Int. J. Human-Computer Studies 86 (2016) 138–148

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Effects of 3D perspective on head gaze estimation with a multiview autostereoscopic display ☆

Ye Pan *, Anthony Steed

Virtual Environments & Computer Graphics Group, Department of Computer Science, University College London, UK

ARTICLE INFO

Article history:
Received 21 January 2015
Received in revised form
17 July 2015
Accepted 21 October 2015
Communicated by B. Lok
Available online 30 October 2015

Keywords:
Gaze
Telecommunication
Autostereoscopic display
Random hole display
Avatars

ABSTRACT

Head gaze, or the orientation of the head, is a very important attentional cue in face to face conversation. Some subtleties of the gaze can be lost in common teleconferencing systems, because a single perspective warps spatial characteristics. A recent *random hole display* is a potentially interesting display for group conversation, as it allows multiple stereo viewers in arbitrary locations, without the restriction of conventional autostereoscopic displays on viewing positions. We represented a remote person as an avatar on a random hole display. We evaluated this system by measuring the ability of multiple observers with different horizontal and vertical viewing angles to accurately and simultaneously judge which targets the avatar is gazing at. We compared three perspective conditions: a conventional 2D view, a monoscopic perspective-correct view, and a stereoscopic perspective-correct views. In the latter two conditions, the random hole display shows three and six views simultaneously. Although the random hole display does not provide high quality view, because it has to distribute display pixels among multiple viewers, the different views are easily distinguished. Results suggest the combined presence of perspective-correct and stereoscopic cues significantly improved the effectiveness with which observers were able to assess the avatar's head gaze direction. This motivates the need for stereo in future multiview displays.

1. Introduction

Gaze has several roles in group communication, including facilitating turn-taking, conveying cognitive activity, and expressing involvement etc (Argyle and Cook). However, standard telepresence systems often distort or destroy gaze cues (see e.g., Nguyen et al., 2005; Vertegaal et al., 2003; Schreere et al., 2005), because the single perspective view of the camera does not preserve the spatial characteristics of the face to face situation. In particular, in group conferencing, when a participant looks into the camera, everyone feels that the participant is looking toward them; when the participant looks away from the camera (for example, toward other participants in the meeting), no one sees the participant looking at them (see e.g., Roberts et al., 2013).

A variety of systems have been developed to support gaze awareness, though the majority use a multiple view 2D video system (see e.g., Nguyen et al., 2005) or a single user virtual reality system (see e.g., Roberts et al., 2009). In particular, the use of autostereoscopic display technologies could support multiple users simultaneously each with their own perspective-correct view without the need for special eyewear. However, these are usually restricted to specific optimal viewing zones. Our telepresence system uses the random hole display design (Ye et al., 2010; Nashel and Fuchs, 2009) which has a dense pattern of tiny, pseudo-randomly placed holes as an optical barrier mounted in front of a flat panel display. This allows observers anywhere in front of the display to see a different subset of the display's native pixels through the random-hole barrier. Additionally, it is technically quite simple to build and can be constructed very cheaply in comparison to holographic displays and volumetric displays.

Recently, avatar-mediated communication, where a remote person is represented by a graphical humanoid, has increased in prevalence and popularity as an emerging form of visual remote interaction. The avatar represents the presence and activities of a remote user and can be visualized using standard displays or projection surfaces in the local room with perspective-correct graphical rendering via head tracking of the local user (Roberts et al., 2009). We developed a view-dependent ray traced rendering method to represent a remote person as an avatar on the random hole display. The method allows multiple observers in arbitrary

☆This paper has recommended for acceptance by B. Lok.

* Corresponding author. Tel.: +44 07403023986.
E-mail address: y.pan@cs.ucl.ac.uk (Y. Pan).

http://dx.doi.org/10.1016/j.ijhcs.2015.10.004
1071-5819/

locations to perceive stereo images simultaneously. We investigated using the random hole display to represent a remote person for group teleconferencing. A study explores the effectiveness with which observers can discriminate the avatar's head orientation when the avatar's eyes are centered in the head, because head gaze is a good indicator of focus of attention in human computer interaction applications (Stiefelhagen and Zhu, 2002; Oyekoya et al., 2012). We compared three different conditions: *conventional 2D*, *perspective-correct*, and *perspective-correct & stereoscopy* across nine varying viewing angles. Results show that the presence of both perspective-correct and stereoscopic cues significantly improved the accuracy with which participants were able to assess the avatar's head gaze in both horizontal and vertical directions. This demonstration motivates the further study of novel display configurations and suggest parameters for the design of teleconferencing systems.

In the following sections, we review related work and present the software and hardware components needed to implement our system. This is followed by an experimental evaluation of our system and results. Finally, we present discussions of the results, implications for future designs, conclusions.

2. Related work

2.1. Autostereoscopic displays for teleconferencing applications

Depth perception, or 3D perception, can add a lot to the feeling of immersiveness in many applications such as 3D teleconferencing. However, a conventional stereo display hardware would require the use of 3D glasses, which are cumbersome and make it difficult to support eye contact perception in two way teleconferencing. Autostereoscopic displays, which present a 3D image to a viewer without the need for glasses or other encumbering viewing aids, can be used to improve the teleconferencing experience.

