HandsIn3D: supporting remote guidance with immersive virtual environments

Weidong Huang[1], Leila Alem[1] and Franco Tecchia[2]

[1] CSIRO, Australia

[2] PERCRO - Scuola Superiore Sant'Anna, Italy

Tony.Huang@csiro.au, Leila.Alem@csiro.au, f.tecchia@sssup.it

Abstract. A collaboration scenario involving a remote helper guiding in real time a local worker in performing a task on physical objects is common in a wide range of industries including health, mining and manufacturing. An established ICT approach to supporting this type of collaboration is to provide a shared visual space and some form of remote gesture. The shared space and remote gesture are generally presented in a 2D video form. Recent research in tele-presence has indicated that technologies that support co-presence and immersion not only improve the process of collaboration but also improve spatial awareness of the remote participant. We therefore propose a novel approach to developing a 3D system based on a 3D shared space and 3D hand gestures. A proof of concept system for remote guidance called HandsIn3D has been developed. This system uses a head tracked stereoscopic HMD that allows the helper to be immersed in the virtual 3D space of the worker's workspace. The system captures in 3D the hands of the helper and fuses the hands into the shared workspace. This paper introduces HandsIn3D and presents a user study to demonstrate the feasibility of our approach.

Keywords: remote collaboration; co-presence, mixed reality, hand gesture, shared visual space.

1   Introduction

It is quite common nowadays for two or more geographically distributed collaborators to work together in order to perform actions on physical objects in the real world. For example, one remote expert might be assisting an onsite maintenance operator in repairing a piece of equipment. Such collaboration scenarios are highly asymmetrical: the onsite operator is co-located with the machine being manipulated or fixed but does not have the required expertise to do the job, while the remote expert does not have physical access to the machine but knows how to trouble shoot and fix it. This type of collaboration scenarios is common in many domains such as manufacturing, education, tele-health and mining.

When co-located, collaborators share common ground, thus being able to constantly use hands gestures to clarify and ground their messages while

communicating with each other verbally. However, when collaborators are geographically distributed, such common ground no longer exists, resulting in them no being able to communicate the same way as they do when co-located. Prior research has shown that providing shared visual spaces and supporting remote gesture can help to build common ground [2, 3]. A shared visual space is one where collaborators can see the same objects at roughly the same time. As a result, a number of remote guiding systems have been reported in the literature to achieve these two goals. While how remote gesture is supported may differ from system to system (e.g., [1, 4]), the shared visual space is generally provided in the 2D format of either video feeds or projection on surfaces.

A recent study on a remote guidance system by Huang and Alem [5] indicated that with 2D shared spaces, helpers had difficulties in perceiving spatial relation of objects. Helpers also had a relatively lower sense of co-presence [6]. Spatial understanding is critical for helpers to make right judgements on objects and guide workers accordingly, while co-presence has been shown to be associated with user experience and task performance [7]. Therefore, these are two important factors and should be addressed properly.

Research has shown that immersive virtual environments (IVEs) help improve spatial understanding [9]. Further, IVEs also bring other benefits [8], such as higher sense of co-presence, improved spatial awareness, more accurate cognitive transfer between simulation and reality and better task performance. Although they have been shown useful in supporting general tele-collaboration in which all collaborators work within the same virtual environment, we wonder if IVEs still help in the context of remote guidance.

We therefore propose a new approach that provides 3D shared visual spaces. A prototype system called HandsIn3D has been developed for the purpose of the proof of concept (see [10] for more details). This system uses a head tracked stereoscopic HMD that allows the helper to be immersed and perform guidance in the virtual 3D space of the worker's workspace. In the remainder of this paper, we introduce HandsIn3D and present a user study of it.

2   HandsIn3D

HandsIn3D is currently running on a single PC. It has two logical sections: the worker space and the helper space (see Figure 1). The worker performs a physical task at the worker space, while the helper provides guidance to the worker at the helper space. The left image shows the layout of the worker space. A user sits at the desk performing a task on physical objects (for example, assembly of toy blocks). A 3D camera is mounted overhead to capture the workspace in front of the user including the hands of the user and objects. A LCD non-stereoscopic monitor is placed on the desk to display the 3D view of the workspace augmented by guiding information.

