Towards 3D-Aware Telepresence: Working on Technologies Behind the Scene.

Òscar Divorra, Jaume Civit, David Espinola
Telefonica Research
Barcelona - Spain
{ode,jaume,davidet}
@tid.es

Rob van Eijk, Wijnand IJsselsteijn
TU/e
Eindhoven – The Netherlands
{R.L.J.v.Eijk, W.A.IJsselsteijn}
@tue.nl

Fei Zuo, Harm Belt
Ralph Braspenning
Philips Research
Eindhoven – The Netherlands
{fei.zuo,harm.belt,Ralph.braspenning}
@philips.com

Ingo Feldmann, Oliver Schreer
Wolfgang Waizenneger
Fraunhofer HHI
Berlin – Germany
{ingo.feldmann, oliver.schreer,wolfgang.waizenneger}
@hhi.fraunhofer.de

Einat Yellin, Pierre Hagendorf
Radvision
Tel-Aviv - Israel
{einat,pierre}
@radvision.com

ABSTRACT

Telepresence has gone a long way since first seminal works on *shared task* and *person spaces* [1]. After a number of technologies, such as broadband internet, high quality HD low-delay video compression, or web applications, have become mature enough, several products have been able to irrupt into the market establishing a solid step forward towards practical true Telepresence solutions. Despite these advances, there is still work to do in what concerns naturality and usability. For instance, current systems are limited to 2D visual communication, limiting proper transmission of body language, and, in general, gaze and eye contact awareness. Telepresence spaces need also more immersive and intuitive interaction with documents and applications for more natural telecollaboration and *task sharing*. In order to tackle these and other shortcomings, emerging technologies such as 3D imaging and multi-touch screens, respectively, are seen to be key for further progress in Telepresence. In this paper, ongoing work in these areas is presented. In particular, the approach taken in EU 3DPresence project is presented as example.

Keywords

Telepresence, 3D Imaging, CSCW, immersive, telecollaboration

1. INTRODUCTION

In virtual collaboration and task sharing within global businesses, teams are geographically dispersed, yet their tasks require collaboration and they share responsibility in working towards a common outcome. These teams are thus critically dependent on the efficacy of mediated communication and collaboration tools. There are several technologies that support virtual collaboration, including computer-mediated communication (CMC), audio conferencing (AC), and video conferencing (VC). State of the art Telepresence systems are typically at the high end of this spectrum, providing audio-visual capture, transmission, and display solutions that attempt to support an illusion that the

The work presented in this paper has been performed within the framework of EU-FP7 3DPresence project -215269-. http://www.3dpresence.eu. The 3DPresence consortium integrated by Telefonica Research, Fraunhofer-HHI, Philips Research, Radvision and Technische Universiteit Eindhoven.

Submitted to CSCW'10, Feb 6–10, 2010, Savannah, Georgia, USA.

Figure 1: State of the art Telepresence system by Cisco.

remote participants are actually together, sharing the same physical space. Telepresence systems have gone a long way since first seminal works on *shared task* and *person spaces* [1],[2],[3]. After a number of technologies, such as broadband internet, high quality HD low-delay video compression, or web applications, have become mature enough, several products have been able to irrupt into the market establishing a solid step forward towards practical Telepresence solutions. Among them, we can count large format videoconferencing systems from major providers such as Cisco Telepresence, HP Halo, Polycom, or telecollaboration suites from leading software companies (see Figure 1). However,

Figure 2: Evolved, more immersive, Telepresence including full 3D display system with multi-perspective experience (green and red perspective viewing cones) and integrated interactive surface for telecollaboration.

current systems still suffer from fundamental imperfections that are known to be detrimental to the communication process.

When communicating, eye contact and gaze cues are essential elements of visual communication, and of importance for signaling attention, and managing conversational flow [4]. It is crucial to establishing rapport and trust in a relationship. Furthermore, immersive integration of telecollaboration interfaces and intelligent *context detection* by the system within *person spaces* appears as a key step in order to allow for the most naturally possible interaction and work with remote conferees. Integration of interactive surfaces, interaction through gestural recognition and room context extraction from multi-modal data (e.g. audio-video) and subjects' gestures, altogether with intuitive and user-centric interfaces are elements that will progressively be introduced on future Telepresence systems.

