

# Estimating 3D Point-of-regard and Visualizing Gaze Trajectories under Natural Head Movements

Kentaro Takemura\*

Yuji Kohashi

Tsuyoshi Suenaga

Jun Takamatsu

Tsukasa Ogasawara

Graduate School of Information Science  
Nara Institute of Science and Technology, Japan

## Abstract

The portability of an eye tracking system encourages us to develop a technique for estimating 3D point-of-regard. Unlike conventional methods, which estimate the position in the 2D image coordinates of the mounted camera, such a technique can represent richer gaze information of the human moving in the larger area. In this paper, we propose a method for estimating the 3D point-of-regard and a visualization technique of gaze trajectories under natural head movements for the head-mounted device. We employ visual SLAM technique to estimate head configuration and extract environmental information. Even in cases where the head moves dynamically, the proposed method could obtain 3D point-of-regard. Additionally, gaze trajectories are appropriately overlaid on the scene camera image.

**CR Categories:** H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles

**Keywords:** Eye Tracking, 3D Point-of-regard, Visual SLAM

## 1 Introduction

Recently, there have been active developments in portable eye tracking systems, and *Mobile Eye* (Applied Science Laboratories) and *EMR-9* (NAC) are examples of them. Researchers are interested in the gaze measurement of a person moving dynamically. When head-mounted eye tracking system is utilized, point-of-regard (POR) is generally measured as a point on the image plane. It is therefore difficult to analyze gaze trajectories quantitatively due to head movements. If we employ extra sensor such as a motion capture system or a magnetic sensor to estimate the 3D POR, measuring range would be limited.

Some methods for estimating 3D POR have been proposed to solve these problems. A head-mounted stereo camera pair has been proposed [Sumi et al. 2004], but the system configuration differs from normal eye tracking system substantially. A method using binocular view lines was also proposed [Mitsugami et al. 2003], but 3D POR was estimated as relative position of observer's head. When the observer's position change dynamically, it is difficult to record 3D POR in world coordinates. A novel method using interest points has been proposed [Munn and Pelz 2008], and the estimation of 3D

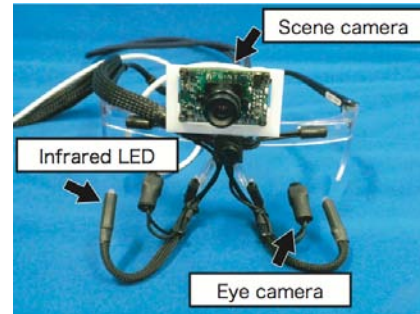


Figure 1: Eye tracking system

POR in world coordinates were achieved. However, when gaze trajectories are analyzed, a visualization method which encourages us to intuitively recognize the observer's 3D POR is needed.

In this paper, we propose a method to estimate 3D POR in real-time and a visualization of gaze trajectories using augmented reality techniques for intuitive understanding. The rest of this paper is structured as follows: Section 2 describes the system configuration, and Section 3 explains the method to estimate 3D POR. In Section 4, we show experimental results and Section 5 concludes this paper.

## 2 Eye Tracking System for Estimating 3D Point-of-regard

The purpose of this research is to estimate the 3D POR of an observer who moves freely. In general, head pose (position and orientation) is estimated by an extra sensor such as a motion capture system or a magnetic sensor, but the measuring range is limited. The measuring range of magnetic sensor is about 5 [m], and many cameras are needed for covering wide area in using motion capture system. To solve this limitations, we propose to estimate the head pose using a scene camera installed in the eye tracking system.

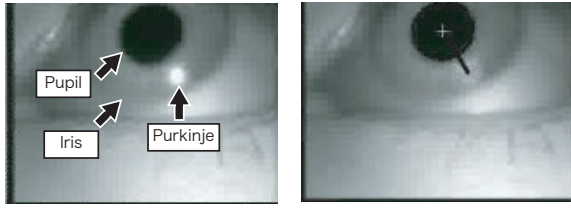
*ViewPoint* (Arrington Research) is used as the head-mounted eye tracking system. Its scene camera is utilized for visibility of POR on image, so it's a low-resolution camera. For this research, the scene camera was changed to FireFly MV (Point Grey Research) as a high resolution camera as shown in Figure 1.

