Real-time algorithms for head mounted gaze tracker

Aleksey Starostenko
Moscow Institute of Physics and Technology
Dolgoprudny, Moscow Region
carrlit@mail.ru

Filipp Kozin
Moscow Institute of Physics and Technology
Dolgoprudny, Moscow Region
filipp.kozin@gmail.com

Roman Gorbachev
Moscow Institute of Physics and Technology
Dolgoprudny, Moscow Region
r.gorbachev@gmail.com

***Abstract*— We introduce a set of real-time algorithms for head mounted gaze tracker consisting of three cameras: two cameras for the eyes and one camera for the scene. The direction of the optical axis of the eye in three-dimensional space is calculated using the reflection of IR LEDs from the cornea. Individual features of the user are taken into account using the short-term calibration procedure. The described algorithms combine high accuracy in determining the point of gaze with high speed.**

The procedure for determining the point of gaze consists of the following algorithms:

estimation of the position of the pupils on the eye cameras frames using of the threshold processing taking into account the histogram of the frame and further approximation of the pupil by an ellipse;

estimation of the IR LEDs glare position on the frames of the eye cameras using threshold processing;

filtration of the glares by brightness, size, circularity, and of the glares beyond the iris, the size of the iris is estimated by the distance from eye camera to pupil position calculated on the previous frame;

indexation of the glares with the template matching;

estimation of the optical axis angles of the eye using a spherical model of the cornea with the nonlinear optimization methods;

estimation of the point of gaze on the scene camera frame using individual user features found during the calibration process.

During calibration, the movement of the ArUco calibration mark and its selection on the scene camera frame are used. To calculate the gaze position on the scene camera, a regression algorithm is used, which implicitly takes into account the individual characteristics of the user.

Keywords — gaze estimation, head mount devices, object detection and tracking, human-computer interaction

I. INTRODUCTION

Different methods of pointing at objects were available to humanity since ancient times. At pre-digital era, people were pointing with hands, pointers, or voice. With the emerging of digital devices a computer mouse and a sensor screen have been added to these common methods. In recent years, there has been a greater effort to explore such new pointing methods as brain-computer interfaces [1] or point of gaze [2]. Price and usability for new pointing methods are also actual problems.

Approaches to determining the point of gaze have been studied in laboratory conditions since the end of the 19th century using wearable lenses, IR LED reflections (glints), stationary and wearable devices, usually called trackers. Nowdays, a number of wearable trackers have been developed [3]. The area of application of trackers is very extensive: neuroscience, psychology, industrial engineering, marketing, computer science [4].

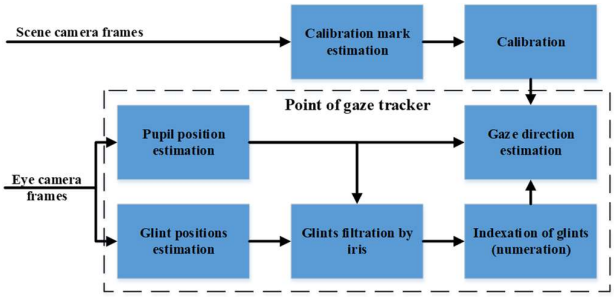

Fig. 1. Algorithm dependency diagram.

This article presents a set of algorithms for determining the point of gaze in real time for a head mounted tracker consisting of three cameras: two cameras for the eyes and one scene camera. One of the most important tasks during the creation of the algorithm set was to obtain the high performance, which in turn allows one to reduce the cost of the device, as well as its power consumption and heat dissipation.

The article is organized as follows: section 2 discusses a set of algorithms used to determine the point of gaze, and section 3 presents the test results.

II. ALGORITHMS

Algorithm dependency diagram is presented in the Fig. 1. Main reason to use the information about IR LEDs positions in the set of algorithms is the opportunity to bind calibration result to specific person, which allows calibration not every time the tracker is used, but once for an individual user. Estimation of the pupil position has a significant impact on the final accuracy of the algorithm set; therefore, refinement stage is introduced.

