

Combined Head - Eye Tracking for Immersive Virtual Reality

Hu Huang Robert S. Allison Michael Jenkin

Center for Vision Research, and the Department of Computer Science and Engineering
York University
4700 Keele St., Toronto, Ontario, Canada, M3J 1P3

huanghu@cs.yorku.ca, allison@cs.yorku.ca, jenkin@cs.yorku.ca

Abstract

Real-time gaze tracking is a promising interaction technique for virtual environments. Immersive projection-based virtual reality systems such as the CAVE™ allow users a wide range of natural movements. Unfortunately, most head and eye movement measurement techniques are of limited use during free head and body motion. An improved head-eye tracking system is proposed and developed for use in immersive applications with free head motion. The system is based upon a head-mounted video-based eye tracking system and a hybrid ultrasound-inertial head tracking system. The system can measure the point of regard in a scene in real-time during relatively large head movements. The system will serve as a flexible testbed for evaluating novel gaze-contingent interaction techniques in virtual environments.

The calibration of the head-eye tracking system is one of the most important issues that need to be addressed. In this paper, a simple view-based calibration method is proposed.

Key words: Immersive VR, eye-tracking, calibration.

1. Introduction

Determination of an observer's line of sight (*l.o.s*) has been the subject of a great number of studies, with applications in different fields including medical research, automobile driving research, weapons guidance, and assistive devices for handicapped people [1][2][3][4].

The movements of the eye can be modeled as rigid body rotation about a hypothetical single center of rotation fixed within the eye [5][6]. Under this assumption, besides translation or rotation of the head, the eyes add three independent degrees of freedom (DOF) for rotational movements. The eyes can be moved horizontally (azimuth), vertically (elevation) and torsionally (roll) with respect to the orbit. In reality the

eye does not have a well-defined center of rotation and any rotation of the eye involves a small translation. However, this small movement can be ignored in most practical situations [7][8][9]. A gaze measurement system should ideally measure head motion with six degrees of freedom (6DOF) and eye motion with three (3DOF). In this paper we are only interested in the elevation and azimuth angle of eye movement as the torsional movements do not redirect the line of sight. Thus, we will ignore the torsion direction.



Figure 1: A subject wearing the head and eye tracker in IVY at York University.

Combined head-eye tracking systems have been studied by a number of researchers [10][11][12]. However, they suffer from different limitations or are not easy to use. For example, the eye-tracker embedded HMD system developed at Clemson University [10] needs to be specially made and suffers from the limitations of the virtual reality HMDs; In Allison's research [11], a magnetic head tracker was used, which restricts it to a relatively small range of head movement due to the range of the head tracking system. The relative position

between the eye and head tracker was measured manually for each subject before the trial, which is inconvenient and error prone; As for Perez's approach [12], the display-fixed combined tracking system suffers from the occlusion and the low update rate of gaze position. These limitations make these systems ill-suited for CAVETM[13]-like immersive applications.

A combined head-eye tracking system is described in this paper. The system can be used in both desktop and immersive applications. However, it is particularly well-suited to large screen projection based virtual reality (VR) systems like CAVETM since it allows tracking during large amplitude natural head and body movements.

2. Approach and Method

Combined System

The combined head-eye gaze-tracking system consists of a video eye-tracking system and a magnetic-inertial head tracking system (Figure 2)[14]. The eye-tracking subsystem used for this work is the VISION2000 developed by EL-MAR Inc. (Toronto, Canada) [15]. It is based on the pupil/corneal reflex technique and uses adaptive real-time image processing to obtain accurate measurements of eye position. Both the horizontal and the vertical eye-movements are measured 120 times per second with a resolution of $\pm 0.1^\circ$. The tracking range covers $\pm 45^\circ$ for both directions [16]. Although both eyes are tracked, only one eye's information (monocular) is used to determine the line of sight in this work.

