mathbf{l}_{B1}$,$\mathbf{l}_{B2}$). The gray shaded box corresponds to physical space occupied by the treatment device.

Fig. 1: Eye tracking hardware (left) and optical arrangement (right)

horizontally in order to have image sections as they would have been made by the virtual cameras behind the mirrors [1]. The two mirrors are placed such, that after the flipping, the left part of the image looks like it would have been taken from $\mathbf{o}_B$, and the right half of the image, as it would have been taken from $\mathbf{o}_A$ (Fig. 1b). (**vi**) To get the absolute positions of the virtual cameras, the operator acquires the appropriate extrinsic camera matrices (homographies [6], [11]) using an image of a checkerboard. The checkerboard itself is co-registered with the world coordinate system of the treatment device. It has approximately the size of an eye and is placed at the same location. (**vii**) Having the two homographies H_A and H_B, the positions of the virtual LEDs ($\mathbf{l}_{A1}$, $\mathbf{l}_{A2}$, $\mathbf{l}_{B1}$. $\mathbf{l}_{B2}$) and the virtual cameras ($\mathbf{o}_A$, $\mathbf{o}_B$) can be determined.

Semi-Supervised Eye Feature Detection. Figure 2 shows a typical image (input) with the overlaid detected features (output) and the appropriate labels after undistortion and horizontal flipping.

To get all required eye features, (**i**) the operator sets two regions of interest (ROI_A, ROI_B) for the current eye tracking session. (**ii**) A couple of seed points have to be set for both ROIs within the pupil area, the iris area and the glints. (**iii**) The algorithm averages the individual seed point sets and defines from that an individual pupil- and glint-threshold for both ROIs: The threshold for the glints is defined by the average of the glint seed point values plus a certain tolerance ($\pm\frac{3}{256}$). The threshold for the pupil consists of the arithmetic mean between

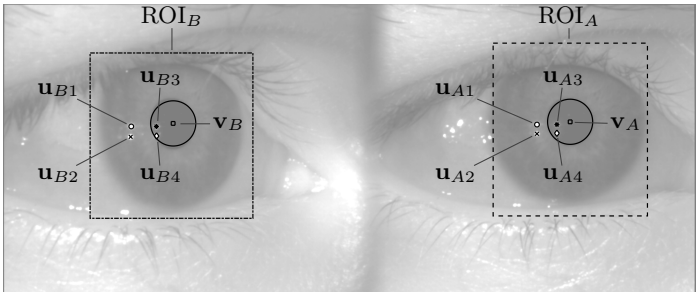


Fig. 2: Virtual stereo camera view onto the same eye with ROI$_A$, ROI$_B$, pupil centers ($\mathbf{v}_A$, $\mathbf{v}_B$), and glints ($\mathbf{u}_{B1}$, $\mathbf{u}_{B2}$, $\mathbf{u}_{B3}$, $\mathbf{u}_{B4}$, $\mathbf{u}_{A1}$, $\mathbf{u}_{A2}$, $\mathbf{u}_{A3}$, $\mathbf{u}_{A4}$)

the average of the iris seed points and the average of the pupil seed points, again plus the mentioned tolerance. Furthermore, (**iv**) the algorithm thresholds the image with the previously set threshold parameters. This results in four binary images, one with glint candidates and one with pupil candidates, both for ROI$_A$ and ROI$_B$. After this, (**v**) a standard 8-connected-component labeling algorithm gets applied on all of the binary images to identify the individual blobs. (**vi**) All the pupil blobs get post-processed with a morphological closing operator to make their border regions more homogeneous. Next, (**vii**) the algorithm extracts the size and the eccentricity (measure of roundness of a certain area) for every blob. (**viii**) All pupil blobs that have an eccentricity above 0.75 are discarded (circle: 0, ellipse: < 1). (**ix**) From the remaining blobs, the algorithm takes the biggest one, builds the convex hull and calculates the centroid of it (pupil center). (**x**) To get the required glint centers, the algorithm extracts the centroids of the glint candidates. The four glints closest to the pupil center are taken and sorted such, that we have them for both ROIs in the order: top-left, bottom-left, top-right, bottom-right. Having this, the algorithm assigns the appropriate labels to the centroids of the glints (Fig. 2). (**xi**) All the coordinates of the extracted features (2D projections) get transformed to the camera coordinate system (3D points) [10]. Afterwards they get transformed with the appropriate homographies H_A and H_B from the camera coordinate system into the world coordinate system, depending on whether they were detected in ROI$_A$ or in ROI$_B$.

Eye Position and Point of Gaze Estimation. The gaze tracking model is based on the method from Guestrin et al. [4], adapted for the catadioptric setup. At this stage of the eye tracking procedure, we have a couple of points in 3D space. Some of them are determined during the hardware calibration, some points were gathered during the eye feature detection. Using these known points we can calculate the eye position and the point of gaze.

