

Eye Contact in Leisure Video Conferencing

Annick Van der Hoest & Dr. Simon McCallum
Gjøvik University College, Norway

19 November 2012

Abstract

This paper presents systems which enable eye contact in leisure video conferencing. The proposed solutions include a software based and a hardware based approach. Two experiments were conducted to test the performance of the systems by measuring the users' sense of presence. The results showed that both solutions succeeded in establishing eye contact but there was no direct indication of an increase in presence with respect to the conventional setup. The hardware solution induced a greater sense of presence and established a greater and more natural eye contact perception than the software solution.

1 Introduction

The concept of presence refers to the sensation of being in a physically distinct location [12]. A sense of presence in a video conferencing session, refers to experiencing togetherness with the remote user rather than being aware of the physical separation. In Knudsen's extensive research in the field of presence [1], the objective was to improve the quality of long distance communication by increasing the sense of presence. This is followed in this work and thus the focus was on enhancing users' sense of presence. According to Witmer & Singer [15] a greater involvement and immersion lead to a greater sense of presence. There are three factors that determine the sense of presence: the individual preconditions present, the characteristics of the content delivered and the sensory environment (i.e. technologies used) [4]. While preconditions are difficult to influence, it is possible to influence the content and the sensory environment. In the field of gaming the content can be managed, however in video conferencing the content is controlled by the users. In both cases we can try to improve presence by altering the sensory environment. This work aims at improving presence in leisure video conferencing.

In conventional video conferencing the camera is on top of the screen, preventing eye contact. This creates a problem because eye contact helps evaluate virtues such as attentiveness, competence and credibility of the speaker [8] and assists with recognising emotions [14]. Eye contact plays an important role in communication as it conveys additional nonverbal information. To improve leisure video conferencing we choose to focus on eye contact in our attempt to improve presence.

This paper was presented at the NIK-2012 conference; see <http://www.nik.no/>.

2 Related Work

Eye contact solutions for video conferencing can be roughly divided into hardware and software. A simplistic hardware approach is to use large screens and have the user sit far away. For this to be effective the user's gaze and the camera location must be separated by less than 6 seconds of arc (vernier acuity [9]). If the angle is small the users do not detect the disparity.

A common hardware solution is based on teleprompter technology, consisting of a display, a camera and a beam splitter. The camera and the display are located at a 90° disparity with the beam splitter in between at a 45° angle to both the display and the camera. The beam splitter deflects the beam emitted from the display, while allowing the light reflected from the user to pass through to the camera. Thus, the camera captures the image of the user through the beam splitter, whereas the user perceives the displayed image as in front of the camera due to the reflection of the beam splitter.

Another hardware approach is using see-through displays, where the camera is behind the display, capturing the image of the user through the display. Shiwa and Ishibashi [11] used switchable liquid crystal displays. Tan et al. [13] used wavelength multiplexing, where the light projected on the display consists of visible wavelengths outside the sensitivity of the camera.

Software approaches are more novel and have a wider application because there is no need for specific hardware. Software solutions use image processing techniques to create a perception of eye contact. A common approach is to generate a view from a virtual camera that is placed at the centre of the screen or at the position of the displayed face of the remote user [16, 2, 3, 5]. The scene is captured by multiple cameras from which 3D information is extracted to model the scene. The scene can then be displayed as if it was captured from a different perspective.

Examples of multiple camera approaches include a system with a pair of calibrated stereo cameras proposed by Yang & Zhang [16]. They use a personalised face model to map the correspondence points necessary for the merging of the camera views, in order to deal with the high disparity of the cameras. Civit & Montserrat [2] proposed a four camera system to decrease the disparity in the images, which enhanced the search for reliable correspondence points. Dumont et al. [3] developed a six camera system, where all computations are done dynamically and in real time using the GPU. The Fraunhofer Heinrich-Hertz Institute [5] developed the Virtual Eye Contact Engine. These systems use multiple cameras, up to six for the Dumont system. This is unrealistic for most casual users.

Single camera approaches include GazeMaster, a system developed by Gemmell & Zhu [6], which segments the eyes and texture maps the face onto a head model with synthetic eyes. However to reach real time performance the tracking of the eyes proved to be too difficult with a single camera (J. Gemmell, personal communication, Januari 24, 2012). Another one camera system was suggested by Jerald & Daily [7]. The camera is mounted on the side of the display and users click on specific points on their eyes to allow for eye tracking. This user interaction is the drawback of their system.

