

# Exact Eye Contact with Virtual Humans

Andrei State

Department of Computer Science, University of North Carolina at Chapel Hill  
CB 3175 Sitterson Hall, Chapel Hill, North Carolina 27599, USA

and

InnerOptic Technology Inc., P. O. Box 824  
Chapel Hill, NC 27514-0824, USA

andrei@cs.unc.edu

**Abstract.** This paper describes a simple yet effective method for achieving accurate, believable eye contact between humans and computer-generated characters, which to the author's knowledge is demonstrated here for the first time. A prototype system provides a high-fidelity stereoscopic head-tracked virtual environment, within which the user can engage in eye contact with a near-photorealistic virtual human model. The system does not require eye tracking. The paper describes design and implementation details, and reports qualitative positive feedback from initial testers.

**Keywords:** Eye Contact, Virtual Human, Gaze Direction, Interactive Virtual Environment.

## 1 Introduction and Related Work

Future Human-Computer Interface technologies will include virtual humans (VHs) capable of fluent, meaningful conversations and collaborations with humans [1]. The effectiveness of such interactions will depend in part on VHs' ability to appear and act like real people [2][3]. Eye contact is essential in human interaction [4] and therefore should be given careful consideration in the design of VHs and other virtual interaction partners. Believable eye contact with such entities will be most important in stereoscopic head-tracked virtual environments (SHVEs), whether using tracked head-mounted displays [5] or head-tracked "fish tank" techniques [6], since the view-point-matched stereoscopic imagery continually maintains the user's awareness of spatial relationships.

Beyond HCI and gaming, other areas that may benefit from accurate eye contact with virtual entities include robotics [7] and advertising, which is likely to conceive intrusive virtual characters that stare at us. Notwithstanding the latter, this paper presents simple yet effective techniques for exact eye contact (EEC) between a human user and a VH. The high accuracy is demonstrated in a precisely calibrated, stereoscopic head-tracked viewing environment. While the current implementation requires a head-mounted tracker, future embodiments may use un-encumbering tracking, such as vision-based head pose recovery. It is important to note that the technique described here does not require pupil tracking; it uses only head pose, which can gener-

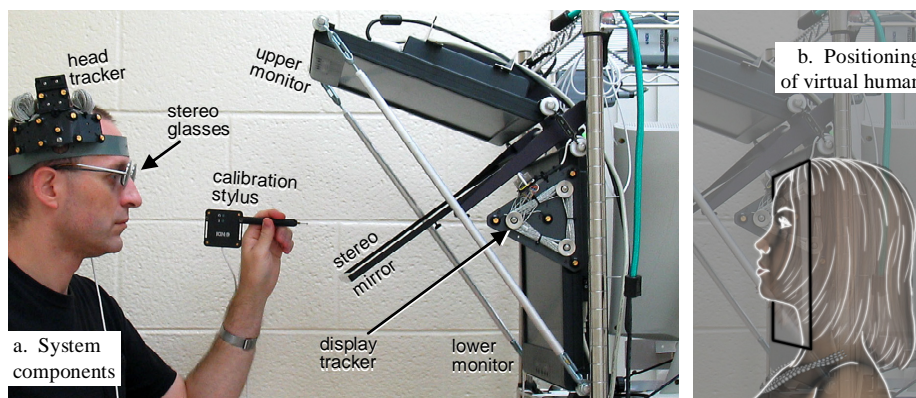
ally be obtained less intrusively, with higher reliability, and from a greater distance away than camera-based pupil tracking. An additional pupil tracker is not required unless the system must know the user's gaze direction, for example in order to record user behavior in training applications [3].

