Interface Design for an Inexpensive Hands-Free Collaborative Videoconferencing System

Nicolas H. Lehment* Katharina Erhardt† Gerhard Rigoll‡

Institute for Human-Machine Communication
Technical University Munich

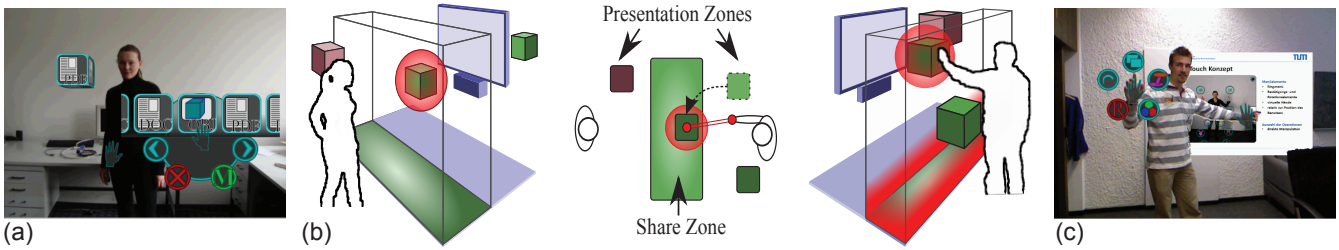

Figure 1: Basic layout of the proposed collaborative system. Virtual objects are augmented into the participants views and can be manipulated by touch. Also shown are the hand centered marker menus (a,c), spatial sharing management (b) and the occlusion handling (c).

Abstract

In this paper an interaction framework for AR enhanced video conferencing is presented. The goal is to provide a cheap and portable system based on a combination of commodity Kinect cameras and regular computer screens. These conditions necessitate the use of contact free interaction methods. The interaction framework presented in this paper is specifically suited for remotely presenting, sharing and annotating visual data such as images, presentation slides and 3D objects. In the proposed system all data is represented by freely manipulable 3D objects which are augmented into the camera views. These representations are integrated into a differentiated ownership scheme, allowing for operations such as spatially managed data sharing. The suitability of different interaction paradigms with regards to this usage scenario is examined. Furthermore, occlusion and collision management between virtual objects and real obstacles is enabled by integrating basic models of the environment.

Index Terms: H.4.3 [Information Systems Applications]: Communications Applications—Computer Conferencing, Teleconferencing, and Videoconferencing;

1 Introduction

We propose a compact and affordable telepresence setup consisting of a single Kinect camera and a regular computer display at each participant's office. The envisioned primary applications are the presentation, exchange and discussion of images, presentation slides and 3D models. To this end we require face-to-face presence with simultaneous display and annotation of the relevant data. In order to increase immersion, we decided to integrate these data directly into the presentation space, resulting in a mixed-reality video conferencing scenario. Additionally, we use on-the-fly environment modeling and occlusion handling to enable realistic interactions between virtual and real objects.

*e-mail: Lehment@tum.de
†e-mail: Ker@mmk.ei.tum.de
‡e-mail: Rigoll@tum.de

IEEE International Symposium on Mixed and Augmented Reality 2012
Science and Technology Proceedings
5 - 8 November 2012, Atlanta, Georgia
978-1-4673-4661-0/12/$31.00 ©2012 IEEE

2 Related Work

The collaborative handling of virtual objects marks a distinct focus of current research. Following the development of a shared space concept by Buxton [3], works by Barakonyi et al. [2] and Kuechler et al. [5] elaborate on this idea. These publications tend to concentrate on collaborative functionality in a well defined task space.

On the other hand we find research dealing with the interaction and rendering of participants. Early work by Prince, Billinghurst et al. [7] eventually led to the life-sized telepresence system developed by Kim et al. [4]. Also of note are recent works by Schreer et. al [8] and Maimone et al. [6]. These works have in common that they motivate their approach with the natural and spatially consistent display of two or more users in conversation scenarios.

Of special interest to our problem is the recent series of advances in pose and gesture recognition [9, 10] which were sparked by the market introduction of the Kinect camera system.

The goal of our work is to combine the essential ideas of these strains of research in order to facilitate the exchange of data and ideas by combining mixed reality and natural user interfaces.

3 System Overview

A schematic overview is given in Figure 1. All data is represented by 3D objects which are placed in a common virtual space encompassing both physical locations. This common virtual space is divided into two presentation zones, where data access for the partner is limited, and the share zone, where the data is available in full to both sides. Annotations can be added by common drawing methods, e.g. rectangles and freehand curves. In order to facilitate annotation, the users are able to switch between viewing their conversation partner and their partner's view of themselves. Interactions are based on hand pose. Since single finger detection becomes unstable in our scenario, we use remote touch selection in depth-triggered marker menus. The 2D object selection uses a 1 second dwell time. Translation is mapped 1:1 to the hand movement, rotation is applied by a fixed degree-per-second ratio. The rotation axis depends on the 2D screen position of the hand relative to the object. Optical cues inform the user of the current system state, e.g. object ownership, object selection etc.

In order to achieve dynamic foreground occlusion, we use the current depth map and a custom ray casting shader to determine object visibility. Furthermore, object-to-world collisions are enabled by background modelling (summarized in Figure 2).