3 Methodology

This work uses a single camera and template matching software to enable eye contact. The system detects and then tracks the eyes, continuously correcting the gaze. This extends the Real Time Eye Tracking and Blink Detection with OpenCV software developed by Nash Ruddin [10]. This included a blink detector to automatically locate an eye and an eye tracker. The blink detector's template matching system was extended to locate both eyes and gaze correction functionality was added. The gaze correction captures a template of the eyes looking at the camera, then blends the template with the current frame.

The software solution was tested against a hardware approach, by constructing a low-budget hardware solution based on teleprompter technology. It consisted of a wooden box with a piece of glass placed at a 45° angle, functioning as a beam splitter. A user looking in the box looks straight into the camera, while perceiving the display in the reflection of the glass.

Experiments were designed and conducted to measure users' sense of presence while engaged in video conferencing. The main experiment used 25 participants conducting a video chat with a member of the research team either in Skype with the teleprompter or with custom video conferencing software with gaze correction. A questionnaire addressing presence was filled out afterwards.

4 Results

The most substantial experimental results are shown as bar graphs in Figures 1 and 2 corresponding to questions addressing immersion and eye contact respectively. For each question a χ^2 test was performed. Participants testing the hardware solution were more engaged (confidence 99%) and less distracted (confidence 95%) from the conversation than the participants testing the software solution.

Eye contact induced by the hardware solution is more natural (confidence 90%) and less distracting (confidence 95%) than the eye contact induced by the software solution. Figure 2 shows that eye contact perception in general was good for both solutions. The χ^2 test was therefore performed to distinguish between a good perception (i.e. rated by number 4) and a very good perception (i.e. rated by number 5). Results show that the hardware solution induced a greater eye contact perception than the software solution (confidence level of 95%).

5 Conclusion

The aim of this work was to develop solutions suitable for leisure video conferencing that allow for eye contact.

The experiment aimed at comparing the software to the hardware solution. From the results of the experiment it can be concluded that the hardware solution induced a greater sense of presence and established a greater and more natural eye contact perception. Based on the experimental results and the positive feedback from the participants both systems show potential. It is our belief that with improvements the solutions could function well as eye contact enabling systems for leisure video conferencing.

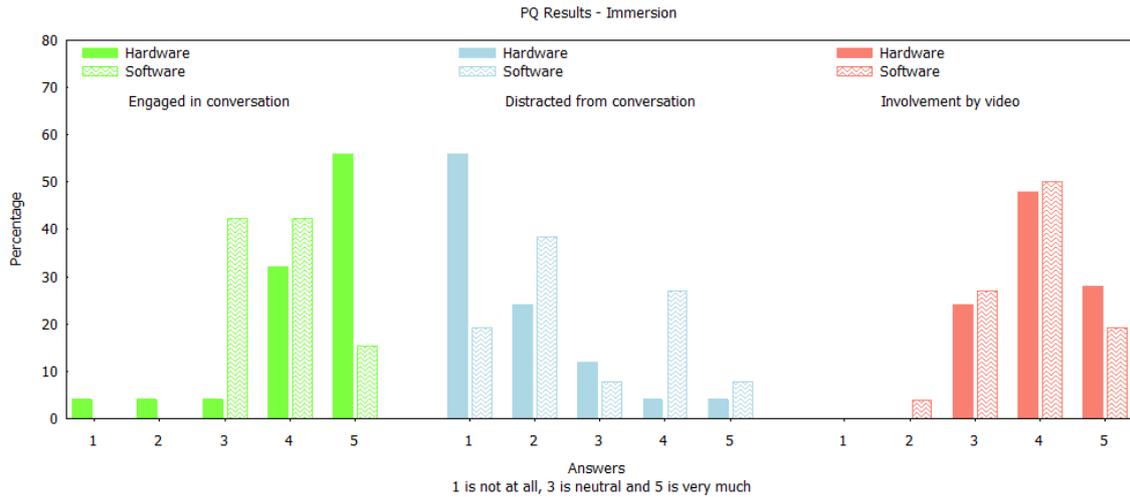


Figure 1: Presence results, comparing the sense of immersion when using the hardware and software solutions.

