A Geometric Model for Measuring Depth Perception in Immersive Virtual Environment

Sunu Wibirama and Kazuhiko Hamamoto
Graduate School of Science and Technology
Tokai University, Tokyo, Japan
sunu_wibirama@yahoo.com

ABSTRACT
Study of three-dimensional (3D) gaze and depth perception in immersive virtual environment (IVE) systems is an important research area. Several approaches have been proposed, but the usage of gaze-screen information for depth perception measurement in IVE has never been reported. We suggest a new approach using geometric model to measure depth perception in IVE. We extend information from binocular gaze-screen mapping and obtain the depth information by line-crossing method. We also investigate error of depth perception by eye camera alignment and front/behind screen condition. Our simulation results show that the 2D gaze tracking estimation error of the proposed method is lower than 1° and yields depth measurement accuracy more than 84% on ±0.6 to ±1 meter depth of front/behind screen condition.

Author Keywords
Geometric method; gaze tracking; depth perception

ACM Classification Keywords
H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – Evaluation/methodology

INTRODUCTION
Recently, virtual environment (VE) has been widely used to simulate real or imagined scenario, such as design and training processes in industrial and collaborative tasks [1-3]. To make the users feel more immersed, interested, and comfortable within these VE systems, optimization of VE content is needed. In VE optimization, studying user’s perceived depth perception measurement is very important to obtain and correct any discrepancy between the measured value and that specified in VE content [4].

Several approaches have been proposed to measure depth perception in VE systems. Poyade et al. [5] investigated the effect size cues of virtual 3D object to depth perception. In this research, the user’s depth perception was measured by asking the user to interactively locate a virtual three-dimensional (3D) object at the same distance with static reference. Bigoin et al. [6] used simulator sickness (SS) method to study depth cues in VE. In this research, SS questionnaire was collected and analyzed after the user viewed virtual world displayed on an immersive semicylindrical projection system. Another method, such as verbal-assisted, time-imagined, and blindfolded walking were also used to investigate several cues that affect depth perception in VE systems [7-9]. Nevertheless, none of these focused on relating gaze direction and screen information with depth perception in IVE.

On the other side, the usage of two-dimensional (2D) gaze tracking as interactive interface and visual search tools in IVE systems has been widely known [10-13]. However, these 2D gaze tracking method merely estimated 2D gaze position on the screen without further analysis for depth perception measurement. In our previous work, we proposed a 3D eye movements tracking by analyzing data captured from dual-camera [14]. However, this method can only be used in a close-view head-mounted device and not suitable for unobstructed-view environment, such as large stereoscopic display or IVE.

To overcome this problem, we suggest a geometric model by extending 2D gaze tracking information. 3D coordinate of pupil center is obtained by transforming and rotating pupil center with bottom left of screen as origin of world coordinate reference. A depth fixation point is then obtained by crossing 3D line of sight from both eyes. Experimental validation was performed to analyze depth perception measurement error by eye camera alignment and front/behind screen condition.

METHODS
Fundamental Procedures
Figure 1 shows fundamental procedures used in this research. We use binocular vision to measure depth of virtual 3D object from screen as shown in Figure 1(a). The eye is assumed as a perfect sphere with known radius performing pure rotations around the center of the eye as shown in Figure 1(b) [15]. There is also assumed that the subject’s head is fixed so that no head movement is produced during depth perception measurement.
Figure 1: Fundamental Procedures

The eye coordinate system is based on modified right-handed Fick-order [16]: y_e-axis (horizontal rotation), x_e -axis (vertical rotation), and z_e -axis (torsional rotation). The optical axis of the eye is initially assumed coinciding with the z_e-axis. P_1 and P_3 are 3D coordinate of pupil center of left and right eye related to each eye coordinate system, respectively. The radius of the eyeball is 13.5 mm as used in conventional gaze tracking research [17-18]. The distance between two eyeballs is assumed 6.5 cm as used in [19].

The origin of the world coordinate is located at the bottom left of the screen. While seeing the virtual 3D object at front or behind the screen, the eye gaze will be mapped on the screen as 3D point P_2 and P_4 for left and right eye, respectively. P_2 and P_4 are related to world coordinate system.

The eye camera is a pinhole camera without any lens distortion. It is initially located at the front of the eyeball with 6 cm distance from the center of pupil to center of camera coordinate. The optical axis of the camera initially coincides with optical axis of the eye. The focal length of the camera is 0.015 pixel with resolution 1280 x 1024 pixels.