This paper presents some of the work performed within EU FP7 3DPresence project where, building upon state-of-the-art Telepresence systems concept, it investigates technologies to go towards more immersive 3D-aware Telepresence experiences. In particular, as depicted in Figure 2, we investigate the use of 3D auto-stereoscopic technologies to allow for multi-user gaze direction awareness, and better gesture perspective preservation. Also, gestural interfaces are investigated, starting by the integration of more immersive distributed collaborative interfaces based on interactive graphic surfaces that exploit the connectivity, flexibility and scalability of the cloud.

The paper is structured as follows: Section 2 revises the limitations in terms of eye-contact and gaze awareness in current Telepresence systems, and discusses about multi-perspective 3D auto-stereoscopic imaging technologies being explored in order to tackle this. An internet-distributed interface based on interactive surfaces and web technologies is then presented in Section 3. Finally, future steps are enumerated in Section 4.

2. NEXT GENERATION VISUALIZATION FOR MORE USABLE TELEPRESENCE: 3D

2.1 Are you talking to me?

Mismatches in perceived eye contact and gesture awareness are socially awkward and may lead to misunderstandings as to who is being addressed during a conversation. These cues appear to be of particular importance in social tasks, such as negotiation or conflict resolution [5],[6]. Current commercial solutions still experience problems in fully supporting eye contact and gaze direction, in multi-party (with multiple conferees per site) videoconferences (e.g., see Figure 3). Additionally, in general, current solutions do not support 3D display of communication partners, thereby lacking an important cue to spatial fidelity and physical presence. Visual 3D supports a more immediate spatial impression and improves the perception of proportion and size of remote conferees [7]. Indeed, Mühlbach, Böcker and Prussog [8] demonstrated that stereoscopy can enhance telepresence in videocommunications. People enjoyed and evaluated positively working with the stereoscopic videocommunication system, despite the drawback that this research was performed using shutter glasses, due to the unavailability of viable auto-stereoscopic solutions at that time (i.e., in 1995)[7]. Current auto-stereoscopic (AS) 3D displays are capable to provide multiple stereoscopic views in order to support some head motion

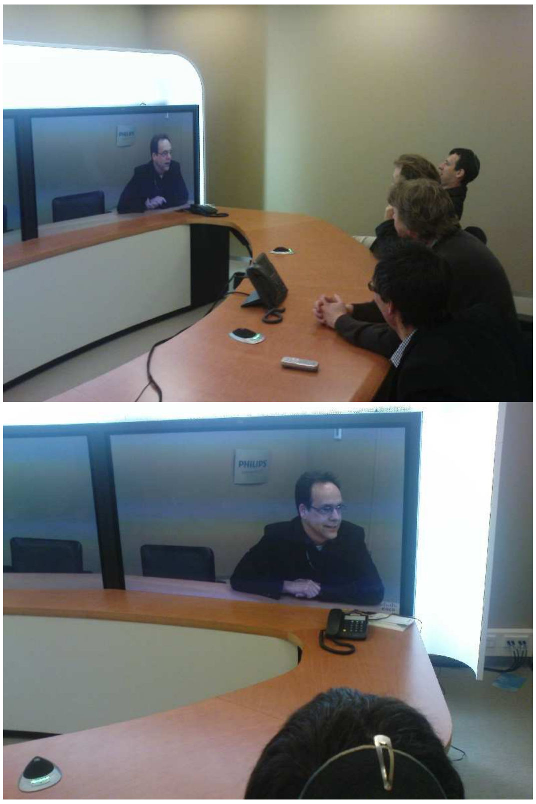


Figure 3 : Lack of directional eye contact and wrong sense of perspective in state of the art 2D Telepresence systems. Both pictures have been taken at the same time. Up: For the middle conferee, gaze direction seems correct. Down: For the conferee in discussion, gaze direction perception is totally flawed.

