A Bi-directional Visual Stereo Interface for Accessing Stereo Matching Results from a Human Brain*

Sheng-Wen Shih and Tzu-Hsuan Lo

Department of Computer Science and Information Engineering
National Chi Nan University
1 University Road, Puli, Nantou 545, Taiwan
E-mail: swshih@ncnu.edu.tw

Abstract

In this paper, a novel interface, termed BVSI, allowing people to input stereo correspondences more efficiently is proposed. The BVSI consists of a stereo display for providing a stereo view of a 3-D scene and a binocular gaze tracker for recording the binocular fixation points of an operator. A calibration method of the BVSI is presented in this paper. The horizontal gaze tracking error of the calibrated BVSI is about ± 10 pixels. Real experiments have been conducted and the results showed that the proposed system is very promising.

1 Introduction

Biological stereo vision systems have been proven to be very successful in the natural world. Therefore, many researchers have payed much attention to stereo matching algorithms in order to build an artificial stereo vision system. However, the major problem of building a stereo vision system is the lack of a robust stereo matching algorithm. Although the stereo matching problem has been studied for many decades, the ability of the existing stereo matching algorithms are still inferior to the stereo vision ability of mankind. Therefore, in many practical applications such as reverse engineering and terrain modeling from aerial images, automatic stereo matching results are verified and are corrected by humans to yield appropriate 3-D models. In some systems, e.g., [3], the stereo correspondences are even directly established manually. In general, editing stereo correspondences is very time consuming because the amount of stereo data is usually very large and the adjustment of the correspondence map has to be done point-wisely. Interestingly, when the operator spending time on maneuvering the pointing devices to adjust the correspondence map, he/she has stared at the correct stereo correspondences for a while. Motivated by this observation, we have developed a more efficient interface for interactive 3-D reconstruction.

In this paper, a *Bi-directional Visual Stereo Interface* (BVSI) for accessing the stereo matching results from a human brain through binocular gaze tracking is proposed. The BVSI mainly consists of a stereo display and a binocular gaze tracker. Because the gaze tracking accuracy is not high enough, we plan to use the BVSI solely to provide initial stereo correspondences for guiding the subsequent automatic stereo matching. In addition to providing an efficient interface for input stereo correspondences, there are several potential applications of the BVSI such as for training stereo matching criteria [5] and for designing fixation selection process in a foviated active vision system.

In the past the gaze tracking systems are mainly used for research purpose in laboratories or for disabled persons because the gaze tracking systems either are intrusive (e.g., requiring users to wear sclera coil or to place electrodes around the eyes) or necessitate users holding their head quite still [4, 2]. Nowadays, many commercial products and improved gaze tracking techniques have been developed to alleviate problems of using gaze tracking systems. Some researchers started to develop new user interfaces with eye-controlled interaction [6]. Besides, many computer graphic systems also utilize eye tracking results to provide realistic 3-D rendering effects [7]. The BVSI is different from the graphic systems equipped with a gaze tracker [6, 7] in that the main function of the BVSI is an input device while the main function of the existing systems are output devices (to display graphics).

2 System Overview

Fig. 1 shows a schematic diagram of the BVSI which is composed of a flat-square CRT stereo display and a binocular gaze tracker. The 19” CRT monitor (CTX-PR960F)

* This work is supported in part by the National Science Concil of Taiwan under Grants NSC 89-2218-E-280-003.