### 2.1 Estimation of Head Pose by Visual SLAM

Our system has two software modules for estimating 3D POR: one for estimating the head pose and one for estimating gaze direction in head coordinates. First, we describe the software for estimating head pose.

Simultaneous localization and mapping (SLAM), which is a technique for estimating robot pose in unknown environments, have been developed in robotics research. A mobile robot with a Laser Range Finder (LRF) could move in unknown environments using SLAM [Smith and Cheeseman 1986]. Moreover, SLAM using a

\*e-mail: kenta-ta@is.naist.jp



**Figure 2:** Pupil and purkinje illuminated by near-infrared light

camera is called Visual SLAM, and the 6 DOF of camera pose can be estimated. In the case of eye tracking, it is necessary to estimate the head pose in an unknown environment, thus Visual SLAM can be utilized. Some techniques for Visual SLAM have been proposed. Among them, MonoSLAM [Davison et al. 2007] and PTAM [Klein and Murray 2007] are representative examples. Particularly, robust estimation was realized by PTAM which utilizes Bundle Adjustment, so we applied it to our eye tracking system. In PTAM, each interest point, extracted using FAST corner detector [Rosten and Drummond 2006], is mapped to 3D coordinates. The camera pose is also estimated this way. In the case of the head-mounted eye tracking system, the scene camera moves along with observer's head movement, thus computing extrinsic parameters of camera is equivalent to estimating the head pose.

## 2.2 Software for Eye Tracking

We now describe the software for eye tracking. OpenEyes [Li et al. 2006] is substituted for the API supplied by ViewPoint for eye tracking. The two software modules could be merged through the use of open source software. It is necessary to get the center of the pupil and the corneal reflection to estimate the gaze direction using OpenEyes. An infrared LED is installed in ViewPoint, so OpenEyes can work on ViewPoint. Figure 2 shows the result using ViewPoint with OpenEyes. Left image is original image captured by eye camera, and right image shows the result of detecting the pupil and the purkinje.

## 3 Estimation and Visualization of 3D Point-of-regard

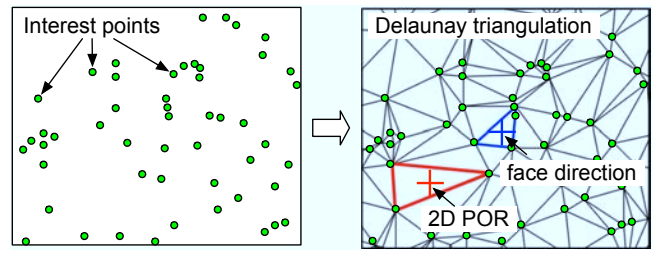
In previous research, 3D POR have been estimated using SLAM [Munn and Pelz 2008]. However, the 3D POR is computed using two key frames to search along an epipolar line. In our proposed method, the 3D POR can be estimated in real-time without two key frames. The following details our approach.

We assume that the 3D POR is located on a plane, which consists of three interest points detected by FAST corner detector, for projecting the POR from 2D to 3D coordinates. The procedures for estimating 3D POR are the following three steps:

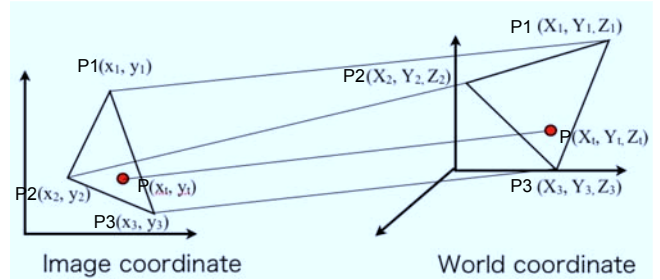
1. Apply Delaunay triangulation to the interest points in Visual SLAM.
2. Determine a triangle which includes the 2D POR estimated by OpenEyes.
3. Compute the 3D POR using the triangle position in world coordinates.