parallax[9] However, considering a situation with at least two users looking at a remote conferee (see Figure 3), an adapted screen should provide proper stereoscopic viewing for both of them as well as head motion parallax with significantly different simultaneous perspectives onto the scene. The geometrical design

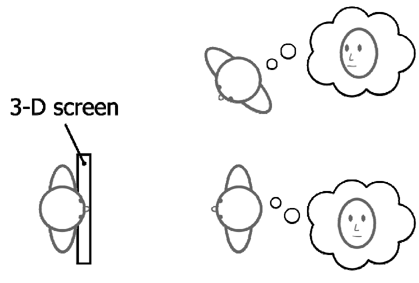


Figure 4: 3D visualization necessary to allow multi-user gaze direction perception for and improved and more natural experience.

of the investigated telepresence system is based upon the well known shared virtual table concept (see Figure 2 and Figure 5). In our particular case, this is supposed to simulate a real conference situation for 3 parties and 6 participants (each remote conferee is replaced by a novel *Multi-perspective AS 3D display*). Eye contact and gesture awareness will be created by virtually adapting the

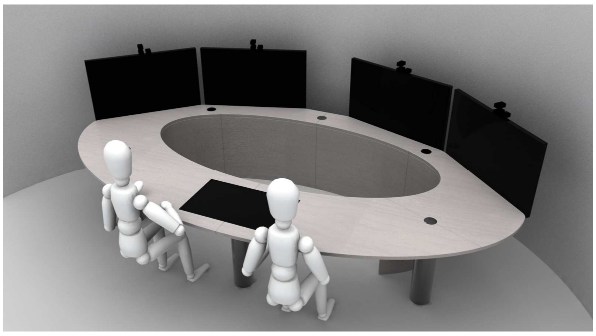


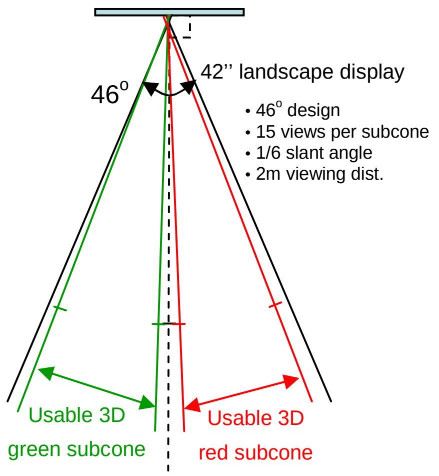
Figure 5 Exact experimental setup under development within the project for shared person and workspace telepresence using prototype multi-perspective 3D Screens, 3D capture and cloud-based interactive graphic surface.

perspective of view, simultaneously, of all remote conferees to each one of the local users.

2.2 3D Visualization for Gaze and Gesture Awereness: Multi-perspective Auto-stereoscopic 3D Displays.

Unlike common properties of usual AS 3D displays [9], the challenge in 3DPresence is to provide stereoscopic viewing for two users and head motion parallax with significantly different perspectives onto the scene. In Figure 4, the conceptual idea of a display with two different viewing cones is presented (viewing cones direction has some degree of freedom for adjustment). The approach taken within the project has been to develop a novel multi-view 3D display which provides two viewing cones providing significantly different perspectives each of them supporting multiple views.

Due to this novel display design, the viewing cones of four displays, related to the four remote conferees, must meet at the correct position of the two local conferees. This has significant impact on the overall geometric design of the system as the overlapping area of all viewing cones related to one local conferee must be as large as possible. The larger the area the better the local conferee can move and change its viewing position. Figure 5



Figures 7: Optical properties of the prototype 3D Display under test for multi-perspective 3D user experience analysis. Each perspective subcon has 7 subviews + 1 transition view.