has been calibrated such that the relationships between the 2-D WCS (Window Coordinate System) and x-y plane of the 3-D SCS (Screen Coordinate System), which is aligned with the screen of the monitor, are known. The relationships between the 2-D WCS and the x-y plane of the SCS are described by four second order binomials to cope with the nonlinear geometric distortion of the CRT monitor, i.e., mapping from the WCS to the SCS:

$$[x_s\ y_s\ 0]^t = \left[{}^S B_{W,x}(x_w, y_w)\ {}^S B_{W,y}(x_w, y_w)\ 0\right]^t, \quad (1)$$

or conversely, mapping from the SCS to the WCS:

$$[x_w\ y_w]^t = \left[{}^W B_{S,x}(x_s, y_s)\ {}^W B_{S,y}(x_s, y_s)\right]^t, \quad (2)$$

where ${}^S B_{W,x}(\cdot,\cdot)$, ${}^S B_{W,y}(\cdot,\cdot)$, ${}^W B_{S,x}(\cdot,\cdot)$ and ${}^W B_{S,y}(\cdot,\cdot)$ are second order binomials, and $(x_s, y_s, 0)$ and (x_w, y_w) are a pair of correspondence points in the SCS and the WCS, respectively. The 3-D coordinates of the three fiducial marks adhered to the monitor have also been measured with respect to the SCS during the screen calibration process. Those three fiducial marks can be used to retrieve the origin and the orientation of the SCS. The other main component of the BVSI is the binocular gaze tracker which consists of a head tracker and a pair of pupil trackers which will be described in detail in the following sections. The computation loads of the BVSI are distributed on three PCs in order to achieve real-time interaction: Stereo rendering is accomplished by the first PC with an Elsa Gloria XXL graphic card. Head tracking and binocular pupil tracking are implemented on the second and the third PCs with frame grabbers, respectively.

3 2-D Pupil Tracker

Fig. 2 shows the stereo goggle that were composed by using an Elsa Revelator, two JAI M536 micro head cameras, and five IR (Infrared) LEDs. Three of the IR LEDs shown in Fig. 2 are used as the tracking targets of the head tracker. The remaining two are used to shine the irises of user's eyes to obtain black pupil images as shown in Fig. 3. Through simple thresholding and connected component analysis, the centers of both the left and right pupil images can be easily computed in video rate (30 Hz). The 2-D coordinates of the detected left and right pupil centers are denoted by p_l and p_r, respectively. The way to map p_l and p_r to binocular 3-D LoSs (Lines of Sight) will be described in section 5.

4 Head Tracker

As shown in Fig. 1, the head tracker is composed of two calibrated stereo cameras with a reference frame called the HCS (Head tracker Coordinate System). The coordinate transformation matrix from the SCS to the HCS, denoted by ${}^H\mathbf{T}_S$, has also been calibrated previously. The head tracker actually track three IR LEDs mounted on the stereo goggle instead of tracking facial features. To improve the robustness of the head tracking results, an IR filter is attached in front of each camera lens. Therefore, the 3-D coordinates of those three LEDs can be easily and robustly estimated in video rate (30 Hz). The reference frame of the goggle, termed as the GCS (Goggle Coordinate System), is defined by using the locations of the three LEDs. Whenever the 3-D coordinates of the three LEDs are computed by the head tracker with respect with the HCS, one can easily determine the coordinate transformation matrix from the HCS to the GCS, ${}^G\mathbf{T}_H$.

5 Binocular 3-D Lines of Sight

In this section, a calibration method is presented. By using the proposed calibration method, system parameters can be easily determined so that the 2-D pupil tracking results and the 3-D head tracking results can be combined to obtain the binocular 3-D LoSs. Fig. 4 shows the calibration arrangement of the BVSI, where an Immersion Microscribe, a low cost coordinate measurement machine, is used in the calibration process.

At first, the Microscribe is used to measure the 3-D coordinates of the fiducial marks of the stereo display with respect to the MCS (Microscribe Coordinate System). Since the 3-D coordinates of the fiducial marks in the SCS are also known, we have three pairs of 3-D correspondence points. Thus, the coordinate transformation matrix from the MCS to the SCS, denoted as ${}^S\mathbf{T}_M$, can be computed by using the algorithm described in [1].

