3D Gaze Characteristics in Mixed-Reality Environment

Kenta Kato and Oky Dicky Ardiansyah Prima
Graduate School of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: g236s002@s.iwate-pu.ac.jp, prima@iwate-pu.ac.jp

***Abstract*— Recently, Head-Mounted Displays (HMDs) have become popular, making it easier to experience Mixed Reality (MR) environments that fuse real and virtual space. With the ability of MR to locate virtual Three-Dimensional (3D) objects in the real space, our 3D perception may also change as this type of 3D experience increases. While there have been studies that measure 3D gaze in either virtual or real space, no studies have discussed how the MR affects 3D gaze. In this study, we developed a See-Through Head-Mounted Display (ST-HMD) to analyze the effect of MR environment on 3D gaze measurement. We conducted experiments in two different physical environments: a room with and without depth cues. Our results showed that there was no significant difference in the measured 3D gaze between rooms with and without depth cues. Experiments of tracking the gaze of a visual target moving from back to front showed that the scanpath of the 3D gaze followed the trajectory of the target's movement.**

Keywords-3D gaze estimation; See-Through Head-Mounted Display; depth perception.

I. INTRODUCTION

The depth cue of binocular disparity was introduced by Wheatstone (1838) for the first time [1]. Since then, the first Three-Dimensional (3D) movie was released in 1922 where anaglyph glasses were used [2]. Later, 3D Televisions were introduced around 2010, but due to low demand, they were discontinued around 2016. 2016 is a symbolic year in which many Virtual Reality Head-Mounted Displays (VR-HMDs) were launched. The HoloLens, a See-Through Head-Mounted Display (ST-HMD), enables users to experience Mixed Reality (MR), new environment and visual representation generated by the fusion of real and virtual world.

Depth perception is classified into binocular cues, which receive three-dimensional sensory information from both eyes, and monocular cues, which are represented in only two dimensions and observed with one eye. Binocular cues include retinal disparity, which exploits parallax and vergence. In contrast, monocular cues include relative size, texture gradient, occlusion, linear perspective, contrast differences, and motion parallax. Displays with head-tracking capability can generate motion parallax to improve the sense of depth.

ST-HMDs are expected to be effectively used in the medical field to allow multiple observers to view 3D medical images in MR [3][4]. These devices are prone to discomfort and fatigue as the viewing time increases [5]. This is due to maintaining the focus of the eye while continuing to gaze at a moving object with vergence eye movements [6]. Analyzing the 3D gaze characteristics of viewers in MR is the key to solve these problems.

3D gaze can be measured using a binocular eye tracker. The representation of 3D gaze can be roughly divided into two categories: the direction of gaze from the eye and the 3D position of the eye-gaze point [7][8]. Recent advances in deep learning technology have made it possible to estimate the gaze direction directly from face images [9], but to estimate the 3D position of the eye-gaze point, a binocular eye tracker with 3D gaze calibration is still needed [10]-[12].

Analyzing 3D gaze in MR environments requires evaluating the relationship between the perceived position of a 3D object and the 3D eye-gaze point of that object. Using the Microsoft HoloLens with a binocular eye tracking, Öney et al. (2020) measured 3D gaze depth during a visual search task of 3D objects in MR environment placed within 1.25m to 5m of the subject. However, their experiment suffered from a significantly large measurement error of more than a meter when the focused object was only 3.5m away from the viewer [13]. To achieve high accuracies in 3D gaze measurements, Kapp et al. (2021) filtered the resulting fixations manually, yielding an approximately 5cm of errors in average within 4.0m measurement distance [14].

In this study, we measure 3D gaze in MR and analyze the influence of the surrounding physical environment on 3D perception cues. Unlike previous studies, we will analyze the characteristics of 3D gaze scan paths of moving targets as well as stationary targets. In addition, we do not perform manual filtering of gaze data to reveal the 3D gaze characteristics in MR environment. The 3D eye tracker with ST-HMD for MR is developed in this study.

The rest of this paper is organized as follows. Section II describes the 3D eye tracker based on ST-HMD developed for this study. Section III explains our experiments in environments with and without depth cues. Section IV describes the results. Finally, Section V presents our conclusion.

