Poster: 3D Referencing for Remote Task Assistance in Augmented Reality

Ohan Oda*         Mengu Sukan*         Steven Feiner*         Barbara Tversky†
Columbia University

ABSTRACT

We present a 3D referencing technique tailored for remote maintenance tasks in augmented reality. The goal is to improve the accuracy and efficiency with which a remote expert can point out a real physical object at a local site to a technician at that site. In a typical referencing task, the remote expert selects a local viewpoint, and the technician navigates to a location from which the target can be viewed, and then to attend to that object. The expert and technician both wear head-tracked, stereo, see-through, head-worn displays, the expert’s hands are tracked by a set of depth cameras. The remote expert first selects one of a set of prerecorded viewpoints of the local site, and a representation of that viewpoint is presented to the technician to help them navigate to the correct position and orientation. The expert then uses hand gestures to indicate the target.

Keywords: Collaborative mixed/augmented reality, referencing technique, remote task assistance, maintenance and repair.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques; K.4.3 [Computers and Society]: Organizational Impacts—Computer-Supported Collaborative Work

1 INTRODUCTION

For participants to collaborate successfully on a spatial/action task, they must be able to generate and interpret effective reference cues. Successful spatial/action coordination is challenged when participants are not colocated and cannot directly gesture in each other’s environments. When spatial communication relies only on language, errors occur even when communication partners think they have agreed [5].

Spatial/action coordination is critical in remote task assistance in which an expert assists someone with less knowledge in that domain (e.g., a technician, in an equipment maintenance domain) and both are in physically separate environments. The remote expert must explain to the technician task procedures involving objects in the task space local to the technician who performs the procedures. Thus, the efficiency and effectiveness of remote task assistance depends heavily on the referencing techniques used.

Referencing in remote task assistance involves two successive steps: localization and selection. In localization, the expert must help the technician navigate to a destination location in the task space where a target object can be clearly seen by the technician. (Note that this step may not be necessary if the task environment is sufficiently simple.) This process can be particularly challenging if the task space is complicated and the target is visible only from a small set of head positions and view angles. Once the target is within the technician’s view, selection occurs when the expert indicates the target to the technician (e.g., by pointing at it).

To address these problems, we are developing an augmented reality referencing technique (Figure 1), based on 3D gestural interaction. The technique is intended for situations in which both the expert and technician wear head-tracked, stereo, see-through, head-worn displays (HWDs), the technician’s head-worn stereo camera views are transmitted to the expert, and the expert’s hands are tracked. Our initial task domain is aircraft engine maintenance.

2 RELATED WORK

Researchers have explored a variety of approaches for supporting remote task assistance. Kirk and Fraser directly embedded the gestural output of an expert in the local view of a technician captured from live overhead video [6], while Goto et al. presented a similar technique using prerecorded video [3]. These techniques may work well on tasks performed on a flat surface, but could be challenging to apply to more complex environments. Kurata et al. [7] demonstrated a system in which the local technician wears a camera and a remote-controlled laser pointer on his shoulder instead of an HWD. The remote expert sees the task space through the camera, and points into the physical environment directly with the laser. Gurevich et al. [4] use a portable projector in the local environment to allow a remote expert to project 2D drawings on objects. Gauglitz et al. [2] apply natural feature tracking to allow the expert to virtually annotate a previously unmodeled environment. All of these approaches can be very effective for selection, but do not address localization. Sodhi et al. [8] use a depth camera and projector to project instructions on a user’s hand. They

---

*ohan,mengu,feiner}@cs.columbia.edu  †btversky@stanford.edu
showed that their technique allows users to follow guided translational movement with controlled speed, but cannot effectively guide rotational movement, which is necessary for many maintenance tasks.