## 2 A Prototype System for Exact Eye Contact

The EEC prototype uses a fish tank SHVE (Fig. 1) consisting of a personal computer and the following main components:

- a Planar Systems SD1710 (“Planar”) stereoscopic display with two 17” LCD monitors and a semi-transparent mirror that reflects the upper monitor’s image onto the lower monitor. The user wears linearly polarized glasses that restrict viewing of the lower monitor to the left eye and viewing of the upper monitor’s reflection to the right eye. The LCDs’ native resolution is 1280×1024. To improve alignment between the monitors, an adjustable stabilization frame was added (Fig. 1a).
- a sub-millimeter precision Northern Digital Optotrak Certus opto-electronic tracking system (“Certus”). Both the Planar and the user’s head are tracked by the Certus in all six degrees of freedom with clusters of infrared (IR) LEDs (11 on the head, 4 on the Planar). The advantage of tracking the display as in handheld augmented reality applications [8] is that both the display and the tracker can be moved with respect to each other while the system is running, for example, to improve LED visibility. The Certus also provides a calibration stylus for precise measurements.

To use the EEC prototype, the user dons the head tracker and first performs a simple, fast eye calibration procedure. This is followed by a live interaction phase, during which the user can engage in eye contact and interact with the VH.

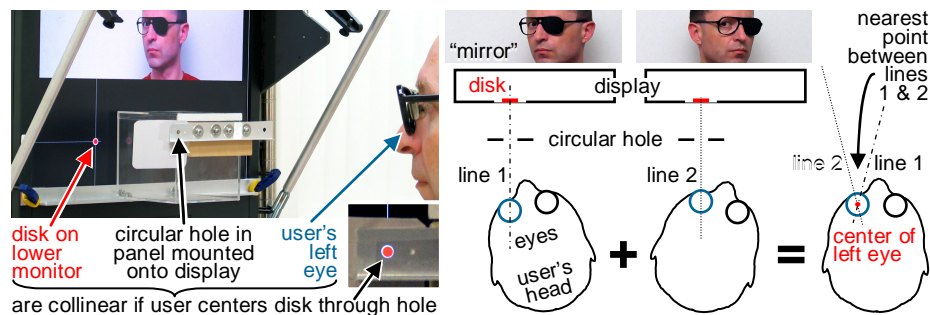


**Fig. 1.** Head-tracked virtual environment for eye contact with a virtual human uses a two-LCD stereoscopic display. The right-hand image shows the positioning of the life-size virtual human within the (rather small) display.

## 2.1 Projection Origin and Eye Calibration

In SHVEs, the calibration between the head tracker and the eyes is usually obtained from measurements such as the user’s inter-pupillary distance (IPD, measured with a pupillometer) [5], the location of the tracker on the user’s head, as well as from assumptions about the most suitable location of the projection origin inside the eye. Popular choices for the latter include the eye’s 1<sup>st</sup> nodal point [6], the entrance pupil [9], and the center of the eye [10]. The EEC prototype uses the eye center [10] because it is easy to calibrate and yields exact synthetic imagery in the center of the field of view regardless of the user’s gaze; a further advantage will be described in 2.2. However, the 1<sup>st</sup> nodal point and the entrance pupil are better approximations for the actual optics within the eye. Therefore, by rendering stereo images from the eye centers, i.e. from a few mm too far back, and thus with a slightly exaggerated separation, the EEC system deforms the stereoscopic field [11] ever so slightly. For higher accuracy, a pupil tracker could detect the user’s gaze directions, and assuming that the user converges onto the virtual object found along those directions, one could move the projection origins forward to the 1<sup>st</sup> nodal point, or all the way to the pupil.

**Calibration.** The eye calibration technique (Fig. 2) was inspired by previous methods [12][13] and modified for the additional display tracker. A small panel with a circular hole is temporarily mounted in front of the bottom LCD panel. Both the hole and the bottom LCD monitor are pre-calibrated (one-time only) to the Planar’s tracker with the Certus stylus. The eye calibration program shows a circular disk on the display. Using a “mirror image” of the user’s head as a guide, the user moves and orients his head to line up the disk through the hole, twice through each eye, under different head orientations. To avoid confusion, users wear frames with one eye masked off, as shown by the “mirror” guides at the top of Fig. 2. The program collects four line equations in head tracker coordinates. In pairs of two, these four lines define the eye centers at their intersections—or rather, at the closest points between them. The entire task takes 1-2 minutes except for inexperienced first-time users, which take longer mostly because they must receive and follow instructions.