II. 3D EYE TRACKER BASED ON ST-HMD

We adopted Moverio (BT-30E, EPSON) as an ST-HMD and a three simultaneous USB camera module (KYT-U030-3NF, KAYETON) to capture images of both eyes and the viewer' scene at the same time. These cameras run at 60Hz. Figure 1 shows our 3D eye tracker. We used Pupil Capture (Pupil Labs), an open-source eye tracking platform, to localize the pupils of both eyes [15]. The software

 ISBN: 978-1-61208-872-3

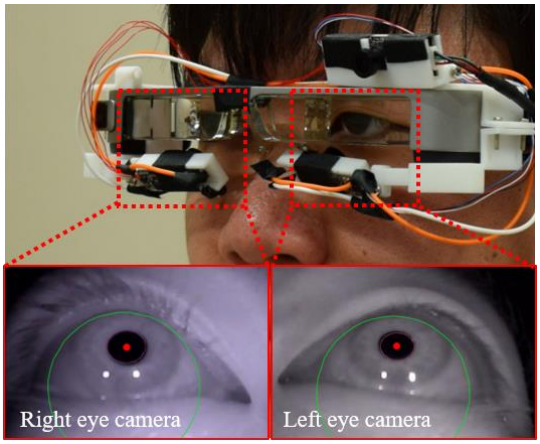


Figure 1. Our 3D eye tracker used in this study.

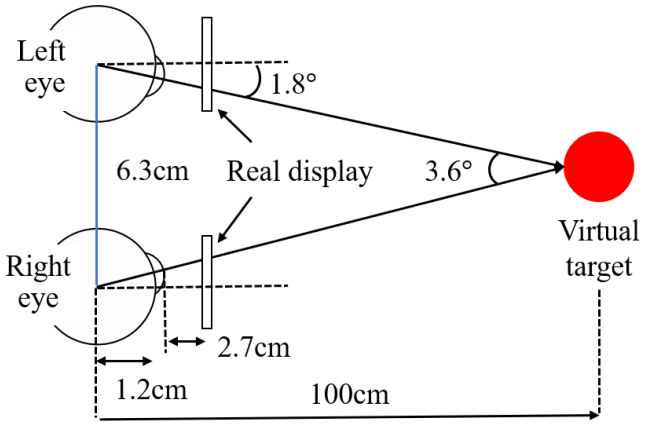
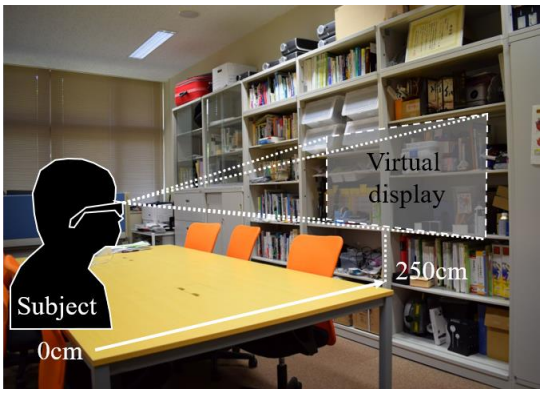
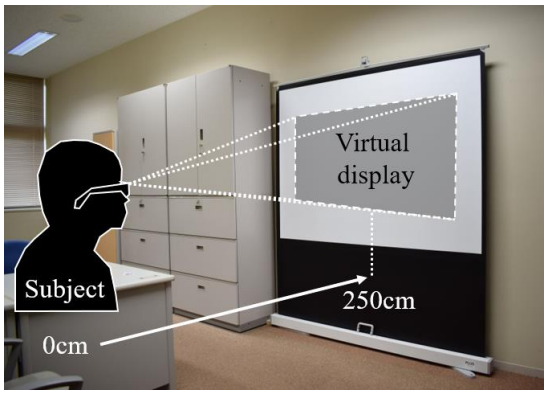


Figure 2. The relationship between both eyes and the virtual



(a) A room with depth cues.



(b) A room without depth cues.

Figure 3. The physical environment for the experiments in this study.

automatically calculates 3D eye vector running from the estimated eyeball center.

