Evaluation of Gaze Tracking Technology for Social Interaction in Virtual Environments

Jan Ciger[*] Bruno Herbelin[†] Daniel Thalmann[‡]
VRlab, Swiss Feredal Institute of Technology (EPFL), Lausanne
http://vrlab.epfl.ch/

Abstract

This paper presents a study of the application of eye tracking technology in the context of social interaction in a virtual environment. We evaluate the reliability and precision of gaze tracking in two different virtual reality applications. In spite of the known drawbacks, the technology still has a potential for interaction with virtual humans.

Keywords: Eye tracking, attention evaluation, human-computer interaction, social

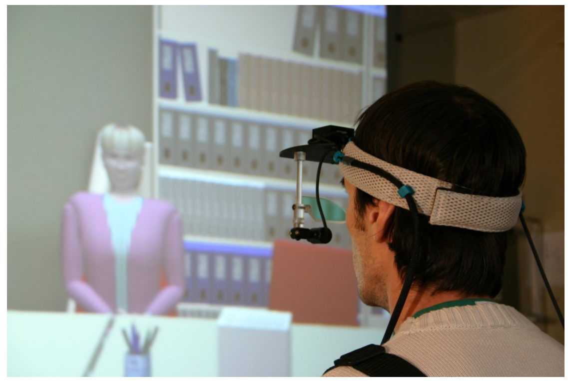

Figure 1: Eye tracker application

1 Introduction

Non-verbal communication is often used in social interactions. However, this additional information channel is rarely used in Human-Computer Interaction (HCI) as it requires complex gesture recognition and context relational models. Gaze tracking may propose an interesting compromise as it covers both aspects while being technologically affordable.

The purpose of this study is to experiment with gaze tracking in context of Virtual Reality immersion figuring a social situation. The objective is to evaluate the limitations of a head-mounted eye tracking device and to outline design issues for the development of interaction paradigms with Virtual Humans (VH).

A quick overview of eye tracking technologies and applications provided us with a sufficient knowledge to start this development with care. We proceeded with experiments to estimate the reliability and the precision of the gaze tracking. Finally, we developed two test applications for collaboration with virtual agents and analysis of human-VH interaction.

2 State of the Art

Eye tracking consist in following the eyes movements and computing the gaze direction in order to integrate this information into a computerized system. According to the very complete analysis made by Yang et al. [1], research in this field really started in the beginning of the 90s. Since then, the technology became more accurate, less cumbersome, and is today available as commercial products. In parallel, the understanding and the modeling of the gaze behavior improved, together with an extension to wide application do-

[*]jan.ciger@epfl.ch
[†]bruno.herbelin@epfl.ch
[‡]daniel.thalmann@epfl.ch

mains; psychiatry, cognitive science, behavioral analysis, medicine and Human-Computer Interaction (HCI).

Although electro-oculography (EOG) can potentially provide the gaze direction [2], computer vision systems, especially with infrared illumination, have much better results. Latest developments achieve acceptable precision and stability by using purely mathematical [3, 4] or connectionist [5, 6] approaches. Thanks to these recent improvements, the setup also became more flexible and can be adapted for desktop, large projection screen or Head Mounted Display [7]. However, large head movements are still impossible, and many system place the camera on the head.

In the context of our study, we focus exclusively on HCI applications. According to a discussion we had with Tom Schnell [8], they abandoned the idea of controlling items in aircraft cockpits using the eyes as the workload "was higher when using gaze-based control than when using old fashioned knobs". Indeed, there are many drawbacks that have to be avoided when using gaze for interaction:

- The "Midas touch": what the user is looking at is not necessarily what he wants to interact with. Although some study try to integrate statistical analysis to better recognize the user's focus of attention [9], using exclusively eyes is not natural and mainly used for assisting disabled people.

- Fatigue: voluntary and precise control of the gaze is tiring and a *passive* use of gaze is preferable to the *active* control: as Krapicher et al [2] propose, eye movements should be "evaluated in the background, without the user noticing any effects or system reactions".

- Perpetual motion: while voluntary eye saccades (1 to 40° of the visual angle) corresponds to the visual search, micro-saccades ($< 1^\circ$) still occur when the eye is focussing on a target. Although we know that the vision is suspended during saccades or eye blink, only few systems are able to really distinguish the fixation phases [10].

Gaze input is more appropriate for multimodal interaction. According to Kaur et al. [4], the combination with speech is very convenient as both can be well synchronized. Many other multimodal systems were experimented with, but the work of Tsui et al [6] is one of the most complete as it integrates five input modes into a fuzzy expert system, relying on gaze to resolve conflict when necessary.

Figure 2: Head mounted eye tracker with a 6 DOF sensor.

