Layered Telepresence: Simultaneous Multi Presence Experience using Eye Gaze based Perceptual Awareness Blending

MHD Yamen Saraiji*, Shota Sugimoto,
Charith Lasantha Fernando, Kouta Minamizawa, Susumu Tachi

Graduate School of Media Design, Keio University, Japan
Institute of Gerontology, The University of Tokyo, Japan

Figure 1: *(a) Specular reflection through an ordinary glass surface, (b) Watching favorite live soccer stream while participating in simultaneous online meetings, (c) Interact with a local co-worker during an immersive multi-party telepresence activity*

Abstract

We propose "Layered Telepresence", a novel method of experiencing simultaneous multi-presence. Users eye gaze and perceptual awareness are blended with real-time audio-visual information received from multiple telepresence robots. The system arranges audio-visual information received through multiple robots into a priority-driven layered stack. A weighted feature map was created based on the objects recognized for each layer, using image-processing techniques, and pushes the most weighted layer around the users gaze in to the foreground. All other layers are pushed back to the background providing an artificial depth-of-field effect. The proposed method not only works with robots, but also each layer could represent any audio-visual content, such as video see-through HMD, television screen or even your PC screen enabling true multitasking.

Keywords: Simultaneous Multi Presence, Perceptual Awareness Blending, Eye Gaze, Peripheral Vision, Depth of field

Concepts: •**Human-centered computing** → **Interaction techniques; Mixed / augmented reality;** •**Applied computing** → *Telecommunications;*

1 Introduction

Multi party teleconferencing services (such as Skype or Google Hangouts) allow a user to connect with multiple people simultaneously. Based on audio gain on each participant, the system automatically frames the active participant as the main screen and other participants are displayed in a small PIP [1] grid. These systems allow the participants to feel the presence of multiple users, however it is difficult to have a sense of presence at each location at the same time. In contrast, Telexistence [Tachi 2010; Fernando et al. 2012] systems allow a participant to have a real-time presence sensation by binding his body with a remote avatar robot. However, it does not allow connections with multiple robots simultaneously. There have been several works [Benford et al. 1994; McNerney and Yang 1999; Bowman and McMahan 2007] which allow a user to have multiple, simultaneous meetings in a virtual reality, mixed media or collaborative virtual environments. Similarly, using transparent display techniques [Lindlbauer et al. 2014] and video-see through technologies [Fan et al. 2014] blending two visual information containing same visual flow is achieved.

In order to achieve the experience of being presented at multiple locations simultaneously and to be capable to have full awareness of these locations, the following requirements should be addressed:

1. Real-time simultaneous representation of body, visuals, and auditory feedback in multiple locations.
2. Natural mechanism of blending the visuals from the multiple sources.
3. Intuitive mechanism for switching the visual perception between the locations, while maintaining the awareness of the other locations.

As shown in Fig 1(a), we sometimes experience the inside and outside view of a building simultaneously due to reflections on the glass surface. The glass surface works as a blended interface between two locations, making a dual presence experience possible. "Layered Telepresence" was inspired by the same phenomena: merge multiple locations in to a single view, providing both background and foreground information of each layer, so that the user can perceive a simultaneous multi-presence. The layers are reorganized based on the users eye gaze. The most significant layer is

*e-mail: yamen@kmd.keio.ac.jp

SIGGRAPH '16, July 24-28, 2016, Anaheim, CA,
ISBN: 978-1-4503-4371-8/16/07
DOI: http://dx.doi.org/10.1145/2945078.2945098

[1] Picture-in-picture

pops out into the foreground while blending other real-time audio-visual information received from multiple telepresence robots into the background. This creates the effect of an artificial depth-of-field. The idea of using blurring effect and depth-of-field to highlight specific information in an image has been previously discussed in [Kosara 2004] and was applied for social contexts in [Yao et al. 2013]. Layered Telepresence uses similar effect to highlight layers in focus. Two uses for Layered Telepresence are proposed: (1) Exocentric mode as in Fig 1(b) in which the layers are observed using an external display and uses an eye tracker mounted on the display. And (2) Egocentric mode as in Fig 1(b) that uses a HMD with embedded eye tracker to deliver an immersive experience.

