Towards Object Based Manipulation in Remote Guidance

Dulitha Ranatunga
Australian National University
CSIRO
Canberra, Australia
dulitha.ranatunga@anu.edu.au

Matt Adcock
CSIRO
Australian National University
University of South Australia
Canberra, Australia
matt.adcock@csiro.au

David Feng
Australian National University
CSIRO
Canberra, Australia
david.feng@csiro.au

Bruce Thomas
University of South Australia
Mawson Lakes, Australia
bruce.thomas@unisa.edu.au

IEEE International Symposium on Mixed and Augmented Reality 2013
Science and Technology Proceedings
1 - 4 October 2013, Adelaide, SA, Australia
978-1-4799-2868-2/13/$31.00 ©2013 IEEE

Abstract
This paper presents a method for using object based manipulation and spatial augmented reality for the purpose of remote guidance. Previous remote guidance methods have typically not made use of any semantic information about the physical properties of the environment and require the helper and worker to provide context. Our new prototype system introduces a level of abstraction to the remote expert, allowing them to directly specify the object movements required of a local worker. We use 3D tracking to create a hidden virtual reality scene, mirroring the real world, with which the remote expert interacts while viewing a camera feed of the physical workspace. The intended manipulations are then rendered to the local worker using Spatial Augmented Reality (SAR). We report on the implementation of a functional prototype that demonstrates an instance of this approach. We anticipate that techniques such as the one we present will allow more efficient collaborative remote guidance in a range of physical tasks.

Author Keywords
Spatially Augmented Reality; Remote Guidance; Object Manipulation; Multi touch interaction, 3D CHI.

ACM Classification Keywords

H.5.1. Information interfaces and presentation: Artificial, augmented and virtual realities; H.5.2. Information interfaces and presentation: Input devices and strategies.

General Terms

Human Factors, Design, Experimentation

Introduction

Remote Guidance takes place when multiple participants in different locations work together to perform a task involving physical objects. Typically the scenario involves an 'expert' with specialized knowledge in the task situated remotely from the physical task environment. They collaborate, using available communication tools, with a 'worker' who is co-located and able to physically interact with the task environment.

There are many different approaches to improving the communication tools and capabilities of the remote helper to enable efficient guiding techniques. These include video-conference situations, view sharing based head mounted display(HMD) systems [7] and through the use of spatial augmented reality. Unfortunately, in video-conference based systems, the expert is only able to provide verbal cues in response to the visual feed. HMD systems are a form of augmented reality however they suffer from encumbering wearable devices that can limit the helpers' freedom to move or operate.

Spatial Augmented Reality (SAR) systems use projectors to display computer generated graphics directly onto the physical environment. SAR provides the benefit of augmentation without wearable or hand-held devices. Furthermore, since the workspace is augmented spatially, the same graphics can potentially be seen by multiple workers with natural stereo affordances.

Some SAR based remote guidance systems have superimposed the expert's sketches or hand gestures into the workers environment but they have typically not used any semantic information about the 3D properties of the physical objects.

We propose a method that encompasses the benefits of SAR and incorporates interaction techniques previously developed for manipulating virtual 3D objects. As shown in Figure 1, users are able to touch objects on a video feed and manipulate a hidden virtual representation of the physical work environment. Changes made in the virtual scene are used to generate augmentations in the real world that act as spatially aligned guidance for the worker. As objects are moved within the physical scene, the virtual model is updated and the augmentations can also be updated accordingly.

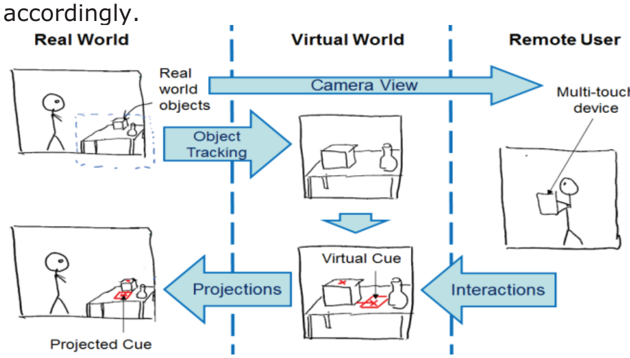

Figure 1: A workflow of how virtual reality can be used to assist in SAR-based remote guidance**.**

Related Work

Our approach draws from a variety of fields, including remote guidance, SAR guidance, and 3D multi-touch.