3 Hardware setup and computation of gaze vector

Most of our virtual reality applications are done with the user standing in front of a large back-projection screen with active stereo capability (fig. 1). The movement of the user is usually tracked using magnetic trackers (wireless MotionStar system from Ascension). In order to make the eye-only tracking usable for full gaze tracking in such setup, we added a 6-DOF magnetic sensor from the MotionStar system to the head-mounted assembly (see figure 2). This allows us to track not only the eye orientation but also the head position and orientation in the workspace (see fig. 3).

The gaze direction and target in the virtual scene is calculated out of several sets of primary data:

- ex, ey from the VisionTrak. The eye tracking system represents the field of view as a $512x512px$ square and returns the coordinates in this coordinate system.

- hx, hy, hz is a position from the MotionStar magnetic sensor. The position is ab-

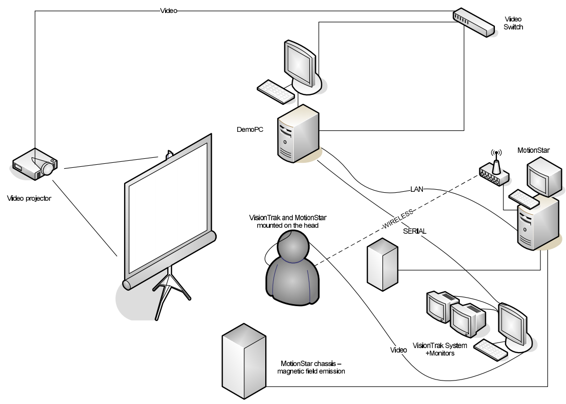

Figure 3: Experimental setup

solute, in the reference frame of the work space.

- x, y, z, w is an orientation returned from the magnetic sensor as a quaternion in the work space frame. Unit quaternion (no rotation) represents the user looking directly into the screen.

The gaze target is calculated as an intersection of the gaze vector with the screen. This is done in four steps.

First, the gaze vector in the eye reference frame is computed using a simple formula:

$$\overrightarrow{d_{eye}} = \begin{pmatrix} ex - 256 \\ 256 - ey \\ -475 \end{pmatrix} \tag{1}$$

We consider ex, ey as coordinates in the eye tracker frame, moving the origin to the center. The z component is calculated using the measured opening angle and the distance to the screen. The angle was measured as $56.6°$, giving $-475px$ as a distance to the screen in eye tracker frame. The z component is negative, because the vector is oriented toward the screen.

Second, the transformation of the gaze vector from the eye tracker frame to the global MotionStar (work space) frame takes into account the rotation of the user's head. The rotation matrix is built from the quaternion received from the MotionStar system and the computed gaze vector from equation 1 is multiplied by its inverse. We can take advantage of the orthogonality property of rotation matrices to simplify the inverse calculation resulting in equation:

$$\overrightarrow{d_{head}} = hrot^T \times \overrightarrow{d_{eye}} \tag{2}$$

An additionnal correction of the gaze vector by the calibration matrix is necessary, because there are differences in the way how the VisionTrak device is worn by different users. The calibration matrix is computed during the calibration procedure. The correction is done according to the formula:

$$\overrightarrow{d_{adj}} = hrinit \times \overrightarrow{d_{head}} \tag{3}$$

Finally, the intersection of the gaze vector with the screen is computed as a solution of the following system of linear equations ($hpos$ is measured head position):

$$hpos + t \cdot \overrightarrow{d_{adj}} = (x, y, 0) \tag{4}$$

The intersection point $(x, y, 0)$ in the work space coordinates is used by the application to compute the coordinates in the camera frame and conversely in the 3D world.

4 Experimental assessment of the tracker accuracy

In order to assess the performance of the combined tracker (eye tracker together with the magnetic system) two tests were performed. We did a reliability test and an precision test. The results of these two tests represent rather well what is possible with the current technology.

4.1 Reliability test

The reliability of the combined tracker was tested on multiple subjects (cca 20) using the following test procedure:

1. The test subject sits down on a chair in front of the screen. The VisionTrak device is mounted on his/her head and adjusted over the eye. The eye should appear in the center of the eye monitor.

2. The person conducting the test starts the data acquisition software, our calibration tool and controls the calibration procedure. The test person only has to look at a blue point projected on the screen.

Diameter	40cm	20cm	10cm	5cm
Active cursor	79%	62.5%	33%	4%
Disabled cursor	74%	59%	27%	2%
Standard deviation	4%	5%	7%	1%

Figure 4: Results of the reliability test

3. Once the device is calibrated, a dry test is carried out in order to verify the calibration of the device.

4. The ”Accuracy test” of the calibration tool is started. The test subject has to gaze successively at the centre of four differently sized circles that are shown in different zones on the projection screen.

We measure the percentage of the time during which the calculated gaze point stayed in the center of the circle. The procedure is performed twice, first with the cursor indicating the calculated gaze position and then without the cursor. This allows us to assess the effect of the visual feedback and the subject consciously “driving” the calculated gaze point to the indicated position.