All points and vectors are denoted with small bold letters and are $\in \mathbb{R}_3$. Known points are: $\mathbf{l}_i$ (light sources), $\mathbf{o}_j$ (nodal points of cameras), $\mathbf{u}_{ij}$ (images of glints on sensor), $\mathbf{v}_j$ (images of pupil on sensor), where i encodes the light sources and j encodes the cameras (Fig. 1b and Fig. 2).

First, the algorithm estimates the nodal point of the eye $\mathbf{c}$ by bringing some of the known points into relation. Coplanarity of points $\mathbf{l}_i$, $\mathbf{o}_j$, $\mathbf{u}_{ij}$, $\mathbf{c}$ can be described with the triple product:

$$\underbrace{(\mathbf{l}_i - \mathbf{o}_j) \times (\mathbf{u}_{ij} - \mathbf{o}_j)}_{\mathbf{w}_{ij}} \bullet (\mathbf{c} - \mathbf{o}_j) = 0 \quad \Leftrightarrow \quad \mathbf{w}_{ij} \bullet (\mathbf{c} - \mathbf{o}_j) = 0\,. \tag{1}$$

Making use of the distributive property, we get:

$$\mathbf{w}_{ij} \bullet (\mathbf{c} - \mathbf{o}_j) = 0 \quad \Leftrightarrow \quad \mathbf{w}_{ij} \bullet \mathbf{c} - \mathbf{w}_{ij} \bullet \mathbf{o}_j = 0\,. \tag{2}$$

Because $a \bullet b = a^T \cdot b$, we can write the above equation in matrix form:

$$\mathbf{w}_{ij}^T \cdot \mathbf{c} = \mathbf{w}_{ij} \bullet \mathbf{o}_j\,, \tag{3}$$

$$\underbrace{\begin{bmatrix} [(\mathbf{l}_{A1} - \mathbf{o}_A) \times (\mathbf{u}_{A4} - \mathbf{o}_A)]^T \\ [(\mathbf{l}_{A1} - \mathbf{o}_B) \times (\mathbf{u}_{B4} - \mathbf{o}_B)]^T \\ \vdots \\ [(\mathbf{l}_{B2} - \mathbf{o}_A) \times (\mathbf{u}_{A1} - \mathbf{o}_A)]^T \\ [(\mathbf{l}_{B2} - \mathbf{o}_B) \times (\mathbf{u}_{B2} - \mathbf{o}_B)]^T \end{bmatrix}}_{\mathrm{M}_2} \cdot \mathbf{c} = \underbrace{\begin{bmatrix} (\mathbf{l}_{A1} - \mathbf{o}_A) \times (\mathbf{u}_{A4} - \mathbf{o}_A) \bullet \mathbf{o_A} \\ (\mathbf{l}_{A1} - \mathbf{o}_B) \times (\mathbf{u}_{B4} - \mathbf{o}_B) \bullet \mathbf{o_B} \\ \vdots \\ (\mathbf{l}_{B2} - \mathbf{o}_A) \times (\mathbf{u}_{A1} - \mathbf{o}_A) \bullet \mathbf{o_A} \\ (\mathbf{l}_{B2} - \mathbf{o}_B) \times (\mathbf{u}_{B2} - \mathbf{o}_B) \bullet \mathbf{o_B} \end{bmatrix}}_{\mathbf{h}}\,. \tag{4}$$

Every row in the above system of equations represents one correspondence between a virtual light source and an image of a glint on one of the virtual camera sensors. This overdetermined system of linear equations can be solved with least squares:

$$\mathrm{M}_2 \cdot \mathbf{c} = \mathbf{h} \quad \Rightarrow \quad \mathbf{c} = (\mathrm{M}_2^{\mathrm{T}} \mathrm{M}_2)^{-1} \cdot \mathrm{M}_2^{\mathrm{T}} \mathbf{h}\,. \tag{5}$$

Having the point $\mathbf{c}$ (nodal point of the eye), the algorithm calculates $\overrightarrow{\mathbf{cp}}$, the geometrical axis of the eye, defined by the points $\mathbf{c}$ and $\mathbf{p}$ (pupil center). The geometrical axis of the eye $\overrightarrow{\mathbf{cp}}$ can also be seen as the line of intersection of two planes, defined by $\mathbf{o}_A$, $\mathbf{v}_A$, $\mathbf{c}$ and $\mathbf{o}_B$, $\mathbf{v}_B$, $\mathbf{c}$:

$$\overrightarrow{\mathbf{cp}} = [(\mathbf{o}_A - \mathbf{v}_A) \times (\mathbf{c} - \mathbf{o}_A)] \times [(\mathbf{o}_B - \mathbf{v}_B) \times (\mathbf{c} - \mathbf{o}_B)]\,, \tag{6}$$

$$\mathbf{s} \coloneqq \overrightarrow{\mathbf{cp}} / \|\overrightarrow{\mathbf{cp}}\|\,. \tag{7}$$