### 3.1 Estimation of 3D Point-of-Regard

First, a 2D POR estimated by OpenEyes have to be associated with interest points extracted by PTAM. Therefore, triangles are built up



**Figure 3:** Delaunay triangulation using interest points



**Figure 4:** Estimation of 3D point-of-regard

based on the interest points, by Delaunay triangulation, as shown in Figure 3. Next, the triangle, which includes the 2D POR, is computed to determine the plane where the 3D POR is. All points on a triangle plane can be presented by forming linear combinations of three points (Eq.1).

$$P = (1 - u - v)P_1 + uP_2 + vP_3 \quad (1)$$

$P, \{P_1, P_2, P_3\}$  are the POR and vertices of triangle in the plane coordinates system respectively, and  $u$  and  $v$  are the ratio of the triangle areas  $(P_1, P_2, P_3)$  and  $(P_1, P_2, P)$ , respectively. When three vertices and 2D POR are given as shown in Eq.2, it is easy to compute the values of  $u, v$  using Eq.3.

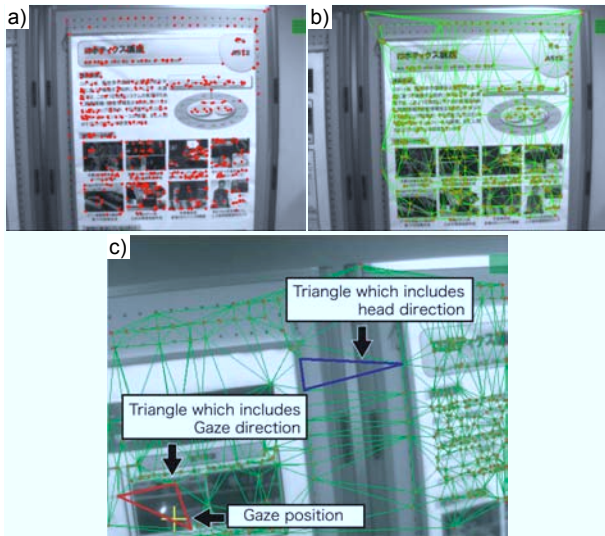
$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = (1 - u - v) \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + u \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + v \begin{bmatrix} x_3 \\ y_3 \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{bmatrix}^{-1} \begin{bmatrix} x_t - x_1 \\ y_t - y_1 \end{bmatrix} \quad (3)$$

$u, v$  are points on the plane coordinate system, thus they can be kept constant for the conversion to 3D coordinates (Eq.4). Then the 3D POR can be estimated as shown in Figure 4. In PTAM, interest points are mapped to 3D coordinates, thus each point has a 3D coordinates.

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} = (1 - u - v) \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} + u \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix} + v \begin{bmatrix} X_3 \\ Y_3 \\ Z_3 \end{bmatrix} \quad (4)$$

If the center of an image is defined as the projected point of head direction, the 3D projected point of head direction can also be estimated. Figure 5 shows the results of our approach, and Figure 5(a) shows the interest points extracted from the scene camera image. Next, Delaunay triangulation is carried out as shown in Figure 5(b). After that, the triangle which includes the 2D POR is determined in Figure 5(c). The 3D POR can be estimated in real-time using these procedures.



**Figure 5:** Delaunay triangulation using interest points. a) Interest points is extracted by FAST feature detector, b) Delaunay triangulation based on interest points, c) Determination of the triangle which includes the 2D POR.