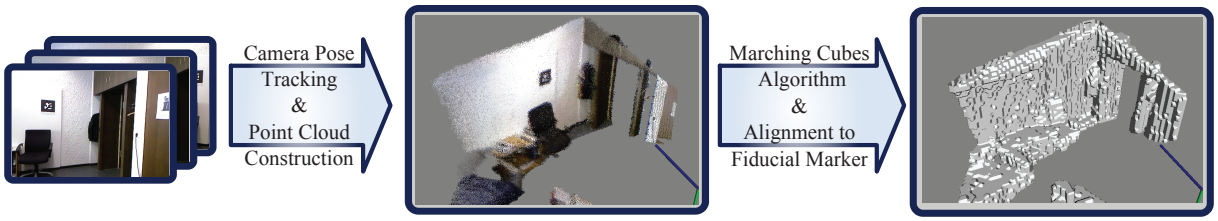

Figure 2: Background modeling based on inter-frame 2D SURF correspondences which are used by an ICP algorithm for camera tracking. The Marching Cubes algorithm then produces a wire mesh model of the static background. A fiducial marker provides a reference transformation.

4 Initial Findings

We found that gesture based interaction in face-to-face conference scenarios poses special challenges: Control gestures must be clearly distinguishable from conversational gesturing in order to avoid confusing both the control system and conversation partners. Additionally, not all efficient control gestures are also socially appropriate in face-to-face conversations. To this end we add visual markers to signal on-going interactions with the control interface to conversation partners. Showing menu outlines around the interacting hand appears to eliminate most misunderstandings.

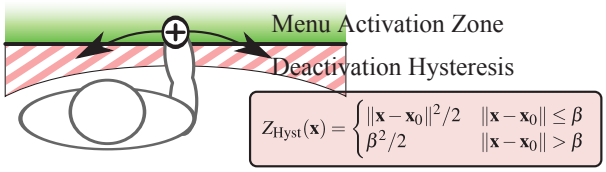

Figure 3: Using a Talwar-function hysteresis to prevent inadvertent menu closing in depth activated dialogues.

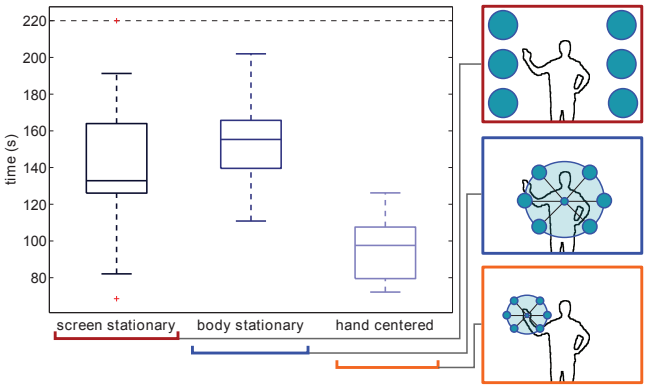

Figure 4: Comparison of required time on a combined manipulation and annotation task for 16 users on three interaction paradigms. Hand centered menus minimize arm movement resulting in low selection times and arm fatigue.

Conventional gesture controlled interfaces tend to misinterpret conversational gestures as input. A viable solution is the use of depth-activated marker menus [1]. The menu appearance is triggered by a fixed depth threshold with a Talwar-function hysteresis in order to compensate for non-linear arm movement (shown in Figure 3). A comparative study of interaction paradigms (summarized in Figure 4) has shown hand centered interfaces to perform best for this application. Since robust single finger detection is hard to achieve at distances greater than 2m, we use remote "touch" by the full hand to select items in these wrist-centered menus.

The spatial data access management found positive response in initial trials. Especially the notion of "handing over" data appears to be an intuitive interaction concept.

5 Conclusions & Outlook

We present a compact and inexpensive remote collaboration tool. Special attention is given to the integration of a hands-free user interface into a mixed reality collaborative telepresence system. A simple and intuitive data sharing mechanism is proposed and implemented in a dual user scenario. While there are open questions concerning the possible conflicts between control gestures and conversational gesturing, the overall concept of integrating mixed reality remote conferencing with gesture based interfaces found a positive response in preliminary user trials.

The integration of gesture based user interfaces into a social interaction scenario continues to raise interesting questions concerning acceptable interface choices. The presented system is thus a suitable vehicle for further studies on user acceptance for various interaction paradigms in a face-to-face conversation.

References

[1] G. Bailly, R. Walter, J. Müller, T. Ning, and E. Lecolinet. Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. In *Human-Computer Interaction INTERACT 2011*, pages 248–262. Springer Berlin / Heidelberg, 2011.

[2] I. Barakonyi, T. Fahmy, and D. Schmalstieg. Remote collaboration using augmented reality videoconferencing. In *Proceedings of Graphics Interface 2004*, pages 89–96, 2004.

[3] W. Buxton. Telepresence: Integrating shared task and person spaces. In *Proceedings of Graphics Interface 1992*, pages 123–129, 1992.

[4] K. Kim, J. Bolton, A. Girouard, J. Cooperstock, and R. Vertegaal. TeleHuman: Effects of 3D perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 2531–2540, 2012.

[5] M. Kuechler and A. M. Kunz. Collaboard: A remote collaboration groupware device featuring an embodiment-enriched shared workspace. In *Proceedings of the 16th ACM international conference on Supporting group work*, pages 211–214, 2010.

[6] A. Maimone and H. Fuchs. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In *Proceedings of the 10th International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 137 –146, 2011.

[7] S. Prince, A. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. Kato. 3D live: Real time captured content for mixed reality. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 316 – 317, 2002.

[8] O. Schreer, I. Feldmann, N. Atzpadin, P. Eisert, P. Kauff, and H. Belt. 3DPresence - A system concept for multi-user and multi-party immersive 3D videoconferencing. In *Visual Media Production (CVMP 2008), 5th European Conference on*, pages 321–334, 2008.

[9] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from a single depth image. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, pages 945 – 952, 2011.

[10] M. Van den Bergh and L. Van Gool. Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, pages 66 –72, 2011.