**Fig. 2.** Eye calibration setup and sequence, shown here for the left eye only.

Since the current head tracker (Fig. 1a) does not guarantee repeatable positioning on the user’s head, the user should not remove it between calibration and the following interactive phase. User-specific head-conforming gear equipped with IR LEDs—

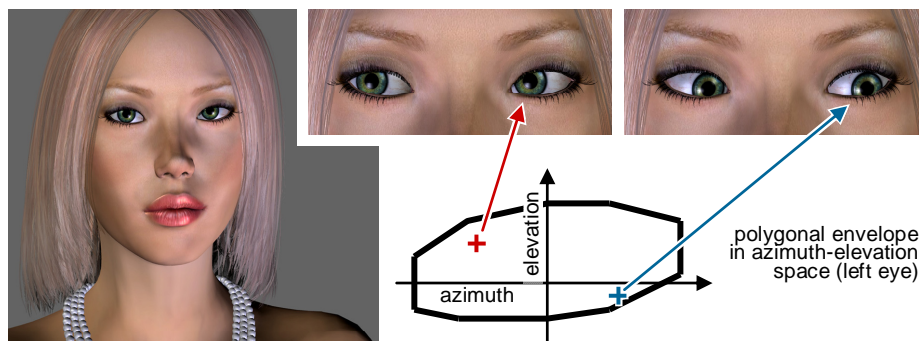
or with passive markers for camera-based head tracking—could remove this restriction and could thus reduce each user’s eye calibration to a one-time procedure.

## 2.2 Interactive Operation

During this phase, the system displays a life-size stereoscopic head-and-shoulders view of a VH as seen from the user’s eye centers. The latter are computed from the head tracker reports and from the eye calibration measurements described above.

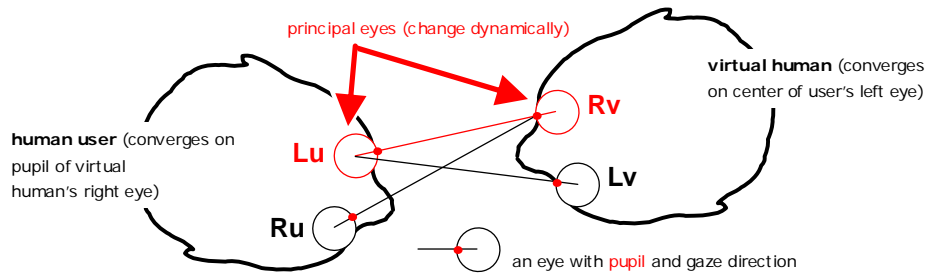
**Virtual Human Model.** The human model (Fig. 3, “Carina” from Handspan Studios, slightly reshaped and edited for fast rendering with the DAZ Studio modeler from DAZ Productions, Inc.), is displayed via OpenGL in separate views for the Planar’s two monitors, using off-center viewports; the model is near-photorealistic but simple enough for real-time rendering. The 2.6GHz dual-core Opteron PC equipped with an NVIDIA 7800GT graphics adapter achieves a frame rate of approximately 60Hz. Only head and shoulders could be fitted in life size within the limited virtual space of the small display (Fig. 1b). The model is completely static except for the eyeballs, which have separate rotation transformations for azimuth and elevation. Thus the eyes can be oriented toward a given target. A polygonal “occlusion envelope” in each eye’s azimuth-elevation space (Fig. 3, right) approximates the eyelid contour and helps detect if the eye “tries” to look at a target “through” skin or eyelids, which could cause the eye to roll up into the head, for example. Whenever either eye encounters such a situation, *both* eyes switch to looking straight ahead. The occlusion envelopes are not anatomically accurate; they simply match the 3D model. The eyes are displayed with high-detail textures for the iris as well as with transparent, specular corneas to catch reflected light (Fig. 3, center).