A. *Virtual Environments*

The virtual targets used for measurement were created by generating binocular disparity to induce vergence. To determine the disparity, the angle between 3D eye vectors is calculated when gazing at a target placed at a certain distance. The ST-HMD used in this study was designed to allow the user to perceive a virtual screen equivalent to 40-Inch displayed at 250cm. The interpupillary distance was fixed at 6.3cm, which is the average interpupillary distance for Japanese [16]. Figure 2 shows the relationship between both eyes and the virtual target placed 100cm in front of the viewer. We defined the radius of the adult eye to be 1.2cm, as there were no significant differences by gender, race, or age group [17].

B. *3D Gaze Estimation*

To calculate the 3D gaze, a polynomial equation is used to determine the relationship between the gaze direction of both eyes and the 3D position of the gazing target when gazing with vergence eye movement. Let (θ_l, θ_r) represent pitch angles and (φ_l, φ_r) represent yaw angles of eye-sight lines coming from the eye-ball center to the pupil center of the left and right eyes, respectively, the coordinate value of the 3D gaze (G_x, G_y, G_z) is calculated by,

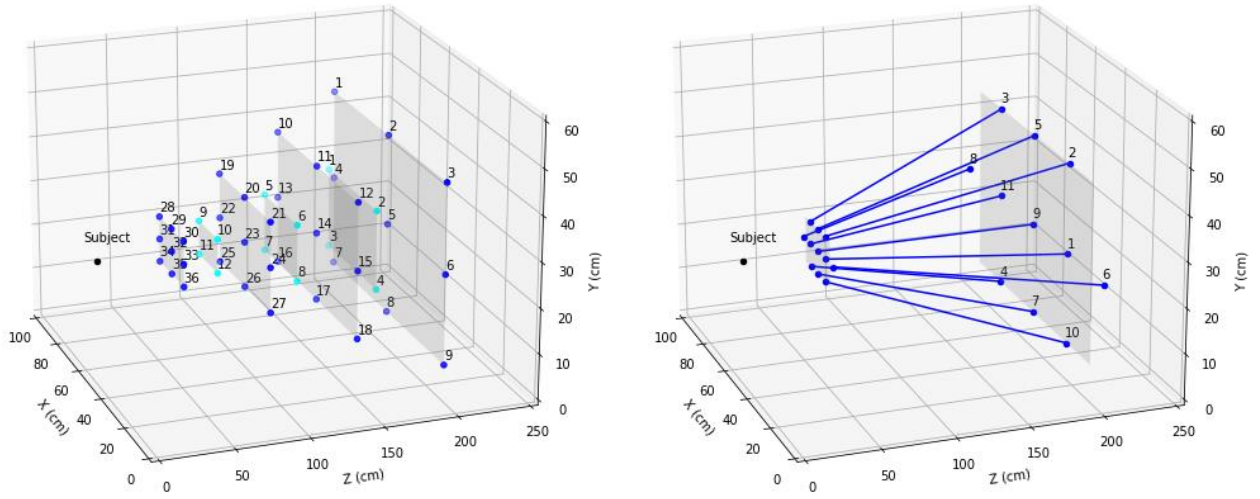
$$G_x = a_1\theta_r^2 + a_2\varphi_r^2 + a_3\theta_l^2 + a_4\varphi_l^2 + a_5\theta_r\varphi_r + a_6\theta_l\varphi_l + a_7\theta_r\theta_l + a_8\theta_r\varphi_l + a_9\theta_l\varphi_r + a_{10}\varphi_r\varphi_l + a_{11}\theta_r + a_{12}\varphi_r + a_{13}\theta_l + a_{14}\varphi_l + a_{15} \quad (1)$$

$$G_y = b_1\theta_r^2 + b_2\varphi_r^2 + b_3\theta_l^2 + b_4\varphi_l^2 + b_5\theta_r\varphi_r + b_6\theta_l\varphi_l + b_7\theta_r\theta_l + b_8\theta_r\varphi_l + b_9\theta_l\varphi_r + b_{10}\varphi_r\varphi_l + b_{11}\theta_r + b_{12}\varphi_r + b_{13}\theta_l + b_{14}\varphi_l + b_{15} \quad (2)$$

$$G_z = c_1\theta_r^2 + c_2\varphi_r^2 + c_3\theta_l^2 + c_4\varphi_l^2 + c_5\theta_r\varphi_r + c_6\theta_l\varphi_l + c_7\theta_r\theta_l + c_8\theta_r\varphi_l + c_9\theta_l\varphi_r + c_{10}\varphi_r\varphi_l + c_{11}\theta_r + c_{12}\varphi_r + c_{13}\theta_l + c_{14}\varphi_l + c_{15}. \quad (3)$$

Coefficients $(a_1 \sim a_{15}, b_1 \sim b_{15}, c_1 \sim c_{15})$ are calculated by the least-squares method based on the correspondence between the gaze direction $(\theta_l, \theta_r, \varphi_l, \varphi_r)$ of each eye and the 3D position of the gazing target obtained by 3D eye calibration.