During the calibration process, the user is asked to stare at each of the 16 target points on the screen while holding his head steady by using a bite bar for five seconds (see Fig. 4). The median value of the collected pupil centers are computed to filter out the micro-saccade effect and to provide stable binocular pupil centers, p_l and p_r. Given the 2-D coordinates of the target point, (x_w, y_w), its corresponding 3-D coordinates in the SCS, ${}^S\mathbf{p}_t = [x_s\ y_s\ 0]^t$, can be computed by using equation (1). Next, we use the tool tip of the Microscribe as a “front sight” to aim at the target point. The 3-D coordinates of the tool tip of the Microscribe, denoted by ${}^M\mathbf{p}_{fs}$, can be directly measured and can be transformed into the SCS as follows: ${}^S\mathbf{p}_{fs} = {}^S\mathbf{T}_M\ {}^M\mathbf{p}_{fs}$. Based on ${}^S\mathbf{p}_{fs}$ and ${}^S\mathbf{p}_t$, the 3-D LoS of user's eye can be easily computed. Since ${}^H\mathbf{T}_S$ has been previously calibrated and ${}^G\mathbf{T}_H$ can be measured online, both the target point and the front sight point can be transformed into the GCS and denoted as ${}^G\mathbf{p}_t$ and ${}^G\mathbf{p}_{fs}$, respectively. Thus, the equation of the 3-D LoS is given by

$${}^G\mathbf{p}_l = {}^G\mathbf{p}_0 + s \cdot {}^G\mathbf{d}, \quad (3)$$

where ${}^{G}\mathbf{p}_0 = {}^{G}\mathbf{p}_{fs} - \frac{\mathbf{e}_3^t \, {}^{G}\mathbf{p}_{fs}}{\mathbf{e}_3^t \, {}^{G}\mathbf{d}} \, {}^{G}\mathbf{d}$, ${}^{G}\mathbf{d} = \frac{{}^{G}\mathbf{p}_t - {}^{G}\mathbf{p}_{fs}}{\|{}^{G}\mathbf{p}_t - {}^{G}\mathbf{p}_{fs}\|}$, $\mathbf{e}_3 = [0\ 0\ 1]^t$, and s is a scalar. Since a 3-D line has only four degrees of freedom, it can be fully specified by the following four parameters: x_0, y_0, d_x and d_y, where $[x_0\ y_0\ 0]^t = {}^{G}\mathbf{p}_0$ and ${}^{G}\mathbf{d} = \left[d_x\ d_y\ \sqrt{1 - d_x^2 - d_y^2}\right]^t$. Because the relative position and orientation of the stereo goggle and user's head is kept constant, there is an one-to-one mapping relation between the pupil center and any of the four parameters of the LoS. During the calibration process, eight second-order binomials are fitted to the calibration data for mapping pupil centers, p_l and p_r, to $x_{0l}\,(p_l)$, $y_{0l}\,(p_l)$, $d_{xl}\,(p_l)$, $d_{yl}\,(p_l)$, $x_{0r}\,(p_r)$, $y_{0r}\,(p_r)$, $d_{xr}\,(p_r)$, and $d_{yr}\,(p_r)$. Notice that, once the BVSI has been calibrated, the user does not require to use the bite bar anymore.

6 Experiments

Before using the BVSI, each user will have to calibrate the binocular gaze tracker first. Normally, it would take 10–15 minutes to complete the calibration. Fig. 5 shows the calibration results accomplished by a subject (the calibration data for the left and the right eyes were acquired separately), where the estimated fixation points and the target points are marked with '$\cdot$' and '$\circ$' signs, respectively. Through this experiment, we found that the horizontal gaze tracking results is more accurate than the vertical ones. This is due to an imperfection of the current design of our stereo goggle—the goggle still can slightly moves in the vertical direction when it is worn on head. However, since the purpose of the BVSI is to detected stereo disparities of the binocular fixation points, horizontal accuracy is our major concern. The current implementation of our gaze tracker offers about ± 10 pixels of horizontal gaze tracking accuracy, which is satisfactory for providing initial stereo matching input.