To minimize the display’s accommodation-convergence conflict, the most important components of the VH, its eyes, are positioned roughly at the LCD panel surface (Fig. 1b), with nose and chin protruding slightly in front of the display. The face looks straight ahead, while the rotating eyes seek out the head-tracked user’s eyes. They can alternately target the user’s left and right eyes, a hallmark of bad acting according to Sir Michael Caine [14], or they can fixate one eye (recommended by Caine).



**Fig. 3.** Near-photorealistic human model with orientable eyes and eyelid occlusion envelope. Left and top images are real-time screen captures from the EEC system. Note highly detailed eyes (model by Handspan Studios).

**Eye Contact Targets.** In the current EEC implementation, the targets for the VH’s eyes are the user’s eye centers; but in real life, people look at irises, not eye centers. Without a pupil tracker, the EEC system cannot know which way the user’s pupils point, so it cannot target them. Fig. 4 shows that only one eye from each person (**Lu** and **Rv** in Fig. 4) looks directly into one eye of the other person; these two “principal” eyes have their centers and pupils aligned on a single line, which is already guaranteed in the EEC system even though it uses eye centers as targets. Hence the VH’s principal eye **Rv** requires no correction. Its other, “non-principal” eye **Lv** is slightly mis-oriented; it points to the center of the human’s principal eye **Lu** instead of to the pupil (whereas the human’s non-principal eye **Ru** fixates **Rv**’s pupil, not its center). Humans have extraordinary acuity for assessing gaze direction [15], and in an extreme close-up situation—probably closer than the current prototype permits—a perceptive user might notice that the VH converges about 10mm behind the user’s pupil. Then again, it may not be possible to detect this slight error since in that situation the user is not directly looking at the slightly misaligned non-principal eye **Lv** but at **Rv**, the correctly oriented principal one. In other words, **Lv** does not appear in the center of either **Lu**’s or **Ru**’s fields of view. Still, **Lv** could be easily corrected if the EEC system were able to determine which of the VH’s eyes is in fact the principal one at a given moment. However, measuring this dynamically changing condition requires pupil tracking.



**Fig. 4.** A closer look at the geometry of eye contact in the EEC system.

**Dynamic Behavior.** As described so far, the EEC SVHE system computes fresh eye gaze targets and associated azimuth and elevation angles for its VH’s eyes for each new stereo frame pair presented to the head-tracked user. This overly simplistic approach results in rather bizarre, neurotic-seeming behavior. For example, as long as the gaze directions towards the user’s eyes fall within the VH’s occlusion envelopes, the VH will continually and instantaneously track the user with unnatural precision; or, if the user dwells at the envelope margins, the VH’s eyes will oscillate, constantly alternating between looking straight ahead and making eye contact with the user. These issues were eliminated by temporally filtering the azimuth and elevation angles over the most recent  $n$  frames (currently  $n=12$ , a fifth of a second at 60Hz). A simple box filter already yields much improvement: the VH’s eyes are stable and follow the user with some lag as he moves about. The lag also creates opportunities to engage in and disengage from eye contact, which more closely approximates human behavior.