In this research, we use HoloStage® Christie Digital System configuration as used in Tokai University, Takanawa campus, Japan [20]. The screen size is 5.4 m x 3 m. The subject’s eye height is assumed 1.7 m as in [21]. The subject is positioned at 3 m distance from the screen.

2D Gaze Mapping

We use center of mass algorithm to detect 2D pupil center coordinate [14]. The 2D coordinate of pupil center detected on eye camera is used to map 2D gaze on the screen. For n to n points mapping, 2D point G_1(u, v) on the eye camera image plane can be transformed to 2D point G_2(u', v') on the screen using homography equation [22] as follows:

\[
\begin{bmatrix}
    u' \\
    v' \\
    1
\end{bmatrix}
= \begin{bmatrix}
    h_{11} & h_{12} & h_{13} \\
    h_{21} & h_{22} & h_{23} \\
    h_{31} & h_{32} & h_{33}
\end{bmatrix}
\begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix}
\]

(1)

\[
\begin{bmatrix}
    u' \\
    v' \\
    1
\end{bmatrix}
= \begin{bmatrix}
    1 & 0 & 0 & -u & -u' & 0 \\
    0 & 1 & 0 & -v & -v' & 0 \\
    0 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
    h_{11} \\
    h_{21} \\
    h_{31}
\end{bmatrix}
= 0
\]

(2)

\[
A h = 0
\]

(3)

A is 2n-by-9 matrix containing both 2D pupil coordinate and 2D gaze coordinate on the screen, h is 3-by-3 homography matrix, and α is scaling factor. We are interested in non-trivial solution of Equation 1 (h≠0), and therefore consider det(A)=0. Singular Value Decomposition (SVD) has been performed to solve h. In this research, a 4-by-3 calibration grid is used which produces 12-by-12 points mapping.

3D Pupil Position Computation

We extend information from 2D gaze mapping. A 3D position of initial left eye gaze on the screen is defined by P_{3i} (x_i, y_i, 0) with x_i = u_i and y_i = v_i. A 3D position of current left eye gaze on the screen is defined by P_{3c} (x_c, y_c, 0) with x_c = u'_c and y_c = v'_c. Given the distance of screen to subject z_sc and radius of thee eyeball r, we can compute horizontal (θ) and vertical (α) rotational angle of the left eyeball as follows:

\[
\theta = \arctan \left(\frac{x_c - x_i}{z_{sc} + r}\right)
\]

(4)

\[
\alpha = \arctan \left(\frac{y_c - y_i}{z_{sc} + r}\right)
\]

(5)

Similar computation is implemented to the right eyeball.

To compute 3D pupil position, translation and rotational matrices of eyeball related to world coordinate have to be computed. Note that initial left pupil 3D coordinate is P_{1i} (x_i, y_i, r). Translation and rotational matrices of the left
eyeball related to world coordinate are defined in homogeneous format as follows:

\[
T_1 = \begin{bmatrix}
1 & 0 & 0 & -x_i \\
0 & 1 & 0 & -y_i \\
0 & 0 & 1 & -z_i \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(6)

\[
T_2 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & -r \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(7)

\[
T_3 = \begin{bmatrix}
1 & 0 & 0 & x_i \\
0 & 1 & 0 & y_i \\
0 & 0 & 1 & z_i \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(8)

\[
T_4 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & r \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(9)

\[
R_1 = \begin{bmatrix}
\cos(-\alpha) & -\sin(-\alpha) & 0 \\
\sin(-\alpha) & \cos(-\alpha) & 0 \\
0 & 0 & 1
\end{bmatrix}
\]  

(10)

\[
R_2 = \begin{bmatrix}
\cos(\theta) & 0 & -\sin(\theta) \\
0 & 1 & 0 \\
\sin(\theta) & 0 & \cos(\theta)
\end{bmatrix}
\]  

(11)

Using translational and rotational matrices, the current 3D left pupil coordinate \(P_{lc}(x_c, y_c, z_c)\) can computed in homogeneous format as follows:

\[
P_{lc} = T_4 * T_3 * R_2 * R_1 * T_1 * \begin{bmatrix}
x_i \\
y_i \\
z_i
\end{bmatrix}
\]  

(12)

where \(P_{lc} = [x_c, y_c, z_c, 1]^T\) is homogeneous formula of \(P_{lc}\). Similar computation is also implemented to compute current position of 3D coordinate of right pupil \(P_{rc}\).