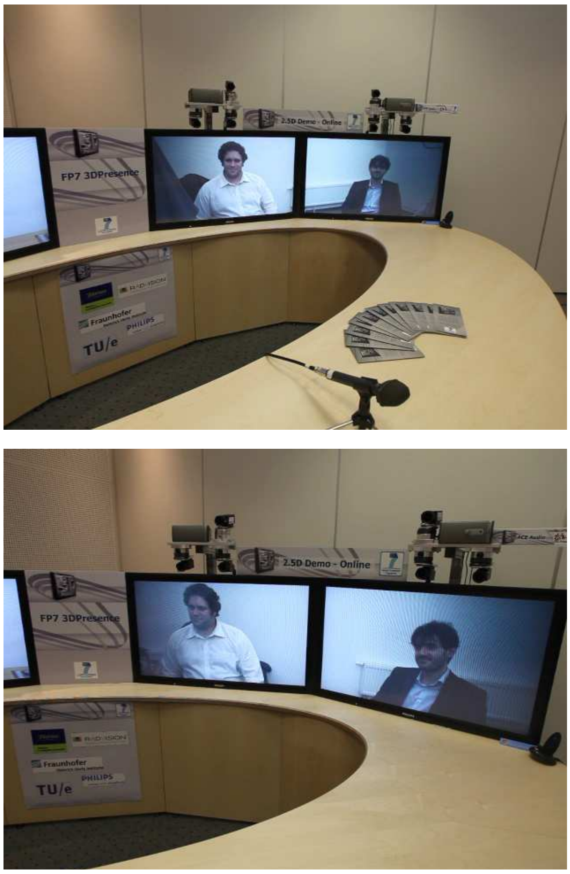


Figure 6 Correct directional eye contact and proper sense of perspective in 3DPresence project experimental setup with multi-perspective screens. Both pictures have been taken at the same time. Up: For the left conferee, gaze direction seems correct (he/she is being looked at). Down: For the conferee on the right, gaze direction perception is coherent with the left one (he/she sees the remote conferees looking at his/her left neighbor).

depicts the display arrangement – notice the slight angle between displays in each pair – in order to achieve the proper overlap among viewing cones (two per display and eight in total). Figure 6 shows the multi-perspective effect produced by the experimental 3DPresence displays on a live around-the-table conference test. Two local conferees can feel simultaneously, and individually, whether they are, or they are not, really looked at.

In another line, efficient multi-view and 3D volumetric scene estimation are investigated [10] in order to enable transmission of AS 3D cues within each viewing cone, as well as to supply input data for room context detection and gesture recognition.

3. CLOUD-CONNECTED MULTI-TOUCH MULTI-USER DESKTOPS FOR MORE NATURAL TELECOLLABORATION

Access to stored digital information for presentation and exchange during a virtual meeting, or sharing a task in the most natural and usable possible manner, can be of critical importance to the success and/or productivity of a meeting. The ability to distribute, discuss and annotate, for example, documents, design drawings or operate an application with anyone in the meeting is part of the

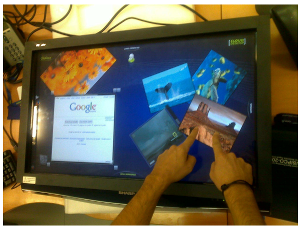


Figure 8: Multi-touch web-application-based interface for multi-user applications, documents and tasks sharing.