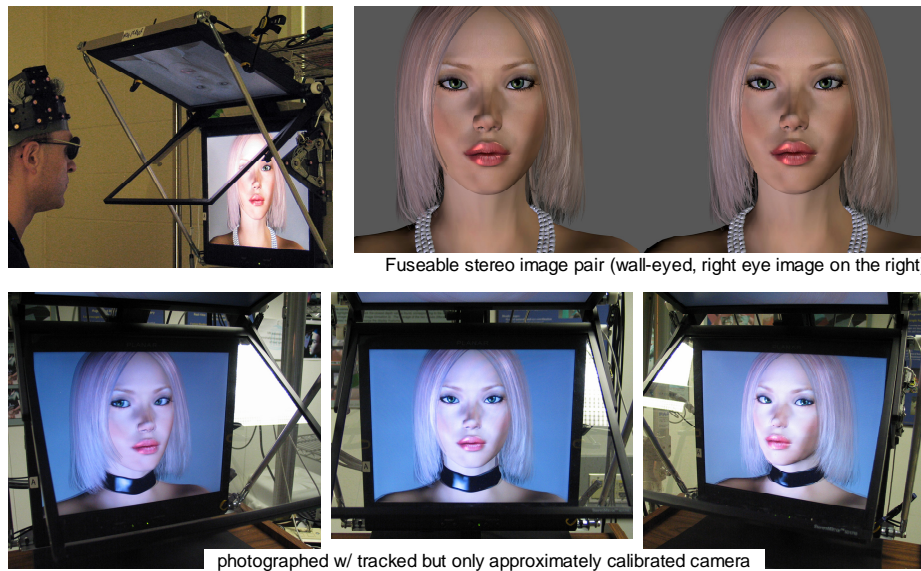


### 2.3 Initial User Experience

The EEC prototype described here is a very simple proof-of-concept system limited to gaze interaction with a VH. A formal study has not been conducted or designed yet, but the EEC system was demonstrated to several experts in the field, as well as to several non-expert acquaintances of the author. All judged the illusion to be quite convincing and the eye contact to be realistic (though none felt *intimidated*). Most were able to tell when the VH was switching targets from one eye to another. One expert in graphics and vision declared that until experiencing the EEC system, he never thought much of stereo displays in general. All testers had their eye centers calibrated as described in Section 2.1 above; for some, an additional sanity check was performed and the distance between their eye centers, as resulting from calibration, was found to be within no more than 1-2mm of their infinity IPD, measured with a pupillometer.

The author also implemented a simple additional interaction in the EEC system: the Certus calibration stylus can act as a light source for the VH, which is programmed to fixate it if the stylus is in range (otherwise it looks for the user as already described). Another simple feature, a software switch that toggles head tracking, turns the system into a simple but compelling educational tool that lets even naïve users instantly comprehend the difference between stereoscopy with and without head tracking.

A video segment accompanies this paper. It shows the EEC system in use, but it obviously cannot convey the live SVHE's sense of presence (Fig. 5). The video can be accessed at <http://www.cs.unc.edu/~us/EEC/>.



**Fig. 5.** As the user moves about the display, the virtual human acquires and attempts to maintain eye contact. The bottom set was photographed through the stereo mirror and shows both eyes' views superimposed (the mirror reflects the right eye view from the top LCD monitor).

### 3 Conclusions

The EEC method described here is simple and easy to implement. It can replace camera-based pupil tracking where gaze direction knowledge is not required. After a brief initial calibration step, the system accurately tracks a user's eyes in a virtual environment with only real-time head pose tracking as input, which can be obtained from any type of tracker, including a non-intrusive, completely untethered computer vision system. As for the display, a much cheaper CRT-based stereo display with active shutter glasses could replace the Planar while still providing high image quality.

The author attributes the positive initial feedback to the extreme accuracy of the optoelectronic tracker, the precise eye calibration, and the high-quality, near photorealistic stereoscopic rendering presented at high frame rates.

The EEC system also confirms that eye centers are suitable projection origins and extends the eye center approximation to small-scale desktop displays; it further shows that eye centers are appropriate gaze targets for eye contact simulations.

### 4 Future Work

Prismatic distortions introduced by users' eyeglasses move the eye centers calculated by the calibration. Nearsighted users' eye centers move forward, farsighted users' backward. Wearing eyeglasses during both calibration and live interaction compensates partially, but the nonlinear nature of the distortion requires further investigation.

The virtual human's appearance leaves vast room for enhancements. To mention only the eye area, anatomical accuracy could be improved through blinking, upper and lower eyelid motion, as well as pupil enlargement and contraction, all of which could be easily synthesized in real time on a modern computer, for example, by means of OpenGL vertex shaders.