The results of the reliability test are summarized in the table 4. The table shows two facts. First, the presence of the visual feedback in the form of cursor has an effect, however not very large. The cursor may still be useful tool to improve the accuracy. Second, the standard deviation is low, the results are reliable with good repeatability.

4.2 Precision test

The accuracy of the combined tracker depends on many factors. Both devices introduce their own inaccuracies – e.g. angle computation from very small motion of the eye, non-homogeneous magnetic field, etc. There are also subjective factors affecting accuracy – exact placement of the device on the user's head is different every time, the device changes position (slips) while in use, concentration and overall fatigue of the user also has an effect.

We used a test pattern to assess the accuracy of the combined tracker. The pattern consists of alternating yellow and blue squares and the test subject is asked to focus on every blue square for five seconds. The resulting data are represented as clusters of red dots and overlayed on top of the test pattern (fig. 5). The red circles mark error larger than 20cm, blue denotes error in the range 10–20cm and green color denotes error less than 10cm. Out of 25 samples, there are 10 with error larger than 20cm, 7 with an error between 10 and 20cm and 8 with an error less than 10cm.

We can conclude several things from the data resulting from the two tests:

- Object cca 40cm large or larger on the screen can be hit with the 80% probability on average.
- Object smaller than 10cm has only 27% chance of being hit.
- The tracker accuracy is larger in the center of the screen and decreases towards the edges.
- The random errors are rare, the data are tightly clustered.
- The error is not uniform (more accurate in the upper right half of the pattern than in the lower left half).

These results indicate that the eye tracking system is a viable option for applications where medium resolution is acceptable. It is comparable to other commonly used pointing hardware (e.g.magnetic trackers, 3D mice).

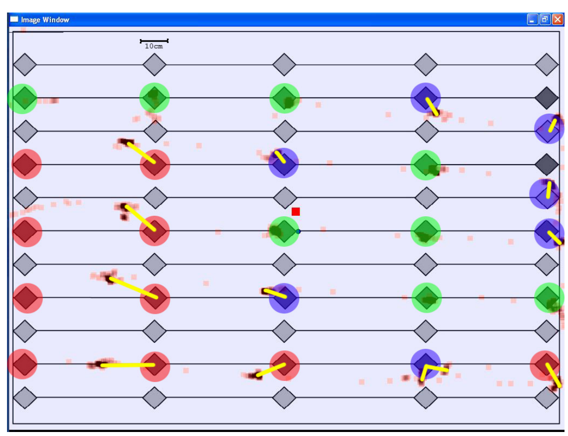

Figure 5: Test pattern with overlayed tracking data

5 Social Interaction with gaze in VR

According to the known drawbacks and to the limitations of our eye tracking system, we evaluated the use of gaze for the interaction with virtual humans in two different scenarios.

5.1 Multimodal control of a virtual character with gaze and selection device

Gaze tracking is difficult to use directly, as mentioned in the previous sections. To work around this problem we combined the gaze tracking together with another input device (gamepad) into a multimodal system. This combination reduces the user's fatigue and allows to track the gaze only when the user really desires to.

We implemented a simple test application with two virtual characters. The user is able to control the characters in two different modes:

- direct control, character following the gaze ("walk where I look!")
- indirect control, character receiving orders ("go there!", the value of "there" being determined by the gaze)

The gaze tracking is used to pick one of the two characters on the screen and then to control him according to the mode. In the direct mode, the selected character just follows the user's gaze. This mode is useful to precisely move the virtual character in the scene, however it requires concentration from the user and it is very tiring. The indirect mode is more practical. The selected virtual character is given orders where to move and the user's attention/concentration is required only for a short moment – to pick the right target position and to activate the order using the gamepad.

There were several insights gained from this application. First, the resolution of the tracker is sufficient to select reliably the virtual character, but it is difficult to pick accurately smaller objects (object farther away, small features). This has to be considered during the application development. If continuous attention of the user is required, the application is very tiring and such designs should be clearly avoided.

5.2 Visual attention analysis in public speaking situation

Public speaking is a common situation where gaze has a great impact on the social interaction. For the needs of the cognitive and behavioral therapy of social phobic patients, we perform virtual reality exposures (VRE) with a virtual scene figuring an assembly of 7 persons [11]. The patient is asked to make a speech in front of this jury and to correct his involuntary gaze avoidance behavior. Eye tracking technology is used here to make a statistical analysis of his attention focus, comparing the time spent looking at the persons and more specifically looking them in the eyes.