The right image of Figure 1 shows the layout of the helper space. In this space, there is a 3D camera that captures the hands of the helper. The helper wears a stereoscopic HMD and sits in front of an optical head tracker. The HMD allows a realistic virtual immersion in the 3D space captured by the camera placed in the worker space, while the tracker tracks the HMD position and orientation.
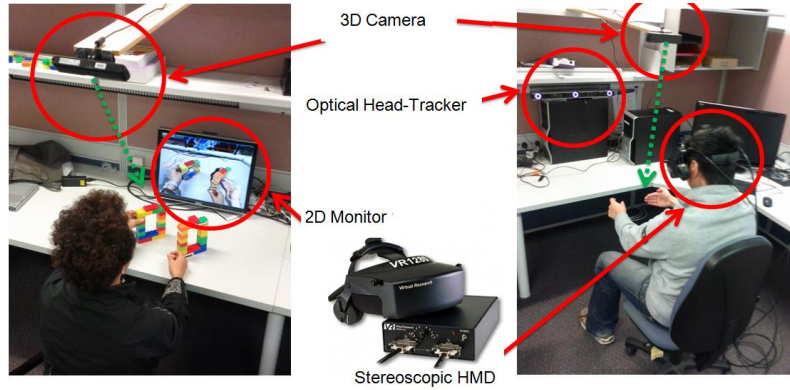
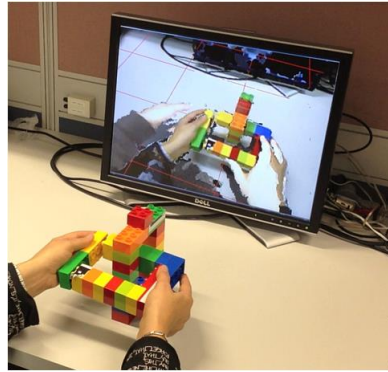**Fig. 1.** Worker space (left) and helper space (right)

Fig. 2. The *shared virtual interaction space* is shown in the LCD monitor. The 3D meshes captured by two cameras are co-located and fused together. Four hands can be spotted in the virtual scene: 2 from the worker and 2 from the helper [10].

The system functions as follows. On the worker side, the worker talks to the helper, looks at the visual aids on the screen, picks up and performs actions on the objects. On the helper side, the helper wears the HMD, looks into the virtual space (and looks around the space if necessary), talks to the worker and guides him by performing hand gesture such as pointing to an object and forming a shape with two hands. During the process of interaction, the camera on the worker side captures the workspace in front of the worker, while the camera on the helper side captures the hands of the helper. The acquired 3D scene data from both sides are fused in real time to form a single common workspace which we call *shared virtual interaction space*. This space is displayed in HMD. The image in Figure 2 is provided to help understand what is presented to the helper during the task execution: by fusing together the 3D meshes acquired by the two cameras, an augmented view of the workspace where the hands of the helper are co-located in the worker space is synthetically created, as shown in the LCD monitor. On the other hand, being

presented with this augmented view, the worker can easily mimic the movements of the helper hands and perform the Lego assembly task accordingly.

The main features of the system therefore include the following:

- Users can speak to each other.
- The helper can see the workspace of the worker via the shared virtual interaction space.
- The helper can perform hand gestures.
- The worker can see the hand gestures of the helper on the screen.
- The two hands of the worker are freed for manipulation of physical objects.

In addition, the shared virtual interaction space also implements features to improve the sense of 3D immersion for the helper. These features include 1) the objects and helper hands cast shadows in the space; 2) the HMD is tracked which allows the helper to see the space from different angles.

3   A User Study

The user study was conducted to evaluate our 3D gesture based interaction paradigm. We were particularly interested in how helpers felt about the 3D user interface.

3.1   Method

Fourteen participants who had no prior knowledge of the system were recruited. Upon their agreement to participate, they were randomly grouped as pairs to perform a collaborative task. In this study, we used the assembly of Lego toy blocks as our experimental task, which is considered as representative of real world physical tasks and has been used for the similar studies. During the task, the worker was asked to assemble the Lego toys into a reasonably complex model under the instruction of the helper. The helper was instructed that he could provide verbal and gestural instructions to the worker at any time. The worker, on the other hand, had no idea about what steps were needed to complete the task.

In order to give users a better appreciation of our new 3D interface in relation to different design options, following the assembly task, the pair was given the opportunity to explore and experience different levels of immersion: 1) no stereoscopic vision, no head tracking and no hands shadow (2D interface), 2) stereoscopic vision, no head tracking and no hands shadow, 3) stereoscopic vision, head tracking and no hands shadow, and 4) stereoscopic vision, head tracking and hands shadow (full 3D interface). This last feature is one that was implemented in HandsIn3D and that participants used in the guiding task at the start of the trial.

Participants were required to complete worker and helper questionnaires after the assembly tasks. These questionnaires asked participants to rate a set of usability metrics, answer some open questions and share their experience of using the system. The usability measures include both commonly used ones and those specific to the system. For more details, see the Results subsection.

3.2 Procedure

The study was conducted a pair after another in a meeting room and was observed by an experimenter. The helper space and worker space were separated in a reasonable distance by a dividing board so that two participants could not see each other. Upon arrival, they were randomly assigned helper and worker roles and informed about the procedure of the study. The helper interface and the worker interface were introduced. They were also given the chance to get familiar with the system and try out the equipment. Then the helper was taken to an office room where he/she was shown a model that needed to be constructed. The helper was given time to think about and plan how to do it and remember the steps.

