3-D Gaze Scan Path by Inside-out Camera System

Hironobu Fujiyoshi†, Makoto Kimura, Shoichi Shimizu††, Yuji Yamauchi†, Takayoshi Yamashita†
†Chubu University, ††Mitsubishi Electric Corporation
hf@cs.chubu.ac.jp, makoto.kimura@gmail.com, Shimizu.Shoichi@ab.MitsubishiElectric.co.jp

Abstract

Measuring the 3-D gaze trajectory of a user moving dynamically in 3-D space is interesting for understanding a user's intention and behavior. In this paper, we present a system for recovering 3-D scan path and scene structure in 3-D space on the basis of ego-motion computed from an inside-out camera. Experimental results show that the 3-D scan paths of a user moving in complex dynamic environments were recovered.

1. Introduction

The goal of First Person Vision(FPV) is to make the system environmentally-aware in order to recognize the key elements of a scene by using the camera, including static objects and the 3-D structure of a scene [1]. Combining these key elements (scene structure) and the gaze point of a user makes it easier to understand a user's intention and behavior. Our goal is to recover the 3-D scan path of a user moving dynamically, and to recover a scene structure in 3-D space by using a inside-out camera system. In this paper, we present a system for recovering 3-D scan path and scene structure in 3-D space on the basis of ego-motion computed from an inside-out camera.

2. Inside-out camera system

Our prototype inside-out camera system consists of two eye cameras installed at the top of the unit for capturing images of the user's eyeballs and two scene cameras installed at the bottom of the unit for capturing the user's visual field [2]. As shown in Figure 1, the inside-out camera achieves an optical configuration in which transparent cameras seem to exist.

Eye camera The eye camera system consists of an infrared mirror, two infrared cameras for capturing the left and right eyeballs, and six infrared LEDs arranged around each camera. Each infrared camera captures a near-infrared image of the user's eyeball from in front of that eye via an infrared mirror at a resolution of 640·480 pixels. The LEDs arranged around each camera emit near-infrared light in a wavelength range of 750 - 900 nm. Since infrared light is invisible, it provides no visual stimuli, enabling images of the eyeballs to be captured unhindered.

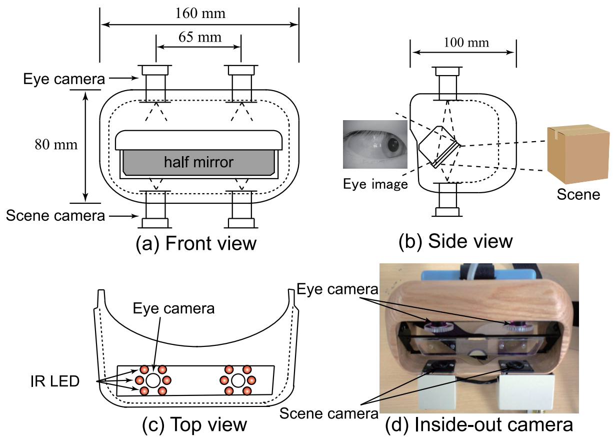

Figure 1. Inside-out camera system

Scene camera The scene camera system consists of a half mirror and two compact CCD cameras for capturing the left and right visual fields. The viewing angle of each CCD camera is about 80 degrees, and the focal length is about 4 mm. The half mirror reflects 50% of incident light and allows the rest to pass. Such use of a half mirror makes it possible to capture images with a transparent camera from a position that is optically nearly the same as the user's view point.

3. 3-D Gaze scan path estimation

Figure 2 shows the process for recovering a 3-D scan path and scene structure [3].

3.1. Ego-motion estimation

First, we recover the ego-motion of a user moving dynamically in 3-D space. Given a set of tracked feature points, $\boldsymbol{m}_i = (u_i, v_i, d_i)^{\mathrm{T}}$, for $i = 1, 2, ..., n$, in the current frame, and the set of corresponding feature points $\boldsymbol{m}'_i = (u'_i, v'_i, d'_i)^{\mathrm{T}}$ in the previous frame, we seek to esti-

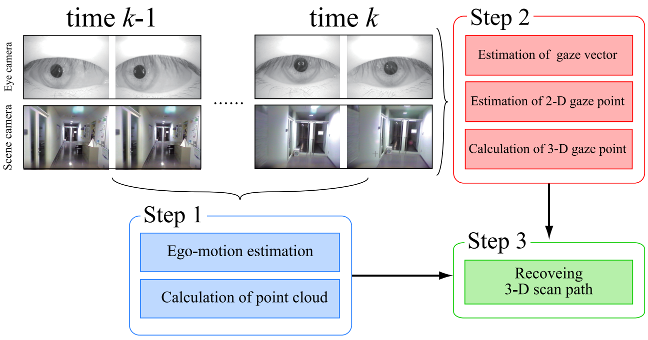

Figure 2. Process for recovering 3-D scan path and scene structure

mate the rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$ by the method of ego-motion estimation proposed in [4]. We also recover the scene structure by identifying image features in common between the two scene cameras and by matching features across images in a video sequence.