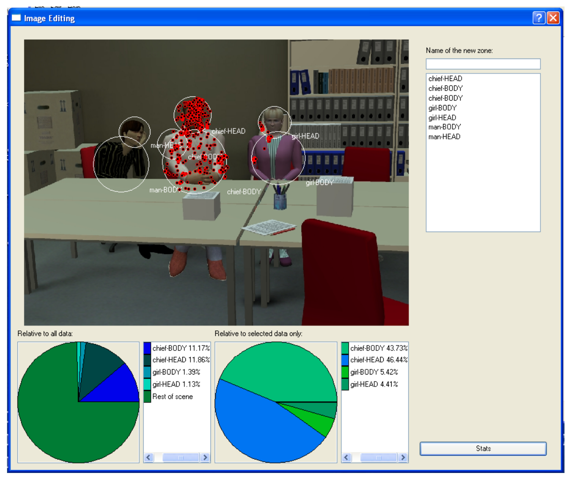

Figure 6: Attention analysis

Within the context of dialogues in a small group, the eye tracking offered a sufficient precision to distinguish not only who the user is speaking to but also which part of the body he is looking at. However, improving the precision up to the part of the face (e.g. the eyes) is only possible if the user gets closer to the virtual character.

For the needs of this experiment, a graphical tool was designed with three main features: record/replay the gaze point, display in overlay the gaze trace as clusters of colored points, compute statistics on the time spent in user-designed zones (see fig.6). The therapist started to use this tool to keep records of patients' performance and to observe their improvements.

6 Conclusion

We conducted experiments with a custom eye tracking setup in front of the wide projection screen. The reliability and precisions obtained seem to be appropriate for natural interaction with virtual humans.We verified that the gaze can be used to select an agent and to give him orders in a simple manner. In the context of social interaction (when close enough to address a virtual character), the precision is sufficient to identify body part the user is looking at and to collect statistics on his behavior.

Furthermore, we would like to utilize the collected statistical data at runtime, enabling natural reactions of the agents based on the created semantic attention map. We are now working on a 3D picking to gather automatically the gaze information, specifically for the observed body parts.Our laboratory is also interested in the believable virtual human simulation based on perception. Optimization of the rendering quality and animation based on attention analysis can help to drastically reduce the amount of needed resources and improve the overall responsiveness of the system.

Acknowledgements

Authors would like to thank the students Luzian Wild and Sebastien Dupré for their excellent work on the tracker and the experiments.

References

[1] Guang-Zhong Yang, Laura Dempere-Marco, Xiao-Peng Hu, and Anthony Rowe. Visual search: psychophysical models and practical applications. *Image and Vision Computing*, 20(4):273–287, 2002.

[2] Christian Krapichler, Michael Haubner, Rolf Engelbrecht, and Karl-Hans Englmeier. VR interaction techniques for medical imaging applications. *Computer Methods and Programs in Biomedicine*, 56(1):65–74, April 1998.

[3] Chern-Sheng Lin, Chia-Chin Huan, Chao-Ning Chan, Mau-Shiun Yeh, and C.-C.Chuang-Chien Chiu. Design of a computer game using an eye-tracking device for eye's activity rehabilitation. *Optics and Lasers in Engineering*, 42(1):91–108, July 2004.

[4] Manpreet Kaur, Marilyn Tremaine, Ning Huang, Joseph Wilder, Zoran Gacovski, Frans Flippo, and Chandra Sekhar Mantravadi. Where is "it"? event synchronization in gaze-speech input systems. In *5th international conference on Multimodal interfaces, ICMI2003*, November 2003.

[5] Zhiwei Zhu and Qiang Ji. Eye and gaze tracking for interactive graphic display. *Machine Vision and Applications*, 15(3):139–148, July 2004.

[6] K.C. Tsui, B. Azvine, D. Djian, C. Voudouris, and L.Q. Xu. Intelligent multi-modal systems. *BT Technology Journal*, 16(3):134–144, July 1998.

[7] Chern-Sheng Lin. An eye behavior measuring device for VR system. *Optics and Lasers in Engineering*, 38(6):333–359, December 2002.

[8] Matthew Rizzo, Johnson Moon, Mark Wilkinson, Kirk Bateman, Julie Jermeland, and Tom Schnell. Ocular search of simulated roadway displays in drivers with constricted visual fields. *Journal of Vision*, 2(7):162, 2002.

[9] Dario D. Salvucci and John R. Anderson. Intelligent gaze-added interfaces. In *CHI 2000 Conference on Human Factors in Computing Systems*, April 2000.

[10] Linda E. Sibert and Robert J.K. Jacob. Evaluation of eye gaze interaction. In *CHI 2000 Conference on Human Factors in Computing Systems*, April 2000.

[11] Bruno Herbelin, Francoise Riquier, Frederic Vexo, and Daniel Thalmann. Virtual reality in cognitive behavioral therapy: a study on social anxiety disorder. In *8th International Conference on Virtual Systems and Multimedia, VSMM02*, 2002.