A. Two-dimensional pupil position estimation

Estimation of the pupil position occur in two stages. At the first stage, the position of the pupil is preliminary estimated using threshold processing of the eye camera frame to detect clusters (blobs) and evaluate their parameters (Fig. 2). The selected clusters are filtered according to size, and then the darkest cluster is selected.

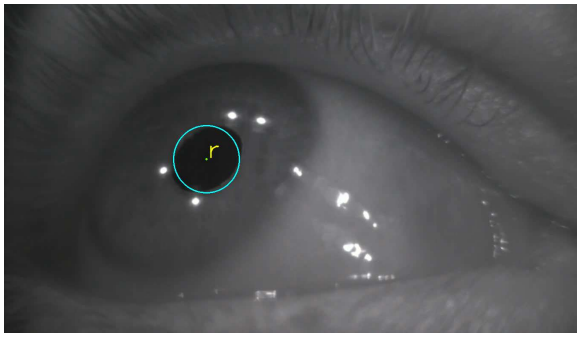

Fig. 2. Blob detection result.

978-1-7281-4326-2/19/$31.00 ©2019 IEEE
DOI 10.1109/IC-AIAI48757.2019.00025

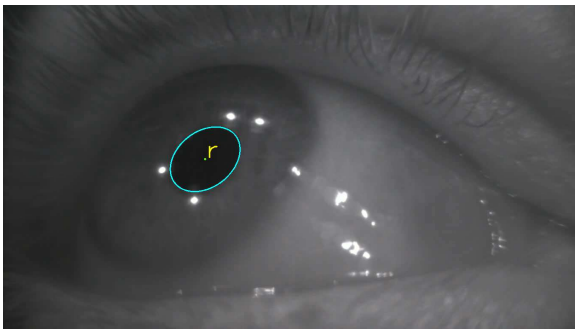

Fig. 3. Ellipse fitting result.

At the second stage, the center of the pupil is determined using the approximation of the pupil by an ellipse, which boundaries are found by threshold processing taking into account the histogram in the region of the previously found cluster (Fig. 3). In order to increase performance, preliminary finding of clusters is carried out on a frame with reduced size.

B. *Two-dimensional glints position estimation and indexation*

During the position estimation of IR LEDs glints on the eye cornea, threshold processing is also used with the selection of clusters [5] and their filtering by brightness, size and circularity.

The iris size is calculated using the average size of the human iris and the distance from the camera to the pupil calculated on the previous frame. Glints outside the iris are filtered out (Fig. 4).

The glints are indexed in the following order: from one of the upper pair closest to the bridge of the nose, in a circle, away from the bridge of the nose (clockwise for the right eye and counterclockwise for the left), see Fig. 5.

Preliminary, for each of the eyes, glints from one frame were indexed. The obtained coordinates in the reference system associated with the camera frame were saved as test ones. Automatic indexation of glints occurs in several passes. On the first pass, two upper glints are selected and, on the basis thereof, the transformation matrix from test indexation to current set of glints is constructed. Pattern matching quality is checked by the distance from the glints to the nearest test glints coordinate, as well as the slope and scale of the transformation matrix. On the second pass, two glints far from the nose are selected and procedure is repeated. If only one of the passes is successful, then its indexation result is chosen.

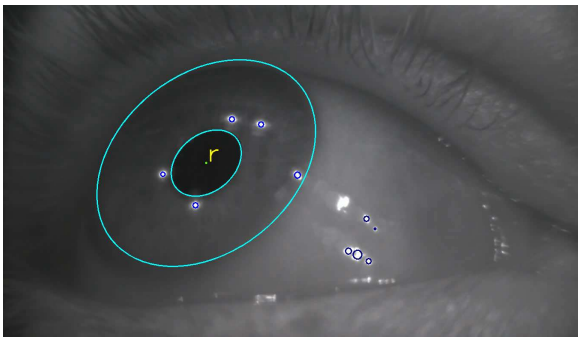

Fig. 4. Glints position estimation.