### 3.2 Gaze Trajectories using Augmented Reality Techniques

Some visualization techniques for gaze trajectories have been proposed [Wooding 2002]. Most of visualization techniques are applied to a specific plane such as displays and pictures. However, when the observer moves freely, gaze trajectories cannot be recorded on a specific plane, making it difficult to analyze gaze trajectories. A method for generating a fixation map, which is suitable for time varying situation, have been proposed [Takemura et al. 2003], but it cannot be properly created when the observer moves. Accordingly, we propose a method which overlays the 3D POR on the scene camera image using augmented reality techniques. Camera pose can be estimated in each frames using Visual SLAM, thus the 3D position of interest points and 3D POR can be backprojected on the scene camera image. Therefore, even when the scene camera moves dynamically, gaze trajectories are kept on the scene camera image using its 3D position.

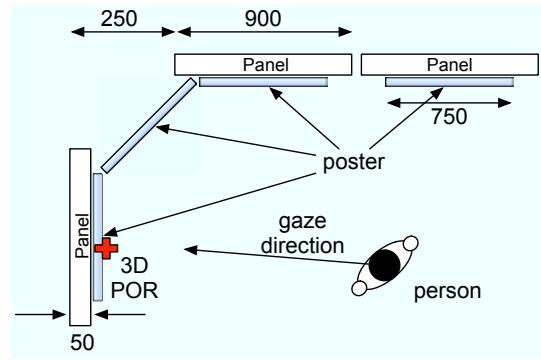
## 4 Experiment of Estimating 3D POR

### 4.1 Experimental Condition

An experiment of estimating 3D POR was conducted for confirming the feasibility of our proposed method. Figure 6 illustrates the experimental environment, and gaze trajectories are measured when the observer looks at some posters. Additionally the observer showed his own POR using a laser pointer to confirm 2D POR, as shown in Figure 7. To calculate the 2D POR in the scene image, calibration was held before measurement. The observer is required to look at nine points on calibration board, thus the observer cannot move during calibration.

### 4.2 Experimental Results

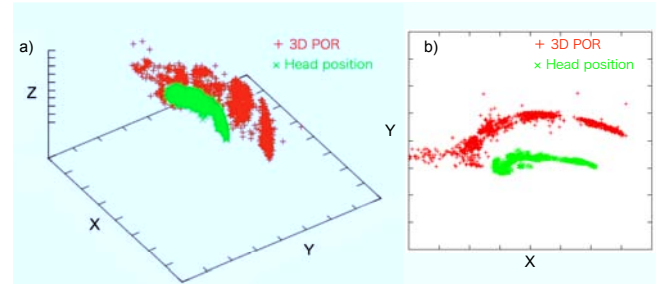
We were able to estimate the 3D PORs. Figure 8 shows the result of estimated 3D PORs and head position. The 3D POR values coincide with the actual location of the four posters in Figure 8,



**Figure 6:** Experimental environment



**Figure 7:** Experimental setup



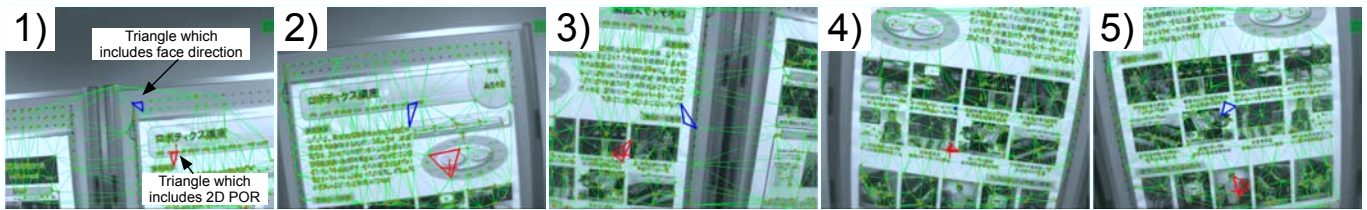
**Figure 8:** Estimated 3D point-of-regard and head position

and we confirmed that the 3D PORs could be estimated appropriately. As shown in Figure 9, triangles, which include 2D POR and the projected points of head direction, could be computed in real-time. Additionally, gaze trajectories are backprojected to the scene image along with the head pose using interest points. When the observer was free to move dynamically, gaze trajectories are kept on the scene camera image as shown in Figure 10. Thus, we could follow the observer's gaze trajectories intuitively. Although we do not have the ground truth of the 3D PORs, the result of overlaid 3D PORs show focus areas such as title and figures on posters, and gaze trajectories are correctly kept on the image even when pose of the scene camera changed. We could confirm the feasibility of our proposed method through an experiment.