An untethered camera-based head tracker would greatly improve user comfort. While head pose tracking alone already provides believable eye contact, an additional pupil tracker would enable correction of the remaining gaze direction error (in the non-principal eye, see 2.2), as well as lead to much additional experimentation. Specifically, with the ability to measure user's gaze behavior, a controlled user study becomes feasible, even though its design would be quite challenging in the author's opinion. Furthermore, the integrated pupil tracker could open the door to interdisciplinary research, for example in the realm of behavioral psychology.

**Acknowledgments.** The author thanks the testers for their insights and suggestions. They include Anna Bulysheva, Charles Burke (MD), Henry Fuchs, Martina Gargard, Bil Hays, Kurtis Keller, Sharif Razzaque, Herman Towles, Greg Welch, and Hua Yang. Anna Bulysheva and Herman Towles reviewed draft versions of the paper. Tabitha Peck prepared the initial version of the GLUT-based framework used by the EEC prototype. Hua Yang prepared interface software for the Certus. The EEC real-time renderer uses libwave by Dave Pape, ported to Windows with help from Hua Yang. Partial funding for this work was provided by the National Institutes of Health (1 R01 CA101186-01A2).

## References

1. Takacs, B., Kiss, B.: The virtual human interface: A photorealistic digital human. *IEEE Computer Graphics and Applications* 23(5), 38--45 (2003).
2. Badler, N.I., Phillips, C.B., and Webber, B.L.: *Simulating Humans: Computer Graphics, Animation, and Control*. Oxford Univ. Press (1993).
3. Raj, A.B., Johnsen, K., Dickerson, R.F., Lok, B.C., Cohen, M.S., Duerson, M., Pauly, R.R., Stevens, A.O., Wagner, P., Lind, D.S.: Comparing Interpersonal Interactions with a Virtual Human to Those with a Real Human. *IEEE Transactions on Visualization and Computer Graphics* 13(3), 443--457 (2007).
4. Argyle, M., Cook, M.: *Gaze and Mutual Gaze*. Cambridge University Press (1976).
5. Meehan, M., Razzaque, S., Whitton, M., Brooks, F.: Effects of Latency on Presence in Stressful Virtual Environments. *Proceedings of IEEE Virtual Reality 2003*, IEEE Computer Society, 141--148 (2003).
6. Deering, M.: High Resolution Virtual Reality. In: J. J. Thomas (ed.) *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '92*, ACM Press, New York, NY. *Computer Graphics*, 26(2), 195--202 (1992).
7. Takayama, A., Sugimoto, Y., Okuie, A., Suzuki T., Kato, K.: Virtual human with regard to physical contact and eye contact. In: Kishino, F., Kitamura, Y., Kato, H., Nagata, N. (eds.) *Entertainment Computing 2005*. LNCS, vol. 3711, 268--278 (2005).
8. Billingham, M., Henrysson, A.: Research Directions in Handheld AR. *Int. J. of Virtual Reality* 5(2), 51--58 (2006).
9. Rolland, J.P., Burbeck, C.A., Gibson, W., Ariely, D.: Towards Quantifying Depth and Size Perception in 3D Virtual Environments. *Presence: Teleoperators and Virtual Environments* 4(1), 24--48 (1995).
10. Holloway, R.: Registration Error Analysis for Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6(4), 413--432 (1997).
11. Lipton, L.: *Foundations of the Stereoscopic Cinema*. Van Nostrand Reinhold (1982).
12. Azuma, R., Bishop, G.: Improving Static and Dynamic Registration in an Optical See-Through HMD. *Proceedings of SIGGRAPH '94*, Computer Graphics, Annual Conference Series, 1994, 197--204 (1994).
13. Fuhrmann, A., Splecht, R., Pikryl, J.: Comprehensive calibration and registration procedures for augmented reality. *Proc. Eurographics Workshop on Virtual Environments 2001*, 219--228 (2001).
14. Michael Caine on Acting in Film. TV miniseries episode produced by BBC Entertainment and Dramatis Personae Ltd. (1987).
15. Symons, L.A., Lee, K., Cedrone, C.C., Nishimura, M.: What are you looking at? Acuity for triadic eye gaze. *J. Gen. Psychology* 131(4), 451--469 (2004).