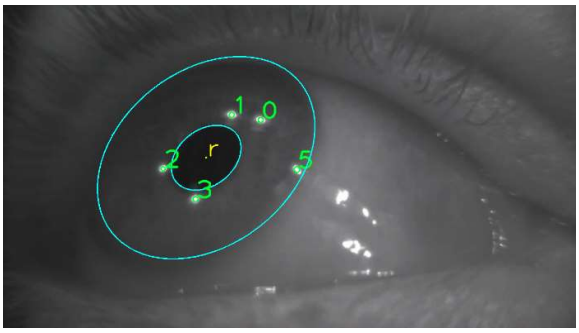

Fig. 5. Diode indexation example.

If both indexation are successful, then we take the average transformation matrix and carry out another indexing with it.

C. *Optical axis estimation*

Vector of optical axis of the eye $\boldsymbol{g}$ is determined by the equation:

$$\boldsymbol{g} = \boldsymbol{c} + k_g \frac{(\boldsymbol{p}-\boldsymbol{c})}{\|\boldsymbol{p}-\boldsymbol{c}\|} \tag{1}$$

where $\boldsymbol{c}$ – nodal point of the eye, $\boldsymbol{p}$ – pupil center, k_g – scalar parameter of linear equation.

Using equations from [2] a system of nonlinear equations is obtained:

$$\begin{cases} [(\boldsymbol{l}_i - \boldsymbol{o}) \times (\boldsymbol{u}_i - \boldsymbol{o})]^T \boldsymbol{b}_n = 0, \quad i = 1, \dots, N \\ \|\boldsymbol{b}_n\| - 1 = 0 \end{cases} \tag{2}$$

Where $\boldsymbol{l}_i$ – three-dimensional position of the i-th IR LED, $\boldsymbol{o}$ – position of the eye camera nodal point, $\boldsymbol{u}_i$ – position of the i-th IR LED image in the focal plane of the eye camera, $\boldsymbol{b}_n$ – direction vector to the center of the spherical surface of the cornea. This system is solved using the Levenberg-Marquardt algorithm for the coordinates of $\boldsymbol{b}_n$.

According to [2], system of the equations to find nodal point of the eye (see Fig. 6):

$$\begin{cases} \boldsymbol{q}_i = \boldsymbol{o} + k_i \frac{(\boldsymbol{o}-\boldsymbol{u}_i)}{\|\boldsymbol{o}-\boldsymbol{u}_i\|} \\ \|\boldsymbol{q}_i - \boldsymbol{c}\| = R \\ (\boldsymbol{l}_i - \boldsymbol{q}_i) \bullet (\boldsymbol{q}_i - \boldsymbol{c}) \cdot \|\boldsymbol{o} - \boldsymbol{q}_i\| = \\ = (\boldsymbol{o} - \boldsymbol{q}_i) \bullet (\boldsymbol{q}_i - \boldsymbol{c}) \cdot \|\boldsymbol{l}_i - \boldsymbol{q}_i\| \\ \boldsymbol{c} - \boldsymbol{o} = k\boldsymbol{b}_n \end{cases} \tag{3}$$

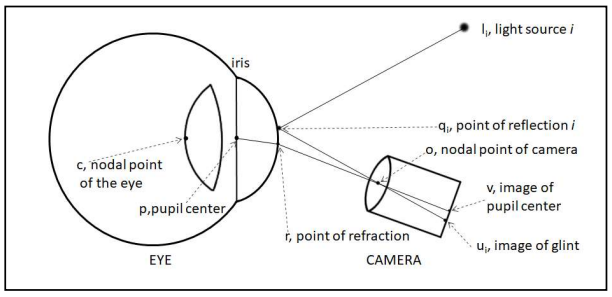

Fig. 6. Camera, light source and eye diagram.