3 Referencing Techniques

We are interested in natural 3D hand-gesture interaction similar to that of Stafford et al. [9], but tailored for maintenance, so that the expert can freely instruct the technician how to perform a maintenance task after conveying the specific target on which the task needs to be performed. Current state-of-the-art hand-pose tracking systems based on depth cameras [12] can track simple 3D gestures relatively reliably without gloves or tracking markers. Our technique uses 3D hand gestures for both the localization and selection steps. Gestures are captured by the 3Gear Systems hand tracking software (http://www.threegear.com), which interactively derives an idealized 3D model of the gesturing hand.

3.1 Localization

Localization can be performed by having the expert either specify a route by which the technician should travel to reach the destination, or specify the destination directly and have the technician find her way there. The former method is more desirable if the task environment is complex and a specific route takes the technician to the destination quicker and more safely than without the route. The latter method is more desirable if the task environment is relatively simple and the specific path taken to the destination is unimportant, since defining a route will require more effort from the expert. We assume that the task environment is relatively complex, and focus on having the expert efficiently and accurately guide the technician to view the task space from a constrained head position and orientation.

Our technique relies on a prerecorded set of viewpoints, similar to those of the SnapAR technique [11]. The viewpoints can be recorded prior to the technician’s inspection if a similar inspection was performed previously or on the fly with the expert’s verbal guidance. Once recorded, they are presented to both users.

The expert initially sees the technician’s environment, including representations of the viewpoints, as a virtual world-in-miniature (WIM) [10] overlaid on a turntable that can be turned to orient the WIM as desired (Figure 1a). The expert then indicates the viewpoint to which she wants the technician to navigate by touching its representation with her index finger. (The technician does not see the expert’s finger.) After the viewpoint is selected, the expert will see the technician’s live stereo camera views and can verbally assist the technician, if necessary, to reach the viewpoint.

The technician initially sees a set of viewpoint representations in her environment. Once a viewpoint is selected by the expert, the technician will see only the selected viewpoint representation, which looks like a splayed window frame through which she can look (Figure 1b). If the viewpoint is outside her field of view, an arrow appears in front of her to guide her to orient her head appropriately. Once the technician places her head within the viewpoint representation, the expert freezes [1] their own copy of the technician’s stereo imagery to prepare to refer to the target. (The technician’s stereo view remains live.)

3.2 Selection

During selection, we render a live 3D virtual model of the expert’s tracked hand, as shown in Figure 1(c), in the expert’s frozen copy of the technician’s view. This allows the expert to point directly at the target with their index finger. Simultaneously, the technician sees the live virtual hand model appear relative to the portion of the task domain seen by the expert.

A pilot study revealed that even with a stereo HWD, it can be difficult to perceive the exact depth of the target and touch it precisely. To address this, we added an arrow originating from the tip of the virtual hand’s index finger in the direction along the first joint of the index finger (Figure 1c); the arrow length is automatically adjusted so that the tip of the arrow always touches the surface of the task space. (Here, we rely on a 3D model of the task space, assuming that the task space is known in advance.) The pilot study also found that if the expert freezes the view when the technician is too close to an object, the virtual hand can enter the object when the expert extends his arm, and would then be occluded. To avoid this, we apply an offset to the virtual hand based on the distance between the view plane of the frozen view and the closest point on the task space within the frozen view frustum.

4 Conclusions and Future Work

We have described a two-step remote spatial referencing system designed to support a remote expert assisting a local technician. The expert first instructs the technician to navigate to a desired location in the task environment by specifying a prerecorded viewpoint and next selects a target object using natural 3D hand gestures. The expert could then use 3D gestural interaction to instruct the technician how to perform an operation on the target object, which we will address in future work. To evaluate our technique, we will conduct a formal study comparing the efficiency and accuracy of our system against other referencing techniques in a remote task assistance domain.

Acknowledgments

This research was funded in part by NSF Grants IIS-0905569 and IIS-0905417. Vuzix generously donated the Wrap 920AR HWDs used for Figure 1. We thank Robert Wang for the 3Gear Systems 3D hand gesture tracking SDK and Steve Henderson for refurbishing and modeling the Rolls-Royce Dart 510 engine.

References