Then the helper went back to the experimental room and put the HMD on and the experiment started. After the assembly task, the pair of participants was asked to experience the different interface features listed in the last subsection in an informal style. The switch between the interface features was controlled by the experimenter. During the process, the participants were told which feature the system was using. They could play with the toy blocks and talk to each other about the assembly steps. But they were not allowed to comment and share how they felt about the system and the features. This was to ensure that their later responses to the questionnaires were not affected by each other's partner and were of their own. After going through all four features, each participant was asked to fill the helper or worker questionnaire for the role played. Then the participants switched roles and the above process was repeated again. Note that this time the model to be constructed was different but with a similar level of complexity.

After finishing the assembly tasks and questionnaires, participants were debriefed about the purposes of the study, followed by a semi-structured interview. They were encouraged to share their experiences, comments on the system, ask questions and suggest improvements. The whole session took about one hour on average.

3.3 Results and Discussion

Observations. It was observed at the beginning, some participants were very shy about wearing a HMD, resulting in very few head movements. Participants needed prompting and encouragement in order to start moving their head around and change their field of view. This indicates that users may need to take some time to get used to system, as one user commented: "It took me about 10 seconds to adapt to the 3D viewpoints. But after that everything is fine."

Apart from this, all pairs of the participants were able to complete their assigned tasks without apparent difficulties. Their communications seemed smooth. Both helpers and workers looked comfortable performing tasks with the system. More specifically, workers were able to follow the verbal instructions from helpers and understand what they were asked to do by looking at the visual aids shown on the screen. Helpers were able to guide workers through the task process with the HMD worn on their head and using hand gestures.

Usability Ratings. Fourteen participants filled two questionnaires each: the helper questionnaire and the worker questionnaire. We had 28 responses in total. A set of usability measures were rated based on a scale of 1 to 7 with 1 being strongly negative, 7 being strongly positive and 4 being neutral; the higher the rating is, the better the usability. The average ratings are illustrated in Figure 3. Note that 1) helpers had two extra items to rate: perception of spatial relationship between objects and sense of immersion; 2) due to the space limitations, we only report here the ratings of the full 3D system.
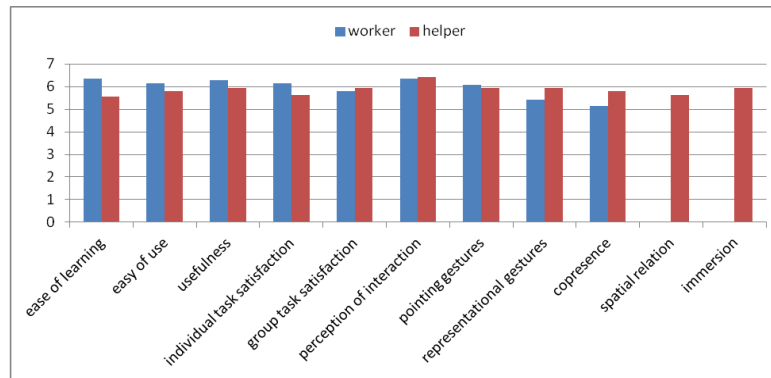

Fig. 3. Average user ratings

As can be seen from Figure 3, despite the slight variations between helpers and workers and across usability items, all items were rated greater than 4, indicating that participants were generally positive about the system. Further, helpers rated the system relatively low for its learning and usage compared to workers. While the system made workers more satisfied with their own individual performance, helpers were more satisfied with the overall group performance. In addition, while helpers gave the same rating for being able to perform both pointing and representational gestures, workers seemed to perceive pointing gestures more easily than representational gestures.

In regard to co-presence, both ratings were over 5, which were higher than those reported by Huang and Alem [6] (just above 4). This indicates that our 3D system offered a higher sense of "being together" for participants. When compared to workers, helpers reported a relatively higher sense of co-presence. Helpers also had positive ratings for perception of object spatial relation and sense of immersion. All these indicate that our 3D design approach did work as expected.

User Experiences. Based on user responses to the open questions and user interviews, participants were generally positive about the system, as one participant stated that "it is very impressive and a great experience to use this system."

More specifically, participants appreciated the feature that workers are able to see and helpers are able to perform hand gestures. A helper commented that "he (the worker) knew exactly what I meant by 'here', 'this one' and 'that one'." A number of

workers simply commented that "(hand gestures were) easy to understand and follow."

Consistent with the usability ratings, the 3D interface has boosted a strong sense of co-presence and immersion for helpers. It was commented that the system had given participants a feeling of being in front of the remote workspace and co-presenting with the remote objects and their remote partner. Comments from helpers include "I feel I was right there at the screen and really wanted to grab the objects." and "I can feel that her hand was in my way, or my hand was in her way. So in this sense, I felt we were in the same space." A few workers also commented that seeing both hands in the field and using words like "this" and "that" during the conversation made a strong visual impression and physical presence.