 ISBN: 978-1-61208-872-3



(a) Visual targets for calibrations and validations

(b) The back-to-front moving visual targets

Figure 4. Experimental environment and target placement

III. EXPERIMENT

In this study, we conducted gazing experiments using the following procedure in viewing environments with and without depth cues.

Step 1. The subject is equipped with the ST-HMD and is seated at 250cm from the front wall.

Step 2. 3D eye calibration is performed using 36 visual targets placed in the virtual space. Each target is displayed sequentially in 3s.

Step 3. Another 12 visual targets are displayed to validate the accuracy of the 3D Gaze calibration.

Step 4. The subject is asked to track a visual target that is approaching from 200cm to 50cm in the virtual space. It takes 5 seconds for the target to travel. This tracking is repeated 11 times with targets coming from different directions.

The above procedure is conducted in two different physical environments: a room with and without depth cues. Figure 3 shows the physical environment for the experiments in this study. Here, to place visual targets for 3D gaze calibration, we set up four planes at 50cm intervals in the virtual space ranging from 50 to 200cm from the subject. In each plane, nine targets were placed. For the validation, three surfaces were set up at 50cm intervals at 75 to 175cm from the subject, and four targets were placed on each surface. The eleven back-to-front moving visual targets are used to examine the change in depth of the 3D gaze.

IV. RESULTS

Four subjects (male, mean age 23.3) participated in the experiment. They were tested for visual acuity using a Landolt ring to confirm that their vision achieved 1.0 or better. They were also asked to fill out a questionnaire to confirm that they had no health concerns.

A. Measurement Accuracy

Accuracies for the 2D gaze measurement Acc_{2D} is measured by

$$Acc_{2D} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\pi}{180}atan\left(\frac{\sqrt{(T_{xi}-G_{xi})^2+(T_{yi}-G_{yi})^2}}{T_{zi}}\right)}, \quad (4)$$

whereas and the 3D gaze measurements Acc_{3D} is measured by

$$Acc_{3D} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_{xi}-G_{xi})^2+(T_{yi}-G_{yi})^2+(T_{zi}-G_{zi})^2}. \quad (5)$$

Here, n is the number of targets used for the measurement, T_{xi}, T_{yi}, T_{zi} and G_{xi}, G_{yi}, G_{zi} are the coordinates of the i-th target and the associated eye-gaze points.

Tables I and II show the Acc_{2D} and Acc_{3D} of the resulted calibrations and validations conducted in two physical environments. The accuracy of the 2D gaze calibration is less than 3 degrees, regardless of the experimental environment. This value corresponds to an error of less than 5cm at an area 1m away from the eye. For validations, we observed that the accuracy decreased in the environment with depth cues. On the other hand, there was no significant difference in accuracy for both calibration and validation of the 3D gaze. These accuracies were analyzed with a 2 × 2, depth cues × visual targets, two-way Analysis of Variance (ANOVA). Effects were not found in either the depth cues (F (1, 12) = 0.192, p = .669) or the visual targets (F (1, 12) = 0.192, p = .669; F (1, 12) = 3.497, p = .086). We conducted post-experimental

TABLE I. ACCURACIES FOR THE 2D GAZE MEASUREMENT (°)

Subject	With depth cues		Without depth cues	
	Calibration	*Validation*	*Calibration*	*Validation*
1	2.96	3.06	1.98	2.01
2	3.20	3.34	3.28	2.22
3	2.12	2.10	2.26	3.89
4	3.02	8.75	2.13	1.91
Mean	2.83	4.31	2.41	2.51
Std. Dev.	0.478	3.006	0.589	0.930

TABLE II. ACCURACIES FOR THE 3D GAZE MEASUREMENT (cm)