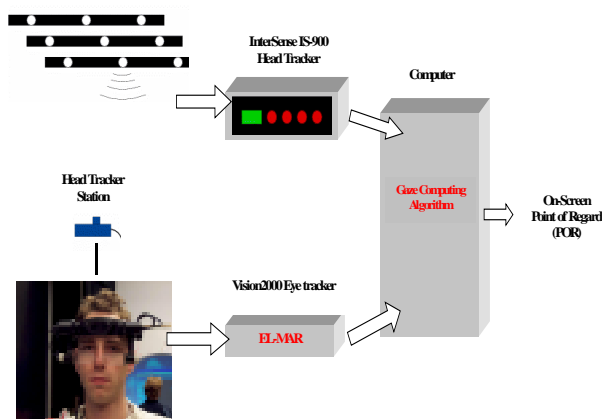


Figure 2: Construction of the combined system

The head-tracking subsystem, InterSense's IS-900, is a hybrid acoustic-inertial 6DOF position and orientation tracking system [17]. It is made up of SoniStripsTM, Tracked Devices or Stations, and a Processor Unit. The tracking stations make use of the InertiaCube, an ultra-miniature sensor module, to simultaneously measure six physical properties, namely angular rates and linear accelerations along all three axes. Changes in orientation and position are calculated by integrating the output of

its gyroscopes and accelerometers, and drift is corrected by using a room-referenced ultrasonic time-of-flight range measuring system. The fusion software then combines the inertial and acoustic range data with a set of algorithms to produce a motion predicted, 6DOF solution of the tracked station.

IVY

Our main immersive virtual environment generator is the Immersive Visual Environment at York (IVY)[18][19][20]. IVY is a fully immersive (six-sided) projective stereo visual environment that can be used for a range of tasks from structure visualization to studying issues related to human perception in real and virtual environments. This CAVE-like display is driven by a nine-workstation Linux cluster. Both the development and 'production' systems are based on the VE library (see Appendix) and are completely compatible with the head-eye tracking system described here.

In IVY, the video eye-tracking system (VISION2000) was used to track the eye movement, while the InterSense's IS-900 tracking system was used to track the user's head movement. The fixed frame (SoniStripsTM) containing the acoustic beacons were mounted in the doorway to IVY, blocking one wall and making IVY essentially a five-walled environment. Figure 3 shows the Immersive Visual Environment and the set-up of IS-900 tracking system in IVY [20].

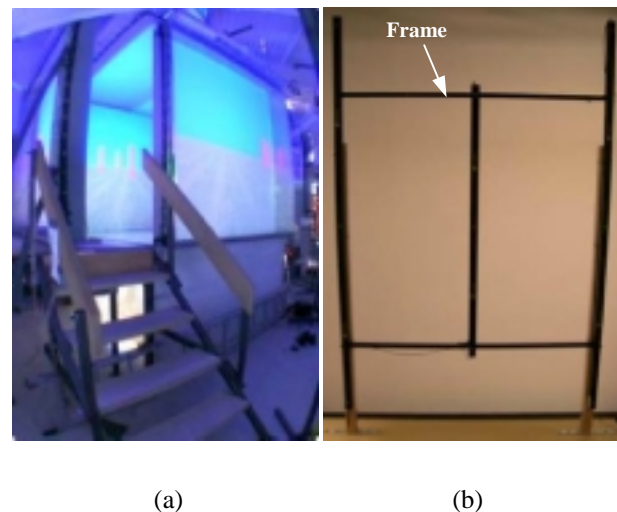


Figure 3: Immersive Visual Environment at York - IVY (a) and the IS-900 tracking frame in IVY (b)

The coordinate systems (CS) of this combined system is shown in Figure 4. Correct estimation of the gaze (point of regard) in a scene in this combined system relies not only on the performance of each individual subsystem but also on the calibration of the whole system.

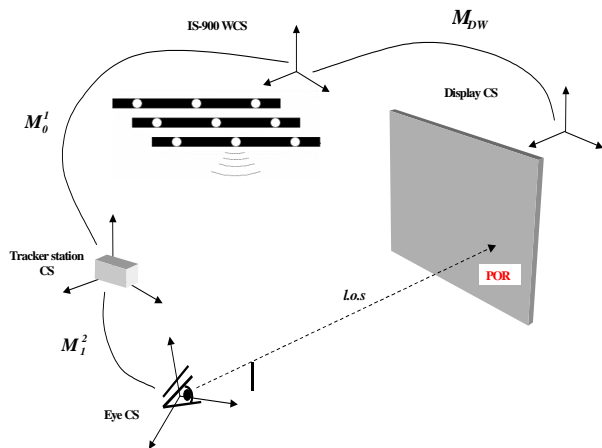


Figure 4: Coordinate systems of the combined system

Calibration

The large area head tracking system – IS-900 was calibrated during the installation process, thus the relationship between the display screen coordinate system and the world coordinate system is considered to be fixed and a one-time calibration can be performed to establish the transformation. For each subject, it is different in headgear placement, so the spatial relationship between the head tracker and the subject's eye varies. Calibration must be performed to find the transformation matrix between them. The calibration of the eye tracker also needs to be performed before each session. Thus the calibration procedure can be divided into three steps.