User comments also provide further evidence that the 3D interface improved perception of spatial relation and participants appreciated that. For example, "You can see the difference between 2 objects with the same base but different heights in 3D." "3D helped to see how the final shape looked. With 2D, I had to ask the worker to show me the piece in another angle." "It gives the depth of the objects, so remote guiding could be easier in some cases."

Participants generally liked the idea of having shadows of hand and objects, commenting that it would be easier to point and gesture in the remote workspace as hand shadows could provide them with an indication of hand location in relation to the remote objects. However, there were mixed responses when participants were asked whether the shadow feature actually helped. For example, "Yes, it helps. It makes a good stimulation effect. So I can do better communication with my partner." "The shadow helps me feel that the view is in 3D. But I think I can still understand without the shadow." "No, there are some real shadows in grey color. The black shadow is artificial and a little bit annoying." "The shadow could sometime cover objects and I think this could potentially lead to something wrong (maybe a transparent shadow)." "Yes, (shadow helps) for pointing only, but not much on rotating etc."

In regard to head tracking, participants commented that it enables helpers to see more of the way that blocks are connected without needing their partner to rotate them and that it makes workers aware of what helpers are looking at.

Further, in comparison with the 2D interface, participants commented that 2D is fine with simple tasks, but 3D offers much richer user experience and is more preferable and useful for complex procedures when a high level of user engagement is required. For example, "3D is more realistic as I can see all angles." "In 2D, it seems like playing a game. When changing my viewpoints into 3D, I got a feeling of going back to real world." "(3D) helps more when I need to give instruction related to space concept." "3D interface makes it easy to match the screen and the physical objects." "3D feels real. 2D interface is enough for simple tasks but 3D interface helps more when the task gets more complicated."

Although the main purpose was to test the usability and usefulness of our 3D concept for remote guidance on physical tasks, user comments also gave some hints for further improvements and more features. These include 1) use a lighter and more easily adjustable helmet; 2) increase image quality and resolutions; 3) differentiate

worker and helper hands by color and make them transparent; 4) provide a more dynamic and more immersive environment for helpers to interact with, such as when the helper moves closer to the objects, they become bigger; 5) enable helpers to have access to manuals while guiding; 6) make shadows grey and transparent.

4   Concluding Remarks

Our user study has showed that the proposed 3D immersive interface is helpful for improving perception of spatial relation of helpers and sense of co-presence and that the system is generally useful and usable, particularly more so for complex tasks.

The study also points out some future research and development directions. We plan to advance the prototype into a close-to-production system so that we can test it in a more realistic setting. For example, separate the two sides of the system and connect them through the internet, instead of hosting them by the same PC. We also plan to compare HandsIn3D with its 2D versions through rigorously controlled studies so that we can have more quantitative and objective information about the benefits of immersive virtual environments in supporting remote guidance.

References

1. Alem, L., Tecchia, F. and Huang, W. (2011) HandsOnVideo: Towards a gesture based mobile AR system for remote collaboration. In Recent Trends of Mobile Collaborative Augmented Reality. 127-138
2. Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E. and Kramer, A.D.I. (2004) Gestures over video streams to support remote collaboration on physical tasks. Human-Computer Interaction, 19: 273-309.
3. Gergle, D., Kraut, R.E. and Fussell, S.R. (2013) Using Visual Information for Grounding and Awareness in Collaborative Tasks. Human-Computer Interaction, 28: 1-39.
4. Gurevich, P., Lanir, J., Cohen, B. and Stone, R. (2011) TeleAdvisor: a versatile augmented reality tool for remote assistance. In CHI'11, 619-622.
5. Huang, W. and Alem, L. (2013) Gesturing in the Air: Supporting Full Mobility in Remote Collaboration on Physical Tasks. Journal of Universal Computer Science.
6. Huang, W. and Alem, L. (2011) Supporting Hand Gestures in Mobile Remote Collaboration: A Usability Evaluation. In Proceedings of the 25th BCS Conference on Human Computer Interaction.
7. Kraut, R. E., Gergle, D. and Fussell, S.R. (2002) The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence. In CSCW'02, 31-40.
8. Mortensen, J., Vinayagamoorthy, V., Slater, M., Steed, A., Lok, B. and Whitton, M.C. (2002) Collaboration in tele-immersive environments. In EGVE'02, 93-101.
9. Schuchardt, P. and Bowman, D.A. (2007) The benefits of immersion for spatial understanding of complex underground cave systems. In VRST'07, 121-124.
10. Tecchia, F., Alem, L. and Huang, W. (2012) 3D helping hands: a gesture based MR system for remote collaboration. In VRCAI'12, 323-328