Subject	With depth cues		Without depth cues	
	Calibration	*Validation*	*Calibration*	*Validation*
1	21.38	24.74	11.48	21.73
2	20.03	23.68	26.65	26.77
3	7.40	11.07	12.85	20.74
4	22.55	34.79	9.78	22.92
Mean	17.84	23.57	15.19	23.04
Std. Dev.	7.037	9.721	7.744	2.643

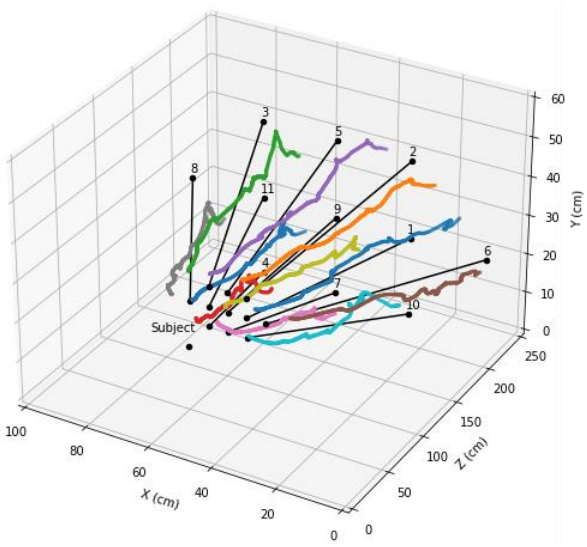


Figure 5. The path along which each target moves and its gazing position.

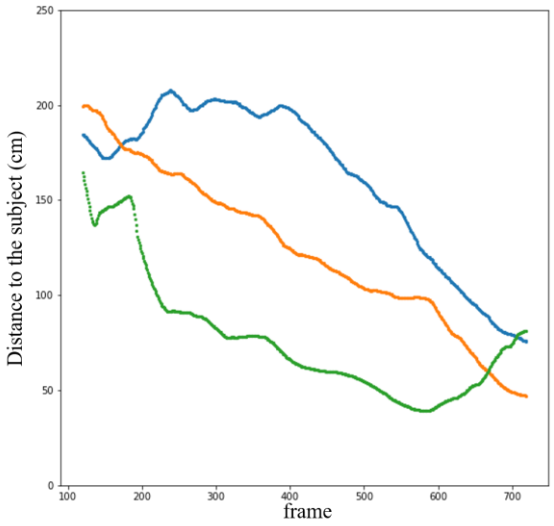


Figure 6. Change in gazing distance with movement of the target.

interviews to find out how the users perceived the 3D visual targets. The post-experimental interview revealed that the fourth subject was unable to converge on some of the validation targets. However, since the accuracy of the calibration was sufficiently high that the data from the subsequent experiments could be used for analysis.

B. *Difference between perceived distance and measured distance*

After the gaze calibration and validation completed, we performed experiments to track the back-to-front moving visual targets with the gaze. Figure 5 shows a typical of subject's scanpath of 3D gaze while tracking the back-to-front targets. The black line indicates the line connecting the initial position of the target in the front-back direction and the position of the subject's eyes. As can be seen, the scan path of the 3D gaze follows the trajectory of the target motion.

Figure 6 shows the profile of typical 3D scanpaths of #7 back-to-front moving visual target. The orange line shows a downward trend, indicating that the distance measured by the 3D gaze gets shorter as the back-to-front visual target gets closer. This is an ideal result, but not all subjects were able to produce these characteristics of eye movements. Subjects observed that their 3D gaze became unstable when the targets started to move, as shown by the blue and green lines. In the post-experimental interviews, we found that this was due to difficulties in tracking the visual target with proper vergence when it started moving.

V. CONCLUSSION

In this study, we measured and analyzed 3D gaze in MR environments. Our experiments showed that there was no significant difference in the measured 3D gaze between rooms with and without depth cues. This result is consistent with that indicated by Öney et al. (2020), but our setup achieves an average accuracy of less than 25cm, which is about four times better than their measurements [13].

Experiments with tracking the gaze of a visual target moving from back to front showed that the 3D gaze scanpath followed the trajectory of the target's movement. However, in some situations, some users found it difficult to track the visual target with proper vergence when it started moving. In MR environments, it is common for viewers to view 3D contents in motion, thus further analysis of the relationship among viewer movements and 3D gaze characteristics is important.