Remote Guidance

A comparison of existing remote gesture technologies was previously conducted by Kirk [6]. Two SAR-like methods were discussed in this paper; projected hands, and projected sketches. In the first setup, the helper's hands were recorded from a top down camera and directly projected onto the surface of the workers location. The second approach extended the first by including a digital whiteboard that allowed for the combination of sketching as well as the hands. In both systems, visual feedback to the expert was via a camera feed of the augmented physical world shared with the worker. There was no model of the scene within the system so it was up to the expert to orient their self, define reference points, and use relative movements.

In some non-collaborative scenarios, SAR has been shown to benefit guidance of physical tasks. LightGuide is a system that projected guidance cues onto a moving hand to follow a particular path [11]. Rosenthal et al. used micro-projectors to complement on-screen instructions for manual task guidance and determined situations in which guidance improved and decreased performance tasks [9].

SAR Guidance

Tsimeris explored a variety of projected cues that could assist in the arrangement of physical objects [14]. While the system did incorporate a 2D GUI authoring tool, capable of real-time manipulation, it did not facilitate a collaborative remote guidance scenario. That research determined that SAR visual cues were faster and more accurate than manual arrangement.

Henderson has shown that dynamic instructions overlaid on or near objects is preferred for psychomotor tasks and is significantly more efficient when compared to a nearby LCD display [2]. That study was specific to HMD based augmented reality and the guidance was entirely pre-authored.

Tecchia employed depth cameras to stream real time virtual representations of the workspace for remote guidance [13]. The system required the expert to wear an HMD showing a depth camera feed from the workspace, while the worker watched a combination of that feed and a depth camera feed taken of the expert's hands.

Suenaga et al.[12] implemented a tele-instruction system which used a shared AR space. The expert would manipulate a single ultrasonic probe via a visualization named a 'Web-Mark' which was projected onto the body of a patient. Hiura[3] developed a tele-direction system that creates a virtual model of objects to allow for projected annotations. Our system extends these ideas by allowing the expert to directly manipulate multiple objects using touch. Our system also does not require the expert to first trace the outline of the target object(s).

3D Multi-Touch

A number of research efforts have explored the ways 2D multi-touch can be used to interact with 3D virtual environments and Liu et al. [8] provide a comparison of current methods. This is a relatively new field, but already we are seeing useful abstractions emerge.

The touch device only displays the video feed from the camera and not the virtual scene. There is a model of the camera, Kinect and projector within the virtual scene used for calculations from their physical perspectives.

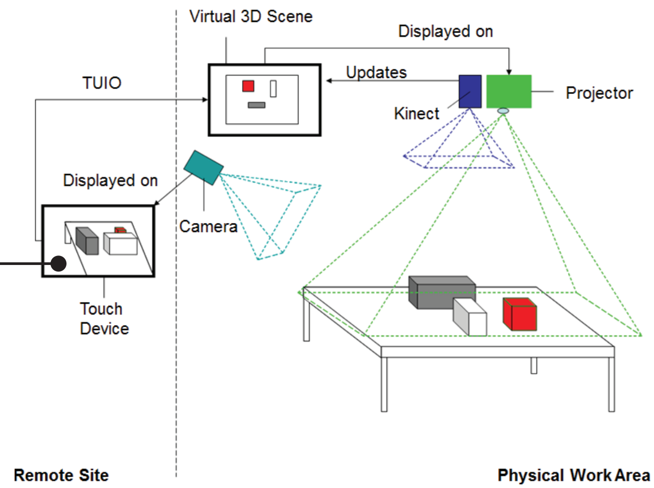

Figure 2: An overview of our system and interactions.

Object Based Remote Guidance

By bringing 3D multi-touch interaction techniques to SAR-enabled remote guidance, we aim to provide a interface for the remote expert that is more akin to 'direct manipulation' [4] than previous 'freehand'. We propose to let users touch objects on a video feed and then specify actions and properties on an object-specific level. For example, a wheel shaped object might only move around a single axis of rotation while a ball could rotate freely. Knowledge of the overall scene also allows for physical constraints to be considered such as limiting the movement of a box to remain on the table. The properties of a given object can be used to automatically generate relevant SAR visual cues to convey the desired action intended by the expert.