is transformed to the system:

$$\begin{cases} \left\| k_i \frac{(\boldsymbol{o}-\boldsymbol{u_i})}{\|\boldsymbol{o}-\boldsymbol{u_i}\|} - k\boldsymbol{b}_n \right\| - R = 0 \\ \frac{\left(\boldsymbol{l_i}-\boldsymbol{o}-k_i\frac{(\boldsymbol{o}-\boldsymbol{u_i})}{\|\boldsymbol{o}-\boldsymbol{u_i}\|}\right)\left(k_i\frac{(\boldsymbol{o}-\boldsymbol{u_i})}{\|\boldsymbol{o}-\boldsymbol{u_i}\|}-k\boldsymbol{b}_n\right)}{\left\|\boldsymbol{l_i}-\boldsymbol{o}-k_i\frac{(\boldsymbol{o}-\boldsymbol{u_i})}{\|\boldsymbol{o}-\boldsymbol{u_i}\|}\right\|} + \frac{\frac{(\boldsymbol{o}-\boldsymbol{u_i})}{\|\boldsymbol{o}-\boldsymbol{u_i}\|}\left(k_i\frac{(\boldsymbol{o}-\boldsymbol{u_i})}{\|\boldsymbol{o}-\boldsymbol{u_i}\|}-k\boldsymbol{b}_n\right)}{\|\boldsymbol{o}-\boldsymbol{u_i}\|} = 0 \end{cases} \quad (4)$$

which is also solved using the Levenberg-Marquardt algorithm for lengths k, k_i, $i=1,...,N$. Here k – length of the vector of the cornea center, k_i, $i=1,...,N$ – lengths vector of reflection points of IR LEDs from the cornea, R – radius of the cornea spherical surface, $\boldsymbol{q_i}$, $i=1,...,N$ – point of reflection of i-th IR LED from cornea.

To find refraction point for the center of the pupil (Fig. 6), the system of equations is used [2]:

$$\begin{cases} \boldsymbol{r} = \boldsymbol{o} + k_r \frac{(\boldsymbol{o}-\boldsymbol{v})}{\|\boldsymbol{o}-\boldsymbol{v}\|} \\ \|\boldsymbol{r}-\boldsymbol{c}\| = R \\ (\boldsymbol{c}-\boldsymbol{o}) = k\boldsymbol{b}_n \end{cases} \quad (5)$$

this system is transformed into quadric equation:

$$k_r^2 - \frac{2kk_r}{\|\boldsymbol{o}-\boldsymbol{v}\|}(\boldsymbol{o}-\boldsymbol{v}) \bullet \boldsymbol{b}_n + k^2 - R^2 = 0 \quad (6)$$

which is solved for k_r. Here $\boldsymbol{v}$ is the pupil center image position in the focal plane of the eye camera, k_r is the length of the vector of pupil center refraction point. This quadratic equation describes the intersection of a line ($\boldsymbol{r}$-$\boldsymbol{o}$) and the corneal sphere, therefore, the distance to the point of refraction corresponds to a smaller root.

To calculate the position of the pupil center (see Fig. 6) in three-dimensional space in [2], a system of equations for ray refraction is used:

$$\begin{cases} (\boldsymbol{r}-\boldsymbol{o}) \times (\boldsymbol{c}-\boldsymbol{o}) \bullet (\boldsymbol{p}-\boldsymbol{o}) = 0 \\ n_1\|(\boldsymbol{r}-\boldsymbol{c}) \times (\boldsymbol{p}-\boldsymbol{r})\|\|\boldsymbol{o}-\boldsymbol{r}\| = \\ = n_2\|(\boldsymbol{r}-\boldsymbol{c}) \times (\boldsymbol{o}-\boldsymbol{r})\|\|\boldsymbol{p}-\boldsymbol{r}\| \\ \|\boldsymbol{p}-\boldsymbol{c}\| = K \end{cases} \quad (7)$$