In particular, parallax displays based on barriers or lenticular lens sheets provide a relatively simple and inexpensive solution for autostereoscopy. A parallax barrier is a flat film composed of transparent and opaque regions, while a lenticular screen is a sheet of cylindrical lenses. Parallax barrier displays occlude certain parts of the screen from one eye while allowing another eye to see them. Systems such as Perlin et al.'s autostereoscopic display (Perlin et al., 2000), Varrier (Sandin et al., 2005), and Dynallax (Peterka et al., 2008) demonstrate this concept. Lenticular displays include Kooima et al.'s work (Kooima et al., 2010) and the MERL display (Matusik and Pfister). Additionally, Kim et al. proposed another approach enabling concurrent dual views on twisted-nematic LCD screens, by exploiting a technical limitation of these LCD screens (Kim et al., 2012).

However, neither autostereoscopic displays nor conventional stereo displays support both vertical motion parallax and multiple arbitrary views. Firstly, most conventional autostereo displays do not offer multiuser motion parallax (multiple distinct views) along the vertical direction. Integral imaging displays using a 2D array of lenslets could generate fullparallax autostereo images, but these have a limited viewing angle and low resolution. Therefore, it would be difficult to provide perspective correct views for observers with different heights. With regular multi-user autostereoscopic displays, untracked viewers must remain in certain viewing areas or they will see incorrect imagery or the same imagery as other viewers. In autostereoscopic display systems with user tracking, multiple viewers are usually not supported because individual display pixels will be seen from multiple views.

Recently, an interesting approach to build multi-view displays is based on viewing the screen through a hole-mask that is placed at a certain distance from the data to serve as a barrier that mediates the view for different users. Kitamura et al.'s Illusion Hole uses a display mask which has a hole in its center (Kitamura et al., 2001). Naschel et al.'s random hole display prototype extends their approach by using a randomized hole distribution parallax barrier (Nashel and Fuchs, 2009). The random hole display design eliminates the repeating zones found in regular barrier and lenticular autostereoscopic displays, enabling multiple simultaneous viewers in arbitrary locations (Nashel and Fuchs, 2009). Ye et al. demonstrate a full multi-user multi-view system using this concept with their Tabletop Autostereoscopic Display (Ye et al., 2010). Instead of using a static hole-mask, Karnik et al.'s MUSTARD uses a dynamic random hole mask allowing coverage of the entire screen by constantly changing the hole-mask from frame to frame (Karnik et al., 2012).

While autostereoscopic and multiple arbitrary views capabilities of a random hole display are novel, the effectiveness of using the random hole display for telepresence is not yet clear. We run an experiment to demonstrate that the random hole display can convey head gaze relatively accurately, particularly for group conferencing.

2.2. Gaze in telepresence systems

Gaze, attention, and eye contact are important aspects of face to face conversation. They help create social cues for turn taking, establish a sense of engagement, and indicate the focus and meaning of conversation (Argyle and Cook). However, perceiving gaze direction is difficult in most teleconferencing systems and hence limits their effectiveness (Nguyen et al., 2005). Chen reported that the perception of eye contact decreases if the horizontal contact angle is greater than 1° or the vertical contact angle is greater than 5° (Chen, 2002).

2.2.1. Telepresence systems

Over the years a number of solutions have been developed to convey gaze direction during multiparty video conferencing, including MAJIC (Okada et al., 1994), Hydra (Sellen et al., 1992), GAZE-2 (Vertegaal et al., 2003), MultiView (Nguyen et al., 2005), cylindrical multiview system (Pan and Steed, 2014), 3D facial display (Nagano et al., 2013), animatronic shader lamps avatars (Lincoln et al.) and One-to-Many System (Jones et al., 2009). Also, a variety of solutions have been devised to explore the preservation of 3D depth cues and motion parallax via a single user head position tracking and the use of shutter glasses, such as, Kim et al. (2012), SphereAvatar (Oyekoya et al., 2012; Pan et al., 2014), PCubee (Stavness et al., 2010), Spheree (Ferreira et al.), 3-d live (Prince et al., 2002) and some CAVE-like environments (Roberts et al., 2009; Gross et al., 2003). However, these systems are currently developed for a single observer.

Our system allows multiple observers to see correct stereo images from arbitrary locations in front of the display.

2.2.2. Perception of head and eye gaze direction

The direction of a person's gaze is one feature that is relevant in judging objects of interest in an environment. Gibson et al. established that gaze direction may be perceived by both the direction in which the head is oriented and the eyes' position relative to the head (Gibson and Pick, 1963). Anstis et al. investigated gaze estimation influenced by three orientations of a TV screen. They found a *TV screen turn* effect such that apparent displacement of the perceived direction in the same direction as the turn of the screen and suggested that the convex curvature of the screen probably caused the TV screen turn effect (Anstis et al.). They also reported an *overestimation* effect such that when gaze was to one side of the participant, the participant judges it to be