### 4.3 Discussion

Our proposed method could estimate 3D PORs in real-time, and gaze trajectories could be overlaid on scene camera image using





**Figure 9:** Snapshots of determining triangles which include the 2D POR by Delaunay triangulation



**Figure 10:** Snapshots of gaze trajectories using augmented reality techniques. When a head-mounted eye tracking system move without restriction, gaze trajectories were drawing on environment appropriately. White arrows show the corresponding points on gaze trajectories between two images.

augmented reality techniques. However, our method has a disadvantage. Our proposed method cannot get the scale of world coordinates because we employed the Visual SLAM which is a bearing only SLAM. It is necessary to solve this problem to evaluate the accuracy of the estimated 3D PORs. If a ground truth marker with known position is used, we can compare the 3D POR to it. However, this would limit the initial position, thus we believe that this problem should be solved by another approach. After solving this problem, we could compare different gaze trajectories of people.

## 5 Conclusion

In this research, we proposed a method to estimate the 3D POR even when the observer moves freely. The head pose is estimated using Visual SLAM, and the 2D POR is associated with interest points to estimate the 3D POR. Our method assumes that the 3D POR is positioned on a plane which consists of interest points. As a result, the 3D POR can be estimated in real-time. In addition to the method of estimating 3D PORs, a visualization technique for gaze trajectories using augmented reality technique is realized. Therefore, even when the scene camera moves dynamically, the gaze trajectories could be kept on environments. We confirmed the feasibility of our proposed method experimentally.

For future work, we need to solve the scale of world coordinates. We believe that it is not necessary to make a map from scratch every time, so localization will be archived using a map that is obtained previously.

## Acknowledgements

This research was supported by Japan Society for the Promotion of Science, Grant-in-Aid for Young Scientists B (No.20700117).

## References

DAVISON, A., REID, I., MOLTON, N., AND STASSE, O. 2007. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 6, 1052–1067.

KLEIN, G., AND MURRAY, D. 2007. Parallel tracking and mapping for small ar workspaces. In *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 1–10.

LI, D., BABCOCK, J., AND PARKHURST, D. 2006. openeyes: a low-cost head-mounted eye-tracking solution. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, 95–100.

MITSUGAMI, I., UKITA, N., AND KIDODE, M. 2003. Estimation of 3d gazed position using view lines. In *Proceedings of the 12th International Conference on Image Analysis and Processing*, 466–471.

MUNN, S., AND PELZ, J. 2008. 3d point-of-regard, position and head orientation from a portable monocular video-based eye tracker. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, 181–188.

ROSTEN, E., AND DRUMMOND, T. 2006. Machine learning for high-speed corner detection. In *Proceedings of the 9th European Conference on Computer Vision*, 430–443.

SMITH, R., AND CHEESEMAN, P. 1986. On the representation and estimation of spatial uncertainty. *The International Journal of Robotics Research* 5, 4, 56–68.

SUMI, K., SUGIMOTO, A., MATSUYAMA, T., TODA, M., AND TSUKIZAWA, S. 2004. Active wearable vision sensor: recognition of human activities and environments. In *Proceedings of IEEE International Conference on Informatics Research for Development of Knowledge Society Infrastructure*, 15–22.

TAKEMURA, K., IDO, J., MATSUMOTO, Y., AND OGASAWARA, T. 2003. Drive monitoring system based on non-contact measurement system of driver's focus of visual attention. In *Proceedings of the 2003 IEEE Intelligent Vehicles Symposium*, 581–586.

WOODING, D. 2002. Fixation maps: quantifying eye-movement traces. In *Proceedings of the 2002 symposium on Eye Tracking Research & Applications*, 31–36.