however, this system could have up to four solutions; thus, a different algorithm is used to determine the pupil center vector. At first, one determines vector $\boldsymbol{a}$, perpendicular to the plane of refraction:

$$\boldsymbol{a} = (\boldsymbol{c}-\boldsymbol{r}) \times (\boldsymbol{r}-\boldsymbol{o}), \quad (8)$$

ray angle after refraction can be expressed as:

$$\sin(\beta) = \frac{n_1}{n_2}\frac{\boldsymbol{a}}{\|\boldsymbol{c}-\boldsymbol{r}\|\|\boldsymbol{r}-\boldsymbol{o}\|} \quad (9)$$

where n_1 and n_2 – absolute refractive indices of air and eyes. Therefore, quaternion of the ray refraction rotation Q is composed:

$$Q = \left(\cos\left(\frac{\beta}{2}\right), \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}\sin\left(\frac{\beta}{2}\right)\right). \quad (10)$$

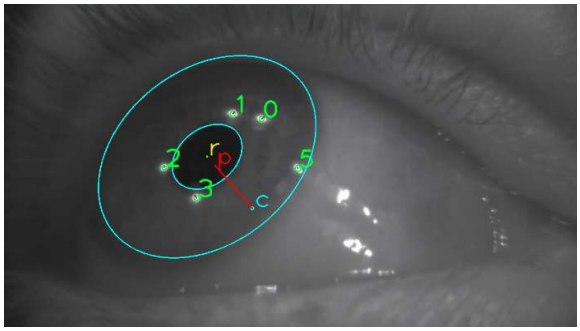

Fig. 7. Optical axis of the eye estimation result.

Applying this quaternion to find the direction vector $\boldsymbol{p}_{norm}$ from the point of refraction to the pupil center point, we obtain:

$$\boldsymbol{p}_{norm} = Q\frac{(\boldsymbol{c}-\boldsymbol{r})}{\|\boldsymbol{c}-\boldsymbol{r}\|}Q^{-1}. \quad (11)$$

Taking into account the calculations above (7) is transformed into quadratic equation:

$$k_p^2 - 2k_p(\boldsymbol{c}-\boldsymbol{r}) \bullet \boldsymbol{p}_{norm} + \|\boldsymbol{c}-\boldsymbol{r}\|^2 - K^2 = 0 \quad (12)$$

where k_p – the distance from the refraction point to the pupil center. Quadratic equation describes the intersection of a line ($\boldsymbol{r}$-$\boldsymbol{p}$) and the sphere with radius K and center in point $\boldsymbol{c}$. The distance from the point of refraction to pupil center corresponds to a smaller root.

Vector of the pupil center is determined with the equation:

$$\boldsymbol{p} = \boldsymbol{p}_{norm}k_p + \boldsymbol{r} \quad (13)$$

Result of optical axis estimation is shown in Fig. 7.

D. Point of gaze estimation

The direction angles of the gaze optical axis relative to the eye camera are calculated from the equation:

$$\frac{\boldsymbol{p}-\boldsymbol{c}}{\|\boldsymbol{p}-\boldsymbol{c}\|} = \begin{bmatrix} \cos\varphi\sin\theta \\ \sin\varphi \\ -\cos\varphi\cos\theta \end{bmatrix} \quad (14)$$

To determine the point of gaze in the reference frame of the scene camera we use polynomial of the third degree for each coordinate:

$$\begin{cases} x = \sum_{i=0}^{i=3}\sum_{j=0}^{i} a_{lij}\theta_l^j\varphi_l^{i-j} + \sum_{i=0}^{i=3}\sum_{j=0}^{i} a_{rij}\theta_r^j\varphi_r^{i-j} \\ y = \sum_{i=0}^{i=3}\sum_{j=0}^{i} b_{lij}\theta_l^j\varphi_l^{i-j} + \sum_{i=0}^{i=3}\sum_{j=0}^{i} b_{rij}\theta_r^j\varphi_r^{i-j} \end{cases} \quad (15)$$

where (x, y) – point of gaze in the reference frame of the scene camera, (θ_l, φ_l) – direction angles of the optical axis of the left eye gaze relative to the left eye camera, (θ_r, φ_r) – direction angles of the optical axis of right eye gaze relative to the right eye camera.