**Depth Perception Measurement**

To compute depth perception, we use line-crossing method [23]. Given two line segments formed by pupil and gaze position \(P_1P_2\) and \(P_3P_4\), joined by shortest line \(P_sP_b\), a fixation point can be find by finding the midpoint \(P_m\) of \(P_sP_b\) as shown in Figure 2. Point \(P_2\) on the line \(P_1P_2\) and point \(P_b\) on the line \(P_3P_4\) are given by equations:

\[
P_2 = P_1 + \mu(P_2 - P_1)
\]

(13)

\[
P_b = P_3 + \eta(P_4 - P_3)
\]

(14)

The shortest line between two crossing lines can be found by minimizing \(|P_b - P_a|\):

\[
P_b - P_a = P_3 - P_1 + \eta(P_4 - P_3) - \mu(P_2 - P_1)
\]

(15)

Since \(P_bP_a\) is perpendicular to line \(P_1P_2\) and \(P_3P_4\), the result of dot product operation between them are zero:

\[
(P_b - P_a).P_2 - P_1 = 0
\]

(16)

\[
(P_b - P_a).P_4 - P_3 = 0
\]

(17)

Substituting Equation 11 to Equations 12 and 13, we obtain:

\[
[P_3 - P_1 + \eta(P_4 - P_3) - \mu(P_2 - P_1)](P_2 - P_1) = 0
\]

(18)

\[
[P_3 - P_1 + \eta(P_4 - P_3) - \mu(P_2 - P_1)](P_4 - P_3) = 0
\]

(19)

Computing Equations 18 and 19 using known \(x, y, z\) values of \(P_1P_2\) and \(P_3P_4\), we obtain \(\mu, \eta, P_a,\) and \(P_b\). The middle point of \(P_aP_b\) can be computed as follows:

\[
P_m = \frac{P_a + P_b}{2}
\]

(20)

Figure 2. 3D intersection of two line segments

**Experimental Validation**

To perform validation, we simulate our geometric model using MATLAB® version R2011b on personal computer with Intel® i7-2600 3.4 GHz processor, 4GB memory, and Windows XP® Service Pack 2 operating system. In this simulation, we assume that 1 MATLAB unit equals to 0.01 meter in real environment. A small red sphere with 0.05m radius is used as virtual 3D object.

The first experiment was conducted to investigate the effect of eye camera installation to depth measurement accuracy. In this experiment, we positioned the eye camera such that the optical axis of the camera performed 0°, 10°, 15°, 20°, and 30° angles related to the optical axis of the eye by maintaining the 6cm distance between the eyeball and the camera center. By using each camera position, we measured
2D gaze mapping on the screen and depth of virtual 3D object from screen. The depth measurement for each angle was conducted at 11 depth positions: 1, 0.8, 0.6, 0.4, 0.2, 0, -0.2, -0.4, -0.6, -0.8, and -1 m from screen. “+” and “-” signs indicate front and behind screen condition, respectively.

The second experiment was conducted to obtain average accuracy of the proposed method in measuring virtual 3D object located at several areas on the screen within various depths from screen parameter. The camera was aligned at 0° degree rotation. We measured depth from screen error using 25 points generated on 2m x 2m area at front and behind screen within ± 0.6m, ± 0.8m, and ± 1m ranges. Each target point was generated such that it was not located at the same horizontal and vertical positions as the calibration grid. The average value of depth from screen error was then computed to confirm depth measurement accuracy.

EXPERIMENTAL RESULTS

This section shows our experimental results. Table 1 shows 2D gaze error by various camera rotations. H and V refer to horizontal and vertical positions, respectively. The average gaze error is about 0.79°. Figure 3 shows comparison of 2D gaze on screen mapping by 0° and 30° camera rotations. The vertical gaze accuracy was found to be affected by camera rotation. For example, if the camera was rotated vertically from 20° to 30°, the average vertical gaze error of left eye increased significantly from 0.98° to 1.32°.

Table 2 shows average depth from screen error by various camera rotations. Figure 4 shows comparison of depth from screen error by various camera rotations. Our proposed method was found to have slightly less depth error when the virtual 3D object was not located near the screen. For example, at 30° camera rotation with virtual 3D object positioned at 0m distance from the screen, the depth error was about 0.1766m. The error decreased when the object was positioned at front or behind the screen.