Firstly, a one-time calibration step establishes the spatial relationship between the display screen coordinate system and world coordinate system, i.e. this process finds the transformation matrix M_{DW} from the display screen coordinate system to the world coordinate system. To do this, four corners of the screen (the fourth one is used for verification purpose only) must be precisely registered in the world coordinate system.

Secondly, a view-based calibration technique is used to find the spatial relationship M_i^2 between the head tracker station and the eye. The subject wearing the head tracker and eye tracker is asked to direct his/her gaze to align the centers of two crosshairs mounted on each end of a tube. The geometry of the crosshairs (position and orientation) is known in the world coordinate system precisely. The boresight alignment is repeated several times with slightly different head orientation. At each alignment, the head tracker station position and orientation are recorded. Each measurement obtains one ray that passes through the eye center. The over-determined set of constraints is used to solve for the position of the eye in the tracker station space, i.e. T_1^2 .

The subject is then asked to sit straight and adjust the sitting position to line up the right eye, the center of the

screen, where a computer-generated point was displayed, and a point in between. This point is carefully configured in advance to ensure the connection line between this point and the screen center is perpendicular to the screen plane. We also assume that the frontal plane of the subject is now parallel to the display screen, thus the normal vector of the screen v_1 is consistent with the subject's line of sight at the primary position (where visual angles along the horizontal and vertical meridians of the eye are defined to be 0). So finding the rotation relationship between the eye and the head tracker station becomes finding the rotation matrix R_1^2 to map the head tracker vector v_2 to the display screen normal vector v_1 . From now on, the subject's head is kept still (with the help of chinrest) until the next calibration step is finished.

Lastly, the calibration of the eye tracker is performed through a dedicated calibration routine. Targets are presented at known visual angles along the horizontal and vertical meridians of the eye. When the subject is fixating the desired target the pixel position of the corneal reflexes and the pupil center are measured. A linear regression algorithm adjusts the gain between the corneal reflex/pupil motion in the image and the known eye movements to fit the horizontal and vertical calibration data.

Line of Sight (l.o.s) and Point of Regard (POR) Estimation

Measurement of eye and head position and orientation allows for estimation of the line of sight and the point of regard (POR). The line of sight is a ray that originates from the center of the eye and follows the eyesight. It is initially described as a vector $v_0 = [0 \ 0 \ 1]^T$ (the z-axis of the world coordinate system) starting at the origin $o_0 = [0 \ 0 \ 0]^T$ (Here, the standard OpenGL coordinate system is used). To measure POR the intersection of the ray with the stimulus plane must be determined.

The eye position (3D) in world coordinate system is determined by the estimate from head tracking system and the spatial relationship M_i^2 found by calibration step 2. The eye tracker allows measurement of the line of sight in the eye coordinate system, which is a head fixed frame. A set of transformations needs to be performed to determine the line of sight in the world coordinate system.

The first transformation is from the world coordinate system to the head tracker coordinate system described by

$$M_o^1 = \begin{bmatrix} R_o^1 & T_o^1 \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

where the translation vector T_o^1 is the position

information obtained from the head tracker (IS-900) station and the rotation matrix R_0^1 is generated based on the quaternion information from the station. The second transformation matrix is from the head tracker station coordinate system to the eye coordinate system

$$M_I^2 = \begin{bmatrix} R_I^2 & T_I^2 \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

which is acquired from calibration step2. Finally, we can calculate the rotation matrix of the eye R_2^3 based on the estimation of azimuth and elevation obtained from the eye tracker. After the series of transformation, the eye position in the world coordinate system is:

$$o_0^3 = o_0^2 = M_0^1 M_I^2 o_0$$

The line of sight is described in the world coordinate system as:

$$l.o.s = v_0^3 = R_0^1 R_1^2 R_2^3 v_0$$

The calculation of an intersection between a ray and a plane ($Ax + By + Cz = D$, $n = [A \ B \ C]^T$) is used to find the point of regard (POR) P_w :

$$P_w = o_0^3 + t \cdot v_0^3$$

where

$$t = \frac{-(n \cdot o_0^3 + D)}{n \cdot v_0^3}$$

In order to display the corresponding gaze point on the screen, conversion of the coordinate from world coordinate system to screen coordinate system is necessary as:

$$P_D = M_{WD} P_w$$

where M_{WD} is the inverse matrix of M_{DW} that we established in calibration step1.

3. Applications

As an illustration of the utility of the combined system, this section describes some applications of this system. The first application is a demo of gaze contingent display system. The second application is a 3D eye-picking demo. And at last applications in IVY are discussed.

Single Screen System (Desktop or Projection based)

In this system, we used a standard Linux workstation (2 AMD Athlon MP 1900+ processors, 1024 MB System RAM, GeForce4 TI4600, Red Hat Linux with kernel version 2.4.18) to calculate the line of sight and point of

regard and to generate the images for the virtual environment

2D rendering application

This application was developed using GLUT (an extension of OpenGL). Figure 5 shows the screenshot of the application. It is a demonstration of using eyetracking in LOD display. Only the part of picture within the high-resolution window (inset) is rendered with high details while the rest of the picture is displayed in low resolution. This is achieved by OpenGL alpha blending functionality. The estimated gaze position was used to control the position of the inset window.

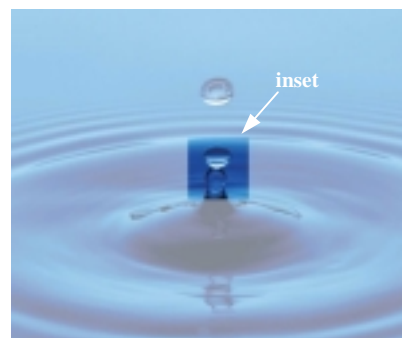


Figure 5: Screenshot of the 2D rendering application

3D picking application

This application was developed using GLUT too. The eye movement was used to represent the mouse movement. Picking is performed in this application by pressing a key on the keyboard or a game pad. Figure 6 is a screenshot of this application. The eye movement and the key press were used to take the place of mouse move and click. Key presses could be replaced by eye blinking, which would be more suitable for the case when both hands are occupied.

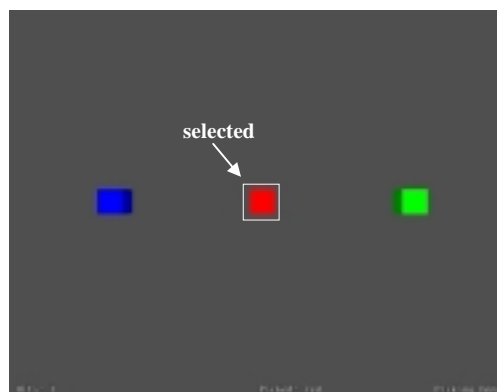


Figure 6: Screenshot of the 3D picking application (red box is selected)

Working in IVY

The similar procedure of estimating the line of sight can

be applied to IVY. And the information of line of sight can be used to find point of regard on the screen or the object of interest within the environment.

Determine the POR on the Screen

To estimate the point of regard on the screen, the method is similar to the one described above. Since there are six projection screens in the system (five with IS-900 tracking system setup), the ray/plane intersection is repeated for each of the six sides of IVY and the intersection point that has the shortest distance to the eye is used (Figure. 7). To make the algorithm work correctly, similar calibration procedures described in the previous section need to be performed before each application.

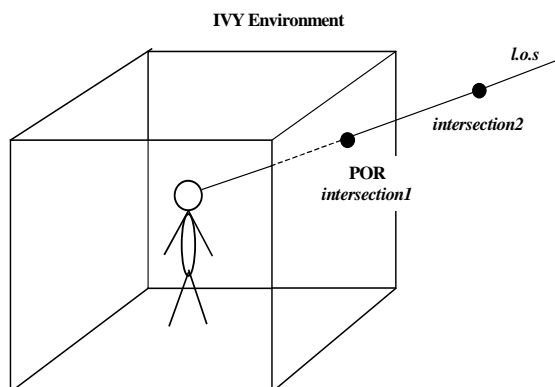


Figure 7: Method in IVY: there are two intersection points in this case, the point of regard (POR) is the nearest one to the subject's eye.