The geometrical axis is only plausible (and accurate), when the mentioned planes have different orientation. Otherwise their normalized normals ($\hat{\mathbf{n}}_1$, $\hat{\mathbf{n}}_2$) are parallel or almost parallel and no (accurate) intersection can be calculated. We detect this special case by calculating the norm of the difference vector between $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$. When this mentioned norm is below a certain tolerance (0.2), the result is marked as implausible (inaccurate).

The geometrical axis can also be expressed with a tilt and a shift angle (φ, θ), relative to the world coordinate system:

$$\varphi = \sin^{-1}(-\mathbf{s}_y), \quad \theta = \sin^{-1}(-\mathbf{s}_x / \cos(\varphi))\,. \tag{8}$$

The visual axis of the eye (line-of-sight) is constructed by connecting the nodal point of the eye with the fovea, the point of sharpest vision on the retina. The subject-specific deviation between the geometrical axis and the visual axis can again be expressed with two angles α and β, for which the algorithm takes tabulated standard values as a first estimate ($\alpha = \pm 5°$, $\beta = 1.5°$). Consequently, the point of gaze is defined by the intersection of the visual axis with a certain plane in space, a computer screen or any other plane containing at least one calibration point. Assuming that the plane of intersection corresponds to the xy-plane ($z = 0$), the point of gaze is:

$$k = \mathbf{c}_z / [\cos(\varphi + \beta) \cdot \cos(\theta + \alpha)] \,, \tag{9}$$

$$\textit{Point of gaze} = \mathbf{c} - k \cdot \begin{bmatrix} \cos(\varphi + \beta) \cdot \sin(\theta + \alpha) \\ \sin(\varphi + \beta) \\ \cos(\varphi + \beta) \cdot \cos(\theta + \alpha) \end{bmatrix} . \tag{10}$$

Having at least one calibration point with well known coordinates, α and β can be calculated, when comparing the true calibration point position with the estimated point of gaze position (calibration of subject-specific parameters) [4].

3 Results

We mounted the eye tracker on an optical table, equipped with 15 calibration points at well defined positions. Additionally, the optical table contained a calibration checkerboard, which defined the world coordinate system and was used to determine the absolute positions of the virtual cameras and the virtual LEDs. Our experimental setup is derived from the treatment facility and enables direct comparison. Having the system calibrated, which takes a few minutes, we let ten healthy volunteers fixate the predefined calibration points using an ophthalmic chin rest. The corresponding images were recorded with the camera. Afterwards, we calculated the point of gaze for every image i and decided whether the estimate is plausible or not with the criterion mentioned above. On average, we discarded five points per volunteer. If plausibility was given, we recorded on the one side the calculated α and β values and the deviation between the estimated point of gaze and the true calibration point location (point of gaze error). The recorded α_i and β_i were averaged per volunteer and used for a second evaluation with the new subject-specific parameters ($\overline{\alpha_i}$, $\overline{\beta_i}$). The resulting point of gaze errors of the second (calibrated) round of evaluation were transformed from millimeters to degrees (measured at point $\mathbf{c}$) and visualized in Fig. 3. This conversion makes the result independent of a specific geometrical setup. Consequently, it enables us to compare our result with the results from [10] and any other eye tracker accuracy. The average error over all volunteers measured at point $\mathbf{c}$ is below 0.96 °. Beside the good point of gaze accuracy, we observed a high precision in point $\mathbf{c}$ estimation. We looked at the bounding boxes containing all points $\mathbf{c}$ per volunteer (one estimation per calibration point). The mean bounding box over all volunteers had the dimensions: $\Delta x = 2.04\,\text{mm}$, $\Delta y = 1.72\,\text{mm}$,

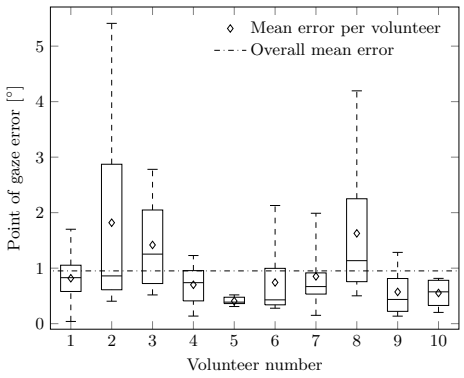


Fig. 3: Point of gaze error

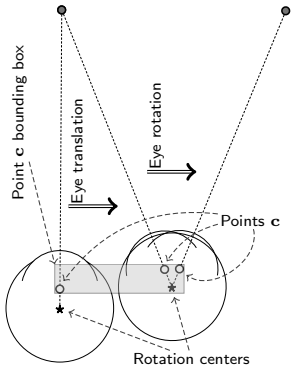


Fig. 4: Eye translation & rotation

and $\Delta z = 1.52\,\text{mm}$. A displacement of point $\mathbf{c}$ in the xy-plane of about 0.8 mm has to be expected, even when the eye is not translated but only rotated. This is because the nodal point $\mathbf{c}$ does not correspond to the rotation center of the eye (Fig. 4). Hence, one portion of the mentioned Δ-values is assumed to originate from slight eye translation.