With the increasing use of VR devices, MR will become more accessible, and there will be more research on the characteristics of vergence eye movements in 3D experiences. In our next experiments, we plan to measure 3D gaze while

 ISBN: 978-1-61208-872-3

the subject is moving to investigate the effect of motion parallax.

REFERENCES

[1] K. R. Brooks, “Depth Perception and the History of Three-Dimensional Art: Who Produced the First Stereoscopic Images?” I-Perception, vol. 8, no. 1, pp. 1-22, 2017, doi:10.1177/2041669516680114.

[2] S. Kunić and Z. Šego, “3D Television,” Proceedings ELMAR-2011, pp. 127-131, 2011, ISSN: 1334-2630, ISBN: 978-1-61284-949-2.

[3] T. Sielhorst, C. Bichlmeier, S. M. Heining, and N. Navab, “Depth Perception – A Major Issue in Medical AR: Evaluation Study by Twenty Surgeons,” Medical image computing and computer-assisted intervention : International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9, pp. 364-372, 2006, doi:10.1007/115866565_45.

[4] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, "Egocentric depth judgments in optical, see-through augmented reality," in IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 3, pp. 429-442, 2007, doi:10.1109/TVCG.2007.1035.

[5] M. Emoto, Y. Nojiri, and F. Okano, “Changes in Fusional Vergence Limit and its Hysteresis after Viewing Stereoscopic TV,” Displays, 25, pp. 67-76, doi:10.1016/j.displa.2004.07.001.

[6] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, “Visual discomfort with stereo displays: Effects of viewing distance and direction of vergence-accommodation conflict,” Proceedings of SPIE - the International Society for Optical Engineering, vol.7863, pp. 1-9, 2011, doi:10.1117/12.872347.

[7] A. Kacete, R. Séguier, M. Collobert, and J. Royan, “Head Pose Free 3D Gaze Estimation Using RGB-D Camera,” ICGIP, SPIE, vol. 10255, 2016, Tokyo, Japan, hal-01393594.

[8] L. Świrski, and N. A. Dodgson, “A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting,” In Proceedings of ECEM 2013.

[9] X. Zhou et al., "Learning a 3D Gaze Estimator With Adaptive Weighted Strategy," in IEEE Access, vol. 8, pp. 82142-82152, 2020, doi: 10.1109/ACCESS.2020.2990685.

[10] S. Weber, R. Schubert, S. Vogt, B. M. Velichkovsky, and S. Pannasch, “Gaze3DFix: Detecting 3D fixations with an ellipsoidal bounding volume,” Behavior Research Methods, vol. 50, no. 5, pp. 2004-2015, 2018, doi:10.3758/s13428-017-0969-4.

[11] T. Pfeiffer, M. E. Latoschik, and I. Wachsmuth, “Evaluation of Binocular Eye Trackers and Algorithms for 3D Gaze Interaction in Virtual Reality Environments,” JVRB - Journal of Virtual Reality and Broadcasting, vol. 5, no. 16, 2008, doi: 10.20385/1860-2037/5.2008.16.

[12] E. G. Mlot, H. Bahmani, S. Wahl, and E. Kasneci, “3D Gaze Estimation using Eye Vergence,” BIOSTEC, pp. 125-131, 2016, doi:10.5220/0005821201250131.

[13] S. Z. Öney et al., “Evaluation of Gaze Depth Estimation from Eye Tracking in Augmented Reality,” ACM Symposium on Eye Tracking Research and Applications, pp. 1-5, 2020, doi:10.1145/3379156.3391835.

[14] S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn, “ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays,” Sensors, vol. 21, no. 6, 2234, 2021, doi: 10.3390/s21062234

[15] M. Kassner, W. Patera, and A. Bulling. “Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction,” Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 1151-1160, 2014, doi:10.1145/2638728.2641695.

[16] M. Kouchi, and M. Mochimaru, “2008: Anthropometric Database of Japanese Head 2001,” National Institute of Advanced Industrial Science and Technology, H16PRO-212.

[17] I. Bekerman, P. Gottlieb, and M. Vaiman, “Variations in Eyeball Diameters of the Healthy Adults,” Journal of ophthalmology, vol. 2014, 503645, doi:10.1155/2014/503645.

 ISBN: 978-1-61208-872-3