A Projected Augmented Reality System for Remote Collaboration

Matthew Tait
The HIT Lab NZ
University of Canterbury
Christchurch, New Zealand
Mdt45@uclive.ac.nz

Tony Tsai
The HIT Lab NZ
University of Canterbury
Christchurch, New Zealand
Tmt27@uclive.ac.nz

Nobuchika Sakata
Human Interface Laboratory
Division of Systems Science and
Applied Informatics
Osaka University, Japan
sakata@sys.es.osaka-u.ac.jp

Mark Billinghurst
The HIT Lab NZ
University of Canterbury
Christchurch, New Zealand
mark.billinghurst@hitlabnz.org

Elina Vartiainen
ABB Corporate Research
Västerås, Sweden
elina.vartiainen@se.abb.com

Extended Abstracts of the
IEEE International Symposium on Mixed and Augmented Reality 2013
Science and Technology Proceedings
1 - 4 October 2013, Adelaide, SA, Australia
978-1-4799-2869-9/13/$31.00 ©2013 IEEE

Abstract

This paper describes an AR system for remote collaboration using a captured 3D model of the local user's scene. In the system a remote user can manipulate the scene independently of the view of the local user and add AR annotations that appear projected into the real world. Results from a pilot study and the design of a further full study are presented.

Author Keywords

Remote collaboration, Kinect Fusion, Augmented Reality, Projection

ACM Classification Keywords

H.5.1 Multimedia Information Systems: Artificial, augmented and virtual reality; H.5.2 User Interfaces: Interaction Styles; H.5.3 Group and Organization Interfaces: Computer-supported cooperative work

Introduction

Remote collaboration is a type of computer supported collaborative work in which two or more people work on the same task using a computer system to mediate communication between them. There are various ways in which remote collaboration can be done such as video-conferencing or audio only conference calls.

Remote collaboration can also involve sharing a view of one or more of the users task spaces. For example, a user could wear a head mounted display (HMD) and a head mounted camera so that a remote user can share the same view. The system described in this paper uses an expert/student relationship with the remote expert observing the scene of the student or local user.

Augmented Reality (AR) can be used to support remote collaboration by tracking the task space of one of the users and overlying virtual annotations onto their view of the real world. However, there has been little research in the field of remote collaboration using AR [1]. In our research we are interested in how AR can be used to enhance remote task space collaboration.

For effective task space communication there needs to be an awareness of the local user's environment. In some systems this is done using image tracking or image based 3D maps [2]. However, fast, cheap and accurate depth sensing technology such as the Microsoft Kinect [8] can be used to build up dense 3D maps of the scene [3]. Tracking and mapping 3D scenes and objects has been extensively researched and used for such applications as AR [4] and autonomous vehicles [5]. Different map densities can be built up using a variety of different techniques. For example, sparse maps are often used for tracking [6] and dense maps for model reconstruction [7]. Our work involves creating a dense depth map of the user’s environment that can have AR cues added to it.

Related Work

In order to develop an AR interface that uses depth information for effective remote collaboration we need to consider earlier related work in scene capture, remote collaboration and user interface. In this section key related work is reviewed in each of these areas.

There are several methods for capturing a 3D model of a scene. A technique that uses a single camera is Parallel Tracking and Mapping (PTAM) [6]. However, PTAM does not create a dense enough or accurate map to create a detailed and model. Some techniques for scene capture use depth sensing cameras, such as the Microsoft Kinect. In [8] the accuracy of the Kinect camera is tested and compared to that of a laser scanner. It was found that the Kinect camera did not have large systematic errors and that the accuracy is acceptable for scene capture. KinectFusion [9] improves depth capture further by combining the Kinect with a high end Nvidia GPU to capture a dense map of the 3D scene that can be exported to a mesh. However KinectFusion has significant hardware requirements.