In the next two experiments, the calibrated BVSI was tested using a pair of stereo images. The stereo images were acquired by using a pair of calibrated stereo cameras. The acquired images were then rectified so that the corresponding epipolar lines are parallel to the x-axis and have the same y coordinate value. Fig. 6 shows the recorded binocular gaze tracking results of a subject. The rectified stereo images are shown in the first row and the recorded binocular gaze tracking results overlaid on the stereo images are shown in the second row. Notice that both the y-coordinates of the detected binocular fixation points have been set to the y-coordinate of the fixation point of subject's dominating (left) eye to make sure the estimated stereo correspondences satisfy the epipolar constraint. The subject first look at the foreground object (a metal can) for about 30 seconds and proceed to look at the background planar object for another 30 seconds. Trajectories of the fixations points recorded when he was staring at the foreground and background objects are plotted using different colors. Fig. 7 shows the 3-D trajectories of the binocular tracking results obtained by triangulation (the can is approximately at z = 1400mm). Although the stereo matching accuracy of the BVSI is not very high, the binocular gaze tracking results contain very rich and important information for automatic stereo matching such as the foreground objects, occluded regions, object boundaries, approximate disparity, etc..

7 Concluding Remarks and Future Work

Although the stereo matching problem has been studied for many decades, there is still no stereo matching algorithm that performs as well as humans do. In this paper, we proposed an interface, namely the BVSI, to access the stereo matching results from a human brain via tracking the binocular fixation points of a person. The BVSI has been calibrated using a method described in this paper such that the gaze tracking error is less than ten pixels in the horizontal direction. Real experiments have been conducted and the results showed that the binocular results can provide rich information, such as the occluding/occluded regions, approximate disparity, etc., for stereo matching. We believe that this kind of interface will be very useful in helping people solving many computer vision problems. Currently, we are developing interactive model for more efficiently guiding stereo matching using this interface.

References

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-square fitting of two 3-d point sets. *IEEE T. PAMI*, 9(5):698–700, 1987.

[2] T. Cornweet and H. Crane. Accurate two-dimensional eye tracker using first and fourth purkinje images. *J. Opt. Soc. Amer.*, 63(8):921–928, 1973.

[3] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *Proc. SIGGRAPH96*, pages 11–20, 1996.

[4] L. A. Frey, J. K. P. White, and T. E. Hutchinson. Eye-gaze word processing. *IEEE T. SMC*, 20(4):944–950, 1990.

[5] G. Pajares, J. M. Cruz, and J. A. Lopez-Orozco. Stereo matching using hebbian learning. *IEEE T. SMC, Part B*, 29(4):553–559, 1999.

[6] S. Pastoor, J. Liu, and S. Renault. An experimental multimedia system allowing 3-d visualization and eye-controlled interaction without user-worn devices. *IEEE T. Multimedia*, 1(1):41–52, 1999.

[7] A. Redert, E. Hendriks, and J. Biemond. 3-d scene reconstruction with viewpoint adaptation on stereo displays. *IEEE T. CSVT*, 10(4):550–562, 2000.

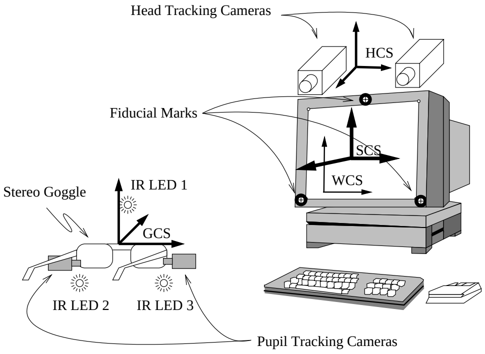

Figure 1. Basic components of the proposed system.