<table>
<thead>
<tr>
<th>Camera rotation (deg)</th>
<th>Mean of left gaze error (deg)</th>
<th>Mean of right gaze error (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
<td>V</td>
</tr>
<tr>
<td>0</td>
<td>0.65</td>
<td>0.75</td>
</tr>
<tr>
<td>10</td>
<td>0.65</td>
<td>0.76</td>
</tr>
<tr>
<td>15</td>
<td>0.65</td>
<td>0.85</td>
</tr>
<tr>
<td>20</td>
<td>0.65</td>
<td>0.98</td>
</tr>
<tr>
<td>30</td>
<td>0.66</td>
<td>1.32</td>
</tr>
<tr>
<td>Average gaze error (deg)</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. 2D gaze error by various camera rotations

<table>
<thead>
<tr>
<th>Camera rotation (deg)</th>
<th>Mean of depth error (meter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1501 ± 0.1427</td>
</tr>
<tr>
<td>10</td>
<td>0.1512 ± 0.1386</td>
</tr>
<tr>
<td>15</td>
<td>0.1534 ± 0.1230</td>
</tr>
<tr>
<td>20</td>
<td>0.1567 ± 0.1166</td>
</tr>
<tr>
<td>30</td>
<td>0.1678 ± 0.8052</td>
</tr>
</tbody>
</table>

Table 2. Average depth from screen error by various camera rotation

<table>
<thead>
<tr>
<th>Depth from screen (meter)</th>
<th>Mean of error (meter)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>± 0.6</td>
<td>0.0939 ± 0.0306</td>
<td>84.35</td>
</tr>
<tr>
<td>± 0.8</td>
<td>0.1070 ± 0.0481</td>
<td>86.62</td>
</tr>
<tr>
<td>± 1.0</td>
<td>0.1144 ± 0.0730</td>
<td>88.56</td>
</tr>
</tbody>
</table>

Table 3. Average depth from screen error and measurement accuracy by front/behind screen condition

Table 3 shows average depth from screen error and measurement accuracy by front/behind screen condition. Figure 5 shows visualization of absolute depth error by ± 1m depth from screen condition. Measurement accuracy was performed by following equation:

\[
\text{Accuracy} = \frac{|\text{depth from screen} - |\text{mean of error}|}{\text{depth from screen}} \times 100\%
\]

It can be shown that increasing depth from screen distance will lead to higher measurement accuracy since the front screen distance of virtual 3D object is getting closer to the subject.
On the other side, gazing to virtual 3D object located near the border of the screen will increase the depth perception error as shown in Figure 5. An unvarying small statistical depth error can be found when gazing to virtual 3D object located near the center of the screen. For example, when the user gazed to a virtual 3D object located at coordinate (1.7, 2.5, 1), the depth from screen error was about 0.2469m. When the gaze was moved to coordinate (3.2, 2, 1), the error decreased to 0.0239m.

**DISCUSSIONS**

Given a known radius of eyeball and distance between the screen and the subject, the proposed geometric model was found to be adequately describing the 2D gaze position since the gaze error of our geometric method was smaller than previous research work as shown in Table 4. Furthermore, our proposed method yielded best depth measurement accuracy about 88.56% for object located within ± 1m range from the screen. In real situation involving human subject, it was reported that depth in virtual environments was often only 50%-80% of the intended distance [24]. Therefore, our geometric method is considered appropriate to be implemented since it has better depth measurement accuracy compared to human depth perception.

It was observed that camera positioning affected both 2D gaze on screen mapping and depth perception measurement accuracy. Since there is only one camera on each eye, the usage of homography transformation for 2D gaze on screen mapping tends to be ambiguous up to scale. The use of homogeneous formula with only one camera means that any multiplication to homography matrix will have the same effect. Thus, the homography transformation in 2D gaze tracking was prone to error affected by any translation or rotation of the eye camera. In the future, investigating further gaze on screen mapping using homography result as initial condition for more accurate gaze mapping algorithm will be conducted. In actual implementation, it is suggested to keep the optical axis of the camera horizontally aligned to optical axis of the eye. Instead of simply installing the eye tracker camera below the eye as in [11], the usage of infrared reflector may be useful to reduce the depth perception measurement error [25].

**CONCLUSION**

A geometric model for measuring depth perception in IVE has been proposed. The model employs an extension of 2D gaze-screen information to obtain 3D coordinate of pupil center. The depth fixation point is then yielded by using 3D pupil position and 3D gaze mapping position as an input of line-crossing method. Accuracy of the proposed method was observed by computing depth error by camera alignment and front/behind screen condition. The experimental results show that the proposed method provides adequate precision compared to previous research work as the 2D gaze estimation error of the proposed method is 0.79°. Additionally, our proposed method yields best depth measurement accuracy about 88.56% for object located within ± 1m range from the screen. In the future, we would like to implement our geometric model to measure user’s depth perception in IVE system.

**ACKNOWLEDGMENTS**

Our thanks to AUN/Seed-Net, Japan International Cooperation Agency (JICA), and Tokai University for supporting research fund and equipments.

**REFERENCES**