An AR remote collaboration system is described in [11] for use in a crime scene situation. The system involves a local user wearing a head mounted display and stereo cameras. A 3D map is built up of the scene using PTAM [6], which also tracks the pose of the user. The remote expert can then attach virtual annotations to points on the map that can be viewed by the local user. The 3D map is freely navigable by the remote user. The main problem encountered in this system was the lack of virtual co-presence, which made it difficult for the users to effectively communicate. The pose tracking system also failed to allow for fast movement or large maps, meaning that the local user had to move slowly in a localized area when using the system.

For the remote expert an important interface element is how they view and interact with the local user’s scene.

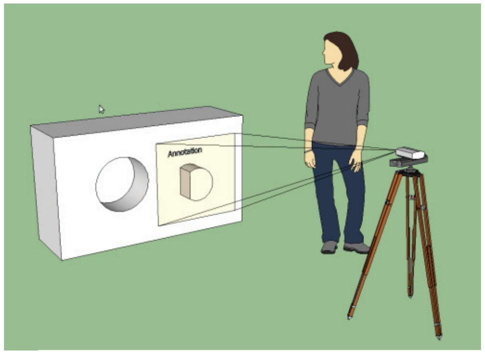

Figure 1. Local user environment

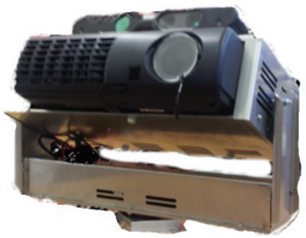

Figure 2. Mounted projector and Kinect system

One method of communicating with the user is through gestures. For example [10] describes a remote collaboration system in which the operator uses a head mounted display and camera. The remote expert can gesture in front of their display, and his/her gestures are captured by a camera and superimposed on the remote operator's display.

One variable in remote user interfaces is whether the expert has some control of the viewpoint as in [2], [11], [14] or whether it is fixed to the local user as in [12], [13]. A way of improving remote expert interaction with a scene is proposed in [13]. In this case the remote expert has the ability to 'freeze' the scene, allowing interaction without movement of the viewpoint, increasing input accuracy and ease of use.

In [2] a system is devised where the camera itself is remotely controlled by the expert. The local user wears the camera on their shoulder and the viewpoint is maintained using gyros and accelerometers. The remote expert has the ability to remote control the camera and change its viewpoint. Gesturing is achieved by attaching a laser pointer to the camera so that the remote expert can also annotate the user's real world. In a study with the system, subjects found it significantly more comfortable to use the shoulder mounted camera than to look at the feed from the head mounted display.

In summary, the currently existing systems allow independent views either through the use of sparse mapping [11] or through an independently controlled camera on servo-motors. The use of a full 3D model generated through depth cameras has not been examined. In our work we want to explore the use of a 3D scene capture system in combination with a projected display for remote collaboration.

System Developed

A system has been developed that allows the creation of a 3D model of the local user's scene that can be manipulated by the remote user (see figure 3). The system involves the local user scanning the scene, creating a textured model of the environment and transmitting it to the remote user. The remote user then can independently view the scene. Capture of the user's environment is done with a Kinect depth sensor mounted on a laser projector.

The image captured by the Kinect is displayed to the remote user in a floating window – the pose of which is determined by the pose of the Kinect and matches the direction of the projector so the remote user is aware of where the local user can see annotations.

The remote user can then place virtual annotations in the local user's view by clicking on the scene where the annotations should be placed. This annotation is then sent back to the remote user and projected back onto the scene using a laser projector. A laser projector is used due to it being very bright, and having a near infinite focal length, meaning it is in focus for all distances. The annotations that can be placed currently includes text and point annotations. How the local user sees the system is shown in Fig 1.

The Kinect and laser projector are mounted on yaw/pitch servo-motors as shown in Fig 2. This means that the remote user can also control the orientation of the Kinect/projector device by the use of the keyboard. Pressing arrow keys allows the remote user to control the yaw/pitch of the device, and to project content anywhere they want in to local users workspace.

The remote user has the ability to shift between having their view fixed to that of the local user or having the view independently controlled through a standard mouse interface. The remote user's graphical user interface is displayed in Fig 3, showing both the 3D model of the scene that has been captured and the floating window of live video that shows the camera pose. The user interface is designed to allow natural interaction with the scene. Navigation is done using a keyboard and mouse interface.