natural flow of meetings, and should be part of any telepresence system that aims to support team collaboration. In terms of manipulating and sharing documents, multi-user multi-touch interfaces appear to have a number of promising features, in particular when multi-user, two-handed, *natural* interaction is supported [11]. As was demonstrated by Leganchuk, Zhai, and Buxton [12], two-handed interfaces have two types of advantages. First, there are manual benefits which are based on the fact that two hands allow twice as many degrees of freedom in movement, which allow for increased time-motion efficiency. Secondly, cognitive benefits occur as a consequence of reducing the load of mentally composing and visualizing the sequence of steps that would be required when using traditional techniques such as those used in current telecollaboration software suites. Furthermore, the benefits of two-handed input increase as the task complexity increases [12]. Remote multi-user tele-collaboration, apart from using most recent evolutions from modern interactive surface concepts (e.g. *TouchLight* [11] and beyond), they need to be connected through the *Cloud* by means of most recent web technologies in order to maximize ubiquitousness, flexibility and convergence. In 3DPresence project, based on a simple multi-touch overlay, a tele-collaboration desktop (*Cloud Desktop*, see Figure 8) has been designed using *Adobe-Flash/Flex* technology and a client-server communication engine architecture (see Figure 9). This interactive web desktop can run web applications being able to run on flash or simply on any browser. At the same time, the client-server architecture distributes among all the connected terminals the necessary information for real-time cross-site interaction. The fact that mainstream web technologies have been used to build such an interactive web desktop for collaboration, allows to any device with a touch-screen and a browser being able to join a telecollaboration meeting based on this application.

4. Future Work

Next steps include future research on 3D capture and display for improved visual telepresence cues, integration of context extraction (e.g. gestures, or attention detection), recognition into the system for an improved immersive experience and user experience assessments for further research development. Part of this work will be prosecuted as described and discussed by Cherubini et al. in [13].

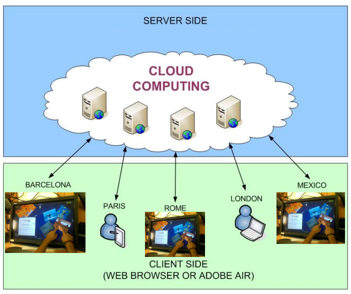


Figure 9: Cloud-based design of a distributed interactive graphic surface to ease usability while being ubiquitous.

5. REFERENCES

[1] Buxton, W.; Telepresence: Integrating Shared Task and Person Spaces

[2] Buxton, W. and Moran, T.; EuroPARC's Integrated Interactive Intermedia Facility (iiif): early experience. Multi-user interfaces and applications, Proceedings of the IFIP WG 8.4, Conference on Multi-user Interfaces and Applications, Heraklion, Crete.

[3] Buxton, W. and Sellen, A. (1991). Interfaces for multiparty video conferences. University of Toronto.

[4] Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*, 78–100.

[5] Fish, R. S., Kraut, R. E., Root, R. W., & Rice, R. (1993). Video as technology for informal communication. *Communications of ACM, 36*, 48-61.

[6] Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.

[7] Harper, B. & Latto, R. (2001). Cyclopean vision, size estimation, and presence in orthostereoscopic images. *Presence: Teleoperators and Virtual Environments, 10*, 312-330.

[8] Mühlbach, L., Böcker, M., & Prussog, A. (1995). Telepresence in videocommunications: A study on stereoscopy and individual eye contact. *Human Factors, 37*, 290-305.

[9] Braspenning, R., Brouwer, E., de Haan, G.. (2005). Visual quality assessment of lenticular based 3D displays. *In Proc. of EUSIPCO*.

[10] Feldmann, I., Atzpadin, N., Schreer, O., Pujol-Acolado, J.-C., Landabaso J.-L., Divorra Escoda, O., (2009), Multi-View Depth Estimation Based on Visual-Hull Enhanced Hybrid Recursive Matching for 3D Video Conference Systems. *In Proc. of IEEE ICIP*.

[11] Wilson, A., TouchLight:, an Imaging and Touch Screen and Display for Gesture-Based Interaction, Siggraph 2005 Emerging Technologies, 2005

[12] Leganchuk, A., Zhai, S., & Buxton, W. (1998). Manual and cognitive benefits of two-handed input: An experimental study. *Transactions on Human-Computer Interaction,* 5(4), 326-359.

[13] M. Cherubini, R. de Oliveira, N. Oliver, and C. Ferran, "Gaze and gestures in telepresence: multimodality, embodiment, and roles of collaboration," in Paper presented at the International Workshop New Frontiers in Telepresence, part of CSCW'10 (G. Venolia, K. Inkpen, J. Olson, and D. Nguyen, eds.), (Savannah, GA, USA), February 7th 2010.