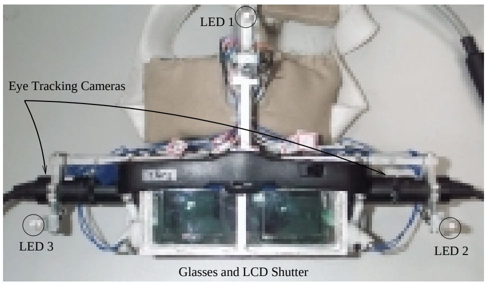

Figure 2. A stereo goggle equipped with a pair of pupil trackers.

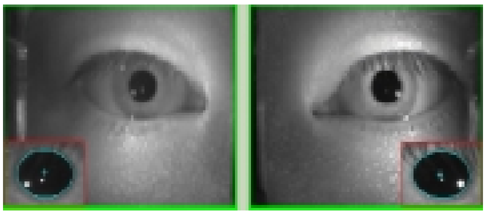

Figure 3. Typical tracked black pupil images of user's eyes.

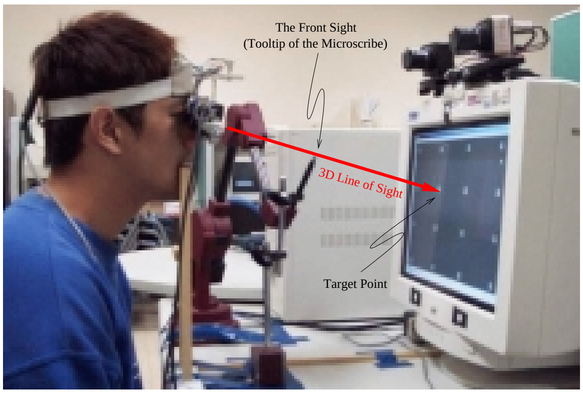

Figure 4. A picture showing the calibration arrangement of the BVSI.

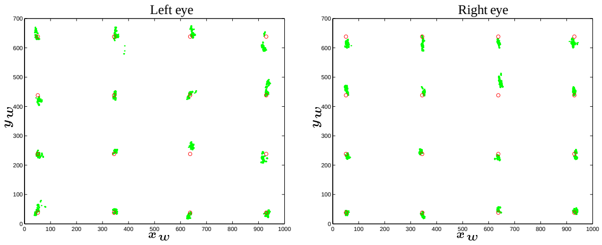

Figure 5. Calibration results of the binocular gaze tracker.

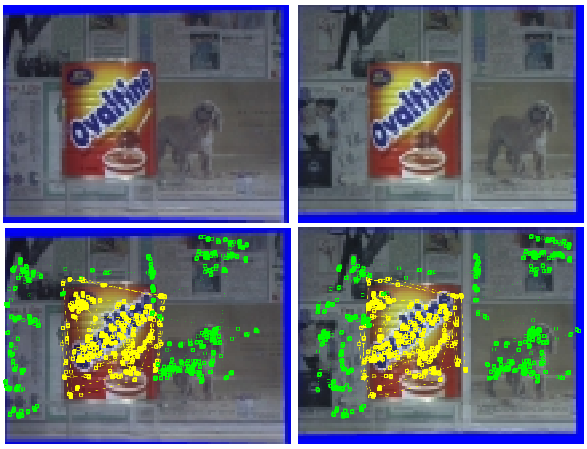

Figure 6. Binocular gaze tracking results for a scene.

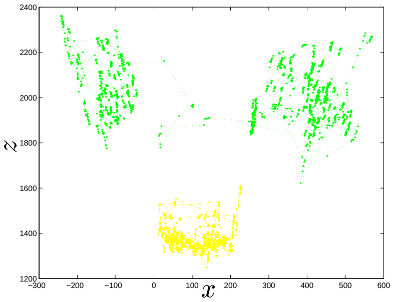

Figure 7. The x-z plot of the 3-D trajectories of the binocular gaze tracking results shown in Fig. 6.