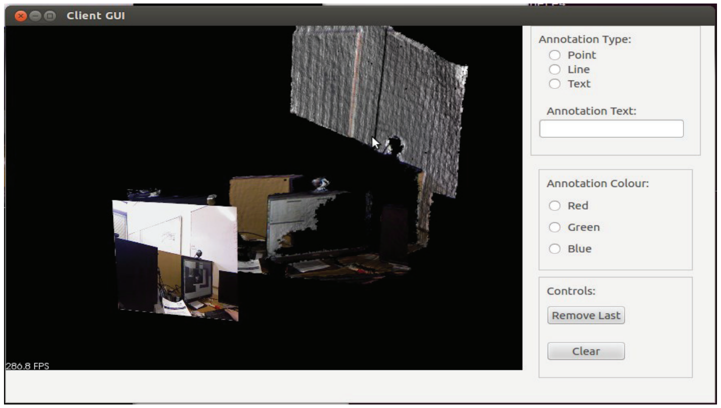

Figure 3. The remote user interface for the system

Placing annotations is done by the remote user holding down the shift key and clicking on the model surface. The type of annotation can be selected from a point, line or text annotation using a series of check boxes ('Annotation Type' in Figure 3). When placing a text annotation the desired text can be written into a text box ('Annotation Text' in Figure 3). The current system supports placing annotations in three colours; red, green or blue. Annotations can be removed either by undoing the last annotation or by clearing all annotations.

The captured point cloud is not segmented in this system, so the modeling, done using the marching cubes algorithm, and texture mapping takes approximately twenty seconds for a desk-sized scene. The system has been set up to work over a network using Google Protocol Buffers. There exists a client-server relationship with the local user being the server and the remote user being the client. Every RGB image frame captured by the Kinect camera is transferred as a JPEG and the pose of the camera calculated using ICP is transferred as a transformation matrix. Annotations are transferred from the remote user to the local user only when they are added. Currently the amount of data being transferred is approximately the size of the JPEG image, between 90 and 100 kilobytes. The scene model that is sent at the beginning of the collaboration is approximately 100 MB for a desktop sized area.

Pilot User Study

A pilot user study has been performed with three pairs of participants to investigate the effects of the remote user being able to control the model of the scene. In this study the local user was situated in a workspace with two walls with five workspaces situated on them as shown in Fig 4.

The remote user was to draw a series of simple shapes in each workspace in set colours, and the local user was then to trace over these shapes. The study used the system as described above with two variables. The first was camera control. In one case the remote user had the ability to manipulate the scene and could see live video floating as described above. In the second case the remote user did not have the ability to manipulate the scene. The remote users view was locked to the camera position and live video was displayed in a separate window.

The other variable was the area required to complete the task. In the first case the local user scene was able to fit within a single camera shot so the camera did not need to be moved. In the second case the local user scene required the camera to be panned to complete all tasks. The large workspace is as shown in Fig 4, with the camera required to move between the workspaces placed on two walls.

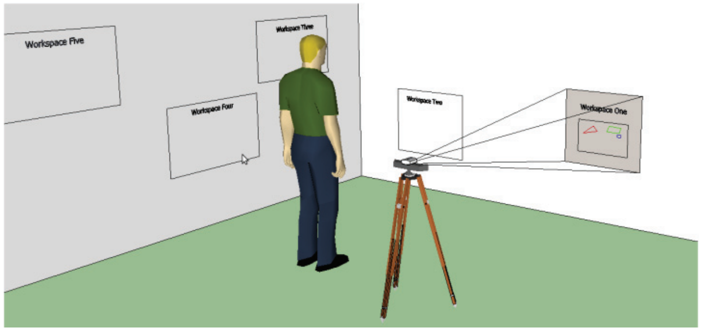

Figure 4. Local user workspace for pilot user study

It was hypothesized that for tasks that involve a sequential process that is spread over a relatively large area the independent view of the model will have faster completion times due to the remote user being able to get ahead of the local user by changing his view. For tasks that involve a small area, such as one that fits in a single frame of the camera, completion time will not be significantly different as the remote user will not gain an advantage from having an independent view.