In order for this to be possible, a virtual world must first be constructed to mirror the real world and it must be updated based on the physical manipulations of the worker. For each virtual object mirroring the real world, a manipulable proxy clone is also created in the virtual scene. This proxy is effectively used by the remote expert to specify the 'end goal' of a physical manipulation for the worker.

Proxy objects in the virtual world can be interacted with through the use of multi-touch gestures. Multi-touch affords the simultaneous manipulation of multiple degrees of freedom, i.e. users are able to simultaneously pan, rotate or draw without using extra controls to change modes. The knowledge of the environment also allows movements to be constrained in relation to other objects to reflect physical properties of the real world. For example, if a remote expert tries to 'drag' a box that is sitting on a table, the 3D location of the box's proxy can be constrained to only slide along the table.

As soon as the location or orientation of a proxy differs from its respective physical object, SAR visual cues are generated to assist the worker in manipulating the physical object accordingly. The system will compare the relative pose between the two objects and indicate the type of transformation that is required between them.

SAR systems typically only permit projection of graphics onto the surfaces of physical objects. In sketching based SAR systems, it may be hard for the helper to visualize a meaningful 2D projection for some manipulations (such as rotations) and from some viewing angles (such as anything other than directly overhead). The proposed system deals with this issue by automatically generating the visualisations.

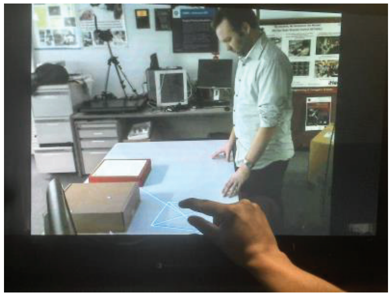

Figure 3a: Expert specifies a rotation and translation of the box.

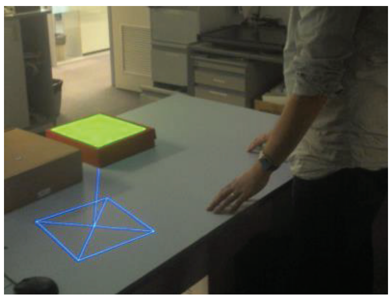

Figure 3b: Rotated and translated visual cue from workers' perspective.

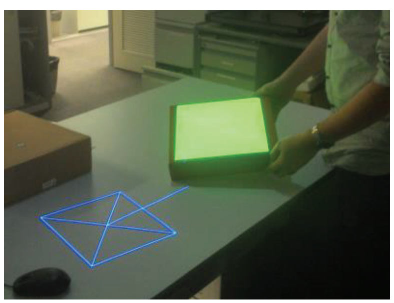

Figure 3c: As the worker moves the box, the SAR cue is updated.

Prototype Implementation

We have built a prototype system that demonstrates an instance of Object Based Remote Guidance. As shown in Figure 2, our setup is implemented with a Kinect depth camera, a projector and an external camera. The remote expert is provided with a multi-touch screen.

The Kinect depth camera is used to track objects in the physical world. A Kinect depth map is used to create a 3D point cloud that is then split into connected component clusters. Each cluster is refined using RANSAC plane fitting and down sampled into a tracked box within the virtual scene. Each box is given a unique id and their previous location is used differentiate between geometrically similar boxes. An initial pass initializes the virtual scene as well as a proxy for each object while future depth map frames update the locations.

The Kinect camera, external camera and projector are pre-calibrated to obtain their intrinsic data and the relative pose between them. A model of each is then represented as viewpoints within the virtual scene constructed in H3D[10]. H3D is also used for rendering the SAR visual cues via the projector.

The external camera feed is separately displayed in full on a remote computer with a 27" Perceptive Pixel touch screen. 2D touch data from the touch screen is sent over the network using the TUIO[5] protocol to the worker's location where it is processed and ray casted into the 3D virtual scene from the perspective of the video camera model.