The coefficients of the system take account of the individual characteristics of the user (angles of between the vision and the optical axis of the eye) and properties of the head mounted tracker (transformation from the eye cameras to

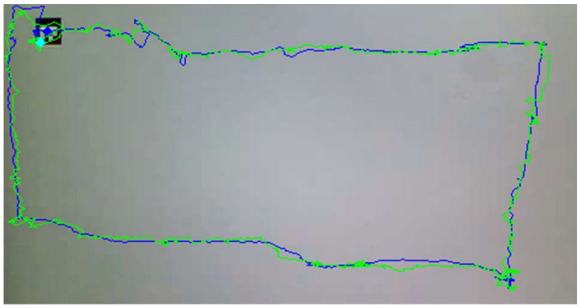

Fig. 8. Calibration result.

the scene camera), these coefficients are determined in the process of calibration.

E. Calibration

The coefficients of polynomial (15) are calculated using regression algorithm.

User calibration undergoes in two stages, first one is the calibration and second one is the verification. Calibration is carried out using the monitor or tablet with the moving AruCo mark. The calibration mark is found using an algorithm from a widespread library AruCo [6]. Calibration result is presented in Fig. 8 (the blue line is detected movement of the calibration mark center; the green line is the point of gaze position estimation).

The investigation of the calibration procedure revealed that a moving mark gives approximately 3 times more accurate results compared to stationary mark, and better retains user attention.

III. SUMMARY AND CONCLUSION

A set of algorithms has been developed that allows to estimate real-time point of gaze for a head mounted tracker.

TABLE I. CALIBRATION RESULTS EXAMPLES

Accuracy, deg	**Precision x, deg**	**Precision y, deg**	**Frames accepted, %**
0.386	0.375	0.263	82.48
0.475	0.447	0.436	85.07
0.43	0.382	0.33	94.65

The root-mean-square (RMS) accuracy exceeds most of the values obtained for wearable devices [7], the RMS deviation is approximately 2 times higher than the similar value for commercial devices, but our calibration time is also as half as long [3]. In the process of algorithm optimization it was revealed that the percentage of recognized frames under good calibration conditions does not significantly affect the result. During operation the frame rate on a personal computer is about 150 Hz in a single thread.

REFERENCES

[1] Sadeghi, S., & Maleki, A., "Recent advances in hybrid brain-computer interface systems: a technological and quantitative review". *Basic and Clinical Neuroscience*, 9(5), 373-388, 2018. http://dx.doi.org/10.32598/bcn.9.5.373

[2] Guestrin, Elias & Eizenman, Moshe. (2006). "General theory of remote gaze estimation using the pupil center and corneal reflections". *Biomedical Engineering, IEEE Transactions* on. 53. 1124 - 1133. 10.1109/TBME.2005.863952.

[3] Matteo Cognolato, Manfredo Atzori1 and Henning Muller, "Head-mounted eye gaze tracking devices: an overview of modern devices and recent advances".*The Journal of Rehabilitation and Assistive Technologies Engineering,* Volume 5: 1–13, 2018. https://doi.org/10.1177/2055668318773991

[4] A.T. Duchowski, "Eye tracking methodology: theory and practice", Springer Verlag, Berlin, 2003.

[5] Suzuki, S. and Abe, K., "Topological structural analysis of digitized binary images by border following". *Computer Vision, Graphics, and Image Processing* 30 1, pp 32-46, 1985.

[6] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion". *Pattern Recogn. 47*, 6, 2280-2292, 2014. DOI=10.1016/j.patcog.2014.01.005.

[7] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," in *IEEE Access*, vol. 5, pp. 16495-16519, 2017. doi: 10.1109/ACCESS.2017.2735633.