In both kinds of task it was hypothesized that the independent view will be more comfortable for the remote user to use, as this has been shown to be the case in other systems where some form of view independence was allowed.

Results

The small pilot study found that the independent view was not preferred by any of the remote users. In both the large and small scene cases the fixed view was preferred by all participants. Two of the participants commented that the control scheme of the independent view was not easy to use and all of the participants rated the independent view as more difficult to operate. One user commented that the task favoured the use of the dependent view due to its simplicity.

Discussion

It was noted that this study did not encourage discovery for the remote user – there was not any point at which the remote user had to locate something in the local users scene. A task with an element of discovery may make better use of the independent camera view. The control scheme for the independent view may have caused the study results to be less reliable in that it made controlling the camera very difficult. This may be a large part of the cause for the preference of the fixed view.

Further Work

Based on the pilot study changes to both the system and a further study were made. The largest change was that the full study would require discovery by the remote user. Instead of blank pieces of paper as workspaces the remote user will instead be required to arrange 3D objects around existing objects in the scene and the local user will be required to place the objects where indicated. This will mean the remote user has to find the object and then place the model.

Another change is that the interface for navigating the scene independently will be simplified. It is planned to use an arcball interface such as described in [15], which has been shown to be easy to use for 3D scene rotation. This should give a better indication of how having an independent view should effect collaboration by making the actual navigation less arduous.

References

[1] H. B.-L. Duh and M. Billinghurst, “Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR,” In Proceedings of ISMAR 2008, pp. 193–202, Sep. 2008.

[2] T. Kurata and N. Sakata, “Remote collaboration using a shoulder-worn active camera/laser,” In Proceedings of ISWC 2004, vol. 69, pp. 62–69, 2004.

[3] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” 2011 IEEE International Conference on Robotics and Automation, pp. 1–4, May 2011.

[4] T. Piumsomboon, “Physically-based interaction for tabletop augmented reality using a depth-sensing camera for environment mapping,” In Proceedings of IVCNZ, 2011.

[5] J. J. Leonard, H. F. Durrant-Whyte, and I. J. Cox, “Dynamic Map Building for an Autonomous Mobile Robot,” The International Journal of Robotics Research, vol. 11, no. 4, pp. 286–298, Aug. 1992.

[6] G. Klein and D. Murray, “Parallel Tracking and Mapping for Small AR Workspaces,” In Proceedings of ISMAR 2007, pp. 1–10, Nov. 2007.

[7] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, “RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments,” The International Journal of Robotics Research, vol. 31, no. 5, pp. 647–663, Feb. 2012.

[8] K. Khoshelham, “Accuracy analysis of kinect depth data,” ISPRS workshop laser scanning, 2011.

[9] S. Izadi, D. Kim, and O. Hilliges, “KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera,” Proceedings of UIST 2011.

[10] S. Lukosch, R. Poelman, O. Akman, and P. Jonker, “A Novel Gesture-based Interface for Crime Scene Investigation in Mediated Reality,” Proceedings of the CSCW workshop on Exploring collaboration in challenging environments, pp. 3–6.

[11] R. Poelman and O. Akman, “As if Being There: Mediated Reality for Crime Scene Investigation,” *Proceedings of CSCW 2012*, no. 5, 2012.

[12] H. Kuzuoka, T. Kosuge, and M. Tanaka, “GestureCam: A video communication system for sympathetic remote collaboration,” Proceedings of the CSCW 1994, 1994.

[13]] J. Ou, X. Chen, S. Fussell, and J. Yang, “DOVE: Drawing over video environment,” Proceedings of ACM Multimedia, 2003.

[14] G. Lee, U. Yang, Y. Kim, D. Jo, and K. Kim, “Freeze-Set-Go interaction method for handheld mobile augmented reality environments,” Proceedings of VRST 2009, pp. 143–146, 2009.

[15] Shoemake, Ken. "ARCBALL: a user interface for specifying three-dimensional orientation using a mouse." Graphics Interface. Vol. 92. 1992.