The touch data is also processed into translation, rotation, selection or annotation gestures. By touching an object in the video feed with one finger, the object is 'selected' and a coloured light is projected onto the object. Dragging the finger along the screen performs a direct translation of the proxy object relative to the surface it is resting on whilst remaining under the fingertips of the helper. By constraining the movement to a surface plane, objects in the virtual world are prevented from translating outside of the projection area. Since the virtual world is hidden from the expert, the visual feedback to the expert is, in fact, exactly the same as the augmented environment provided to the worker. This enforces a strict shared view of the workplace.

In our current system, rotations are relative movements that involve two finger drags and the associated cues are projected in a similar manner to translations as shown in Figure 3a-c.

We expect that, for some tasks, sketch based remote guidance will still be useful and therefore we have aimed for our object-based implementation to be compatible with sketching. In our prototype, sketching is implemented in a similar way to the Sticky Light system [1], in which annotations remain on the physical objects even if they move around.

Conclusions and Future Work

We have introduced the concept of object based manipulation for spatially augmented remote guidance. We have implemented a prototype system that allows an expert to use multi-touch gestures to translate, rotate and annotate an object on a video feed causing meaningful projections into the real world.

This new approach raises a number of new research questions. There is a need to explore which spatial manipulation constraints in the virtual world lead to more efficient remote guidance. Also, we chose visual cues previously shown to be effective in guiding a single user to perform a task. A separate set of cues may be more effective when considering both the worker and the view provided to the remote expert.

Acknowledgements
We would like to thank Stuart Anderson and Chris Gunn for their support and advice.

References
[1] Gunn, C. and Adcock, M. Using sticky light technology for projected guidance. *Proceedings of the 23rd Australian Computer-Human Interaction Conference on - OzCHI '11*, ACM Press (2011), 140–143.

[2] Henderson, S.J. and Feiner, S.K. Augmented reality in the psychomotor phase of a procedural task. *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, (2011), 191–200.

[3] Hiura, S., Tojo, K., and Inokuchi, S. 3-D tele-direction interface using video projector. *ACM SIGGRAPH 2003 Sketches & Applications*, ACM (2003), 1.

[4] Hutchins, E., Hollan, J., and Norman, D. Direct Manipulation Interfaces. *Human-Computer Interaction 1*, 4 (1985), 311–338.

[5] Kaltenbrunner, M., Bovermann, T., Bencina, R., and Costanza, E. TUIO: A protocol for table-top tangible user interfaces. *Proc. of the The 6th Int'l Workshop on Gesture in Human-Computer Interaction and Simulation*, (2005).

[6] Kirk, D. and Stanton Fraser, D. Comparing remote gesture technologies for supporting collaborative physical tasks. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, ACM Press (2006), 1191.

[7] Kondo, D., Kurosaki, K., Iizuka, H., Ando, H., and Maeda, T. View sharing system for motion transmission. *Proceedings of the 2nd Augmented Human International Conference on - AH '11*, c (2011), 1–4.

[8] Liu, J., Au, O.K.-C., Fu, H., and Tai, C.-L. Two-Finger Gestures for 6DOF Manipulation of 3D Objects. *Computer Graphics Forum 31*, 7pt1 (2012), 2047–2055.

[9] Rosenthal, S. and Kane, S. Augmenting on-screen instructions with micro-projected guides: when it works, and when it fails. *Proceedings of the 12th ACM international conference on Ubiquitous computing*, (2010).

[10] SenseGraphics AB. H3D API. *http://www.h3d.org*, .

[11] Sodhi, R., Benko, H., and Wilson, A. Lightguide: projected visualizations for hand movement guidance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (2012).

[12] Suenaga, T., Umeda, T., Kuroda, T., Oshiro, O., and Chihara, K. A tele-instruction system for ultrasound tele-diagnosis. *Proceedings of the International Conference Artificial Reality and Tele-existence, Tokyo, Japan*, (1999), 84–91.

[13] Tecchia, F., Alem, L., and Huang, W. 3D Helping Hands: a gesture based MR system for remote collaboration. *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '12) 1*, 212 (2012), 323–328.

[14] Tsimeris, J.A. Visual Cues for the Instructed Arrangement of Physical Objects Using Spatial Augmented Reality (SAR). *University of South Australia*, (2010).