The required time to process one image was on average 100 ms for the feature extraction and below 1 ms for calculating the eye position and the point of gaze. Our algorithm is implemented in MATLAB so far, intended for post-processing of already recorded images.

4 Discussion

The achieved point of gaze accuracy is more than 1° better as compared to [10], where only one instead of two mirrors was used. The precision of the eye position estimation is very high, and especially better in the depth compared to the reference system [10]. This can be explained by the triangulation angle, which is mainly responsible for the depth information and which is wider in the proposed solution due to the virtual stereo frame. Additionally, the proposed algorithm is faster and more robust, because no optimization of nonlinear systems is required. Furthermore, the patient specific calibration can be achieved with one calibration point and does not require a time-consuming procedure.

By using a setup with two hot-mirrors, we combine the benefits of a single camera system (only one camera to calibrate, no camera synchronization needed) with the advantages of a stereo camera system (simplified eye tracking model, fewer patient specific parameters, simplified calibration algorithm, better accuracy and precision).

The only limitation of our method, as mentioned above, is the fact that problematic geometrical constellations can occur when the geometrical axis $\overrightarrow{\mathbf{cp}}$ points towards the line connecting both virtual cameras (implausible result). This problematic constellation can be limited or even completely avoided by

tilting the individual mirrors or the whole eye tracker such, that the mentioned line connecting the virtual cameras does not cross the area, where high accuracy point of gaze estimation is required. This requirement is easy to fulfill in ocular tumor therapy and therefore no limitation.

5 Conclusion

Ocular tumor therapy can considerably be improved by integrating an eye tracking system. The whole tumor therapy can potentially be made noninvasive [10]. For this, we developed a novel stereo eye tracker with a single physical camera. Stereo eye tracking is more accurate and stable than an eye tracker based on a single camera. Our setup with two mirrors and one camera has several advantages as compared to a setup with two physical cameras: It is more compact, camera synchronization is not needed, and only one camera has to be calibrated. Our results show that we are much more accurate than with a conventional single camera setup. Therefore, our proposed eye tracker is eminently suited for ocular tumor navigation and other applications where both compactness and accuracy are needed.

References

1. Corballis, M.C.: Much ado about mirrors. Psychonomic Bulletin & Review 7(1), 163–169 (2000)
2. Gluckman, J., Nayar, S.K.: Catadioptric Stereo Using Planar Mirrors. International Journal of Computer Vision 44(1), 65–79 (2001)
3. Goitein, M., Miller, T.: Planning proton therapy of the eye. Medical Physics 10(3), 275–283 (May 1983)
4. Guestrin, E.D., Eizenman, M.: General theory of remote gaze estimation using the pupil center and corneal reflections. Biomedical Engineering, IEEE Transactions on 53(6), 1124–1133 (Jun 2006)
5. Hansen, D.W., Ji, Q.: In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32(3), 478–500 (Mar 2010)
6. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision (2003)
7. Narcizo, F.B., Rangel de Queiroz, J.E., Gomes, H.M.: Remote Eye Tracking Systems: Technologies and Applications. In: 2013 26th Conference on Graphics, Patterns and Images Tutorials. pp. 15–22. IEEE (2013)
8. Nene, Sameer A, Nayar, Shree K: Stereo with mirrors. In: Computer Vision, 1998. Sixth International Conference on. pp. 1087–1094. IEEE (1998)
9. Via, R., Fassi, A., Fattori, G., Fontana, G., Pella, A., Tagaste, B., Riboldi, M., Ciocca, M., Orecchia, R., Baroni, G.: Optical eye tracking system for real-time noninvasive tumor localization in external beam radiotherapy. Medical Physics 42(5), 2194–2202 (May 2015)
10. Wyder, S., Hennings, F., Pezold, S., Hrbacek, J., Cattin, P.: With Gaze Tracking Towards Noninvasive Eye Cancer Treatment. Biomedical Engineering, IEEE Transactions on PP(99), 1–1 (2015)
11. Zhang, Z.: A flexible new technique for camera calibration. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(11), 1330–1334 (Nov 2000)