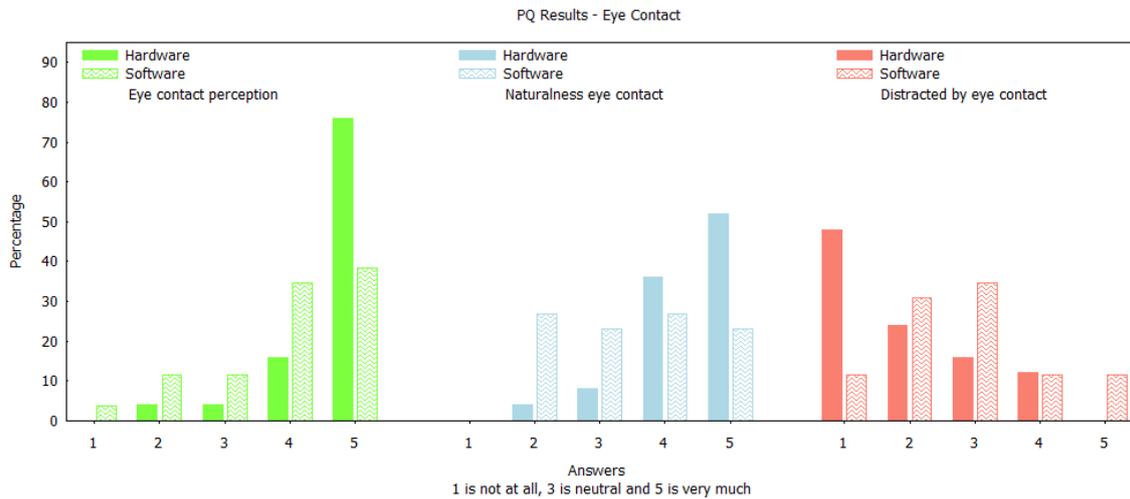


Figure 2: Presence results, comparing the sense of eye contact when using the hardware and software solutions.

6 Future Work

The main technical improvement required relates to tracking the eyes more accurately. Using multiple cameras would also provide more information on the scene. External cameras are becoming cheaper and more common as are multiple integrated cameras (for example the LG Optimus 3D smartphone, Microsoft Kinect and Fujitsu LifeBook AH572 laptop). A feature that should be added to the software is gaze localization such that when the user does not look at the image of the remote user the gaze is not corrected. This could be accomplished by making use of a stored template of the user looking at the image of the remote user. For hardware, developing a tablet protective case with integrated teleprompter functionality would provide a very user friendly solution.

References

- [1] *Presence Production*. PhD thesis, 2004.
- [2] J. Civit and T. Montserrat. Eye Gaze Correction to Guarantee Eye Contact in Videoconferencing. *IEEE Latin America Transactions*, 7(3):405–409, JUL 2009. 13th Conference on Software Engineering and Databases, Gijon, SPAIN, 2008.
- [3] M. Dumont, S. Rogmans, S. Maesen, and P. Bekaert. Optimized two-party video chat with restored eye contact using graphics hardware, 2009.
- [4] N. Enlund. The production of presence - distance techniques in education, publishing and art. In *ACS'2000 Proceedings*, 2000.
- [5] Fraunhofer. Virtual eye contact engine.
<http://www.hhi.fraunhofer.de/en/departments/image-processing/applications/virtual-eye-contact-engine/>, 2011.
- [6] J. Gemmell and D. Zhu. Implementing gaze-corrected videoconferencing. In *Proceedings of CIIT*, 2002.
- [7] J. Jerald and M. Daily. Eye gaze correction for videoconferencing. In *ETRA*, pages 77–81, 2002.
- [8] C. L. Kleinke. Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1):78–100, 1986.
- [9] A. Roorda. *Encyclopedia of Imaging Science and Technology*, chapter Human visual system - image formation. 2002.
- [10] N. Ruddin. Real time eye tracking and blink detection with opencv (software).
<http://nashruddin.com/Real-Time-Eye-Tracking-and-Blink-Detection>, 2009.
- [11] S. Shiwa and M. Ishibashi. A large-screen visual telecommunication device enabling eye contact. *SID Dig.*, 22:327–328, 1991.
- [12] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *JOURNAL OF COMMUNICATION*, 42:73–93, 1992.
- [13] K.-H. Tan, I. N. Robinson, B. Culbertson, and J. Apostolopoulos. ConnectBoard: Enabling Genuine Eye Contact and Accurate Gaze in Remote Collaboration. *IEEE Transactions on Multimedia*, 13(3):466–473, JUN 2011.
- [14] R. van Eijk, A. Kuijsters, K. Dijkstra, and W. IJsselsteijn. Human sensitivity to eye contact in 2d and 3d videoconferencing. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 76 –81, june 2010.
- [15] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240, June 1998.
- [16] R. Yang and Z. Zhang. Eye gaze correction with stereovision for videoconferencing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):956–960, JUL 2004.