Determine the Object of Interest

As the nature of VE library, the eye origin and orientation can be easily abstracted from the eye frame. The line of sight can be considered as a 3D line in the virtual environment. So it can be directly used to perform a ray/polygon intersection to find the object that the user is looking at (Figure 8).

4. Conclusion and Future Work

A calibrated combined head-eye tracking system has been developed. This system can calculate the line of sight and determine the point of regard on a desktop screen or a CAVE-like display while allowing free head and body movement. Evaluation experiments will be performed in future work.

5. Acknowledgements

The financial support of Center for Research in Earth and Space Technology (CRESTech), Canadian Foundation for Innovation (CFI), and Natural Sciences and Engineering Research Council (NSERC), is gratefully acknowledged. We also thank Professor Moshe Eizenman for the valuable advice and help with this project.

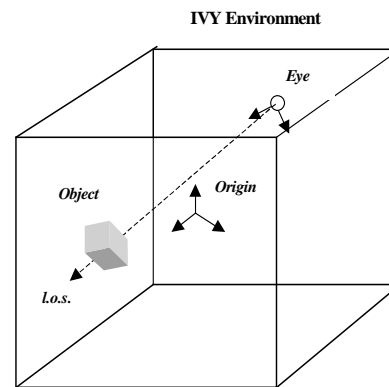


Figure 8: Determine the object of interest in IVY environment

Appendix: VE Library

Applications for IVY can be written in any graphical software package that runs on SGI or Linux. In order to simplify software development a package – VE library has been built to abstract the various display and input technologies required by applications [21].

The VE library is largely concerned with managing an *environment*. An environment is a physical space in the real world in which the user is located. This environment interfaces with the virtual world by means of input devices and walls/windows. *Walls* are portals in the virtual world which are rendered in the real world as *windows*. For example, a wall's center may have a particular location in the virtual world and is oriented so that its normal is aligned with the z-axis and "up" is considered to be in the direction of the positive y-axis. The image that would be generated if the virtual world were projected onto that wall would be displayed in a window, which is a real output device - e.g. a particular area on an X display. A window can be thought of as the output of a wall (see Figure 9).

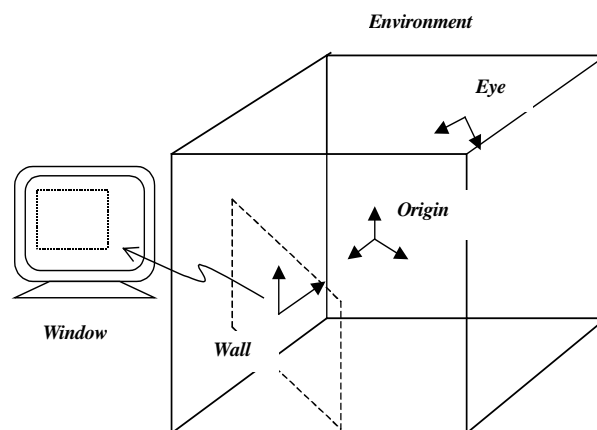


Figure 9: Concept of VE

There are several coordinate systems or *frames* need to be considered. They are *world frame* which is the coordinate system of the virtual world; *environment frame*

which is the physical real-world coordinate system of the virtual environment, it is where the screens, input devices, the user, etc. are located; *Eye frame* which is the view of the world from the perspective of the camera or viewer and *screen frame* which is the view of the world from the perspective of the screen onto which the image is rendered. The frame of the eye is expressed in environment coordinates system and thus relates the eye to the environment. One must combine frames in order to relate the eye to the world. The origin and the direction of eyesight can be easily abstracted from the eye frame.

By altering the environment's frame, the environment can be moved through the virtual world. This is typically how one could simulate vehicle motion - i.e. motion independent of the user's movements within the environment. An eye frame typically represents the user's location within the environment. This information is critical to generating the correct projections of the world on the walls.