3.2. 3-D gaze point estimation

The 3-D gaze point is estimated by the eye cameras and scene cameras. The gaze vector $\boldsymbol{v}_k = (u_v, v_v)^{\mathrm{T}}$ at time k, which base is taken to be the cornea curvature center $\boldsymbol{c}_k$, can be calculated by the equation $\boldsymbol{v}_k = \boldsymbol{p}_k - \boldsymbol{c}_k$. The 2-D gaze point $\boldsymbol{g}_k = (u_g, v_g)^{\mathrm{T}}$ on the image plane of the scene camera can be calculated by using the estimated gaze vector and a conversion equation $\boldsymbol{g}_k = \boldsymbol{a}\boldsymbol{v}_k + \boldsymbol{b}$, where $\boldsymbol{a} = (a_u, a_v)^{\mathrm{T}}$ is the slope and $\boldsymbol{b} = (b_u, b_v)^{\mathrm{T}}$ is the intercept of the linear equation.

Since these 2-D gaze points are points on two scene cameras, the 3-D gaze point can be calculated as a problem in stereo matching. We can solve the problem of the 2-D gaze vectors not intersecting in the 3-D space by treating it as a problem in stereo matching. Finally, the 3-D gaze point $\boldsymbol{G}_k^c = (X^c, Y^c, Z^c)^{\mathrm{T}}$ in the camera coordinate system is calculated by using stereo matching. The accuracy of its 3-D gaze point is equal to that of stereo matching.

3.3. Recovering 3-D scan path over time

Finally, the 3-D gaze point $\boldsymbol{G}_k^c$ at time k in the camera coordinate system is converted to the global coordinate system by using ego-motion which consists of the rotation matrix $\boldsymbol{R}_k$ and the translation matrix $\boldsymbol{t}_k$ computed in section 3.1 by $\boldsymbol{G}_k^w = \boldsymbol{R}_k \boldsymbol{G}_k^c + \boldsymbol{t}_k + \boldsymbol{p}_{k-1}$, where $\boldsymbol{p}_{k-1}$ is the world location of the camera (head) at time $k-1$. The point cloud, which is the 3-D structure of the scene is recovered by identifying image features between the two scene cameras. This point cloud recovered at time k, is also converted to the global coordinate system as same as the 3-D gaze point.

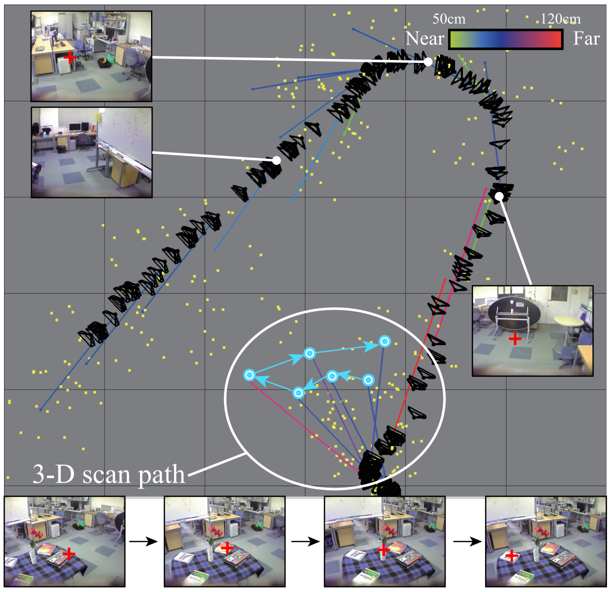

Figure 3. The 3D points of the scene structure are shown with yellow color. The blue circles show the estimated 3-D gaze points, and the blue arrows show the 3-D scan path obtained by connecting the 3-D gaze points at each frame.

4. Experimental Results

Figure 3 shows the 3-D scan path of a user walking around a table. It can be seen that these scan paths and the point cloud are precise enough to understand their spatial relations.

5. Conclusion

In this paper, we described a system for recovering 3-D scan paths and scene structures in 3D space on basis of ego-motion obtained from an inside-out camera. The 3-D scan paths and scene structures obtained from our system can be used to understand a user's intentions, such as the level of interest in an object. Our future work includes finding applications for the 3-D information of gaze and scene structures by using our inside-out camera system.

References

[1] T. Kanade and M. Hebert. First-person vision. *Proc. of the IEEE*, 100(8):pp.2442–2453, 2012.

[2] S. Shimizu and H. Fujiyoshi. Acquisition of 3d gaze information from eyeball movements using inside-out camera. In *Proc. of the 2nd Augmented Human Int. Conf.*, p.6, 2011.

[3] Y. Goto and H. Fujiyoshi. Recovering 3-d gaze scan path and scene structure from inside-out camera. In *Proc. of the 4th Augmented Human Int. Conf.*, pp.198–201, 2013.

[4] H. Badino and T. Kanade. A head-wearable short-baseline stereo system for the simultaneous estimation of structure and motion. In *IAPR Conf. on Machine Vision Applications*, 2011.