2 System Description

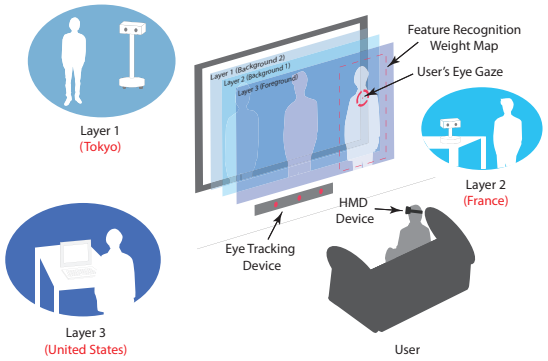

Figure 2: *System Overview*

As shown in Fig 2, Layered Telepresence system currently accepts multiple real-time audio-visual information, such as multiple telepresence robots, 360 video cameras, video see-through HMDs, web cameras, or PC screen. The system arranges the media information into a priority driven layered stack. The media are analyzed for feature points using image processing techniques (emgu CV under Unity). A weighted feature map is then created for each layer, based on the objects recognized, followed by a clipping mask around feature objects. The procedure of generating the weight maps of each layer is summarized in Fig 3, two stages feature detection are applied for each layer: optical-flow detection using Lucas-Kanade method for dynamic contents, and facial detection using Haar cascades classifier for presented people in the scene. The system currently supports up to 3 layers where the user can select the content for each layer.

In our examples we use real-time binocular video information from 3 DOF robot heads. Users can experience the system in immersive 3D using a HMD (Oculus DK2) or with a PC Screen. An eye tracking device (Tobii EyeX) or eye tracking enabled HMD is used to detect the users gaze. For example, as shown in Fig 2, when the user gazes on the right side person, the system takes the gaze point and compares all weighted feature maps around it. Once it matches with a layer (layer 3), a clipping mask is applied around the weighted area and the layer is popped to the foreground. Other layers are pushed back to the background. Layer order is determined with appropriate clipping masks and blending parameters applied to audio-visual information for each layer, creating an artificial depth-of-field effect. In this way, the user can experience a clear view of the gazed person, while others will be blurred out, and can see what is happening on each layer. The result is shown as in Fig 4 where the right side person becomes the foreground (in-focus) and left side person (Layer 2) is pushed back (out-of-focus). In this example Layer 1 is a PC screen, where the user is watching a soccer match.

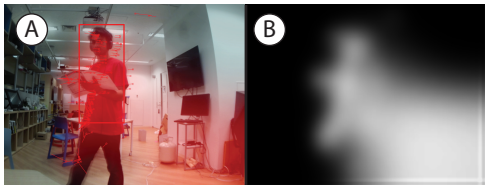

Figure 3: *The process of generating saliency maps and combining the layers.*

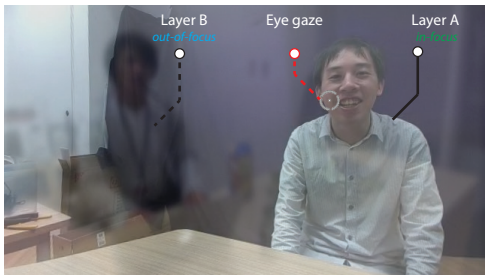

Figure 4: *Final layers blending of two different remote locations.*

3 Acknowledgement

This research is supported by the JST-ACCEL Embodied Media Project.

References

BENFORD, S., BOWERS, J., FAHLÉN, L. E., AND GREENHALGH, C. 1994. Managing mutual awareness in collaborative virtual environments. In *Proceedings of the conference on Virtual reality software and technology*, 223–236.

BOWMAN, D. A., AND MCMAHAN, R. P. 2007. Virtual reality: how much immersion is enough? *Computer 40*, 7, 36–43.

FAN, K., HUBER, J., NANAYAKKARA, S., AND INAMI, M. 2014. Spidervision: extending the human field of view for augmented awareness. In *Proceedings of the 5th Augmented Human International Conference*, ACM, 49.

FERNANDO, C. L., FURUKAWA, M., KUROGI, T., HIROTA, K., KAMURO, S., SATO, K., MINAMIZAWA, K., AND TACHI, S. 2012. Telesar v: Telexistence surrogate anthropomorphic robot. In *ACM SIGGRAPH 2012 Emerging Technologies*, ACM, Los Angeles, CA, USA, SIGGRAPH '12, 23:1–23:1.

KOSARA, R. 2004. *Semantic Depth of Field-Using Blur for Focus+ Context Visualization*. PhD thesis, Kosara.

LINDLBAUER, D., AOKI, T., HÖCHTL, A., UEMA, Y., HALLER, M., INAMI, M., AND MÜLLER, J. 2014. A collaborative see-through display supporting on-demand privacy. In *ACM SIGGRAPH 2014 Emerging Technologies*, ACM, 1.

MCNERNEY, M., AND YANG, R. Y., 1999. System for implementing multiple simultaneous meetings in a virtual reality mixed media meeting room, Dec. 7. US Patent 5,999,208.

TACHI, S. 2010. *Telexistence*. World Scientific.

YAO, L., DEVINCENZI, A., PEREIRA, A., AND ISHII, H. 2013. Focalspace: multimodal activity tracking, synthetic blur and adaptive presentation for video conferencing. In *Proceedings of the 1st symposium on Spatial user interaction*, ACM, 73–76.