References

- [1]Reuter, B. W. and Schenck, U. (1985), "Data acquisition in visual screening of microscopical samples". In Proceedings of Image Science '85 Acta Polytechnica Scandinavica, Applied Physics Series, pp.206-209.
- [2]Salvucci, D. D., Liu, A. and Boer, E. R., (2001), "Control and monitoring during lane changes", Vision in Vehicles 9.
- [3]Birt, J. A. and Furness, T A.(2003), "Visually Coupled Systems".<http://www.airpower.maxwell.af.mil/airchronicles/aureview/1974/mar-apr/birt.html>.
- [4]Yamada, M. and Fukada, T. (1987), "Eye Word Processor (EWP) and Peripheral controller for ALS patient". In Proceedings of IEEE, part A: Physical Science, Measurement and Instrumentation, Management and Education, Reviews, pp. 328-330.
- [5]Duke-Elder, S. and Wybar, K. (1973), "Ocular Motility and Strabismus". In system of Ophthalmology. S. Duke-Elder (ed.), Vol. VI, Mosby: St.Louis.
- [6]Westheimer, G. (1957), "Kinematics of the Eye". J Opt. Soc. Am. Vol. 47(10), pp.967-974.
- [7]Alpern, M. (1962), "Movements of the Eyes in: The Eye", Hugh Davson (ed.), vol. 3, pp. 7-27, Academic Press.
- [8]Westheimer, G. (1981), "Donder's, Listing's and Hering's Laws and Their Implications". In Models of Oculomotor Behaviour and Control. B. Zuber (ed.), pp. 149-159.
- [9]vonNoorden, G.K. (1990), "Binocular Vision and Ocular Motility", Fourth Edition. Mosby: St. Louis.
- [10]Duchowski, A. T., Medlin, E., Gramopadhye, A., Melloy, B. and Nair, S. (2001), "Binocular eye tracking in VR for visual inspection training", Proceedings of the ACM symposium on Virtual Reality Software and Technology, pp. 1- 8.
- [11]Allison, R. S., Eizenman, M., Cheung, B.S.K. (1996), "Combined head and eye tracking system for dynamic testing of the vestibular system", IEEE transaction on biomedical engineering, VOL. 43, NO. 11.
- [12]Perez, A., Cordoba, M. L., Garcia, A. Mendez, R., Munoz, M. L., Pedraza, J. L. and Sanchez., F. (2003), "A Precise Eye-Gaze Detection and Tracking System", In Proceedings of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2003.
- [13]Carolina, C, Sandin, D. J., DeFanti, T. A., (1993), "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE". Proc. SIGGRAPH '93, pp. 135-142.
- [14]Huang, J., Schumacher, J., Allison, R. S. and Zacher, J. E. (2003), "What are you looking at?" CRESTech's Annual General Meeting.
- [15]Manual of EL-MAR VISION2000, "Video Eye-Tracking System". Version 5.0, June, 2002.
- [16]DiScenna, A. O., Das, V., Zivotofsky, A. Z., Seidman, S. H. and Leigh, R. J. (1995), "Evaluation of a video tracking device for measurement of horizontal and vertical eye rotations during locomotion", Journal of Neuroscience Methods 58, pp. 89-94.
- [17]InterSense's IS-900 VET & VWT Manual, Appendix A: IS-900 Theory of Operation and Specification, Doc. No. 072-00060-0F00 Revision 1.0.
- [18]Robinson, M., Laurence, J., Zacher, J. E., Hogue, A., Allison, R. S., Harris, L. R., Jenkin, M. and Stuerzlinger, W.(2002), "IVY: The Immersive Visual environment at York", IPT Symposium (held in conjunction with VR2002).
- [19]Robinson, M.,Laurence, J., Hogue, A., Zacher, J. E., German, A. and Jenkin, M., (2002), "IVY: Basic design and construction details", Proc ICAT 2002, Tokyo, Japan.
- [20]Yan, L., Allison, R. S. and Rushton, S. K. (2004), "New Simple Virtual Walking Method -Walking on the Spot". 9th Annual Immersive Projection Technology (IPT) Symposium.
- [21] <http://www.cs.yorku.ca/~ivy/ve/>, The VE User Manual.