TeleAdvisor: A Versatile Augmented Reality Tool for Remote Assistance

Pavel Gurevich, Joel Lanir, Benjamin Cohen, Ran Stone
IBM Haifa Research Lab
Mt. Carmel, Haifa, 31905, Israel
{pavelg, yoell, cohen, ranst} @ il.ibm.com

ABSTRACT
TeleAdvisor is a novel solution designed to support remote assistance tasks in many real-world scenarios. It consists of a video camera and a small projector mounted at the end of a tele-operated robotic arm. This enables a remote helper to view and interact with the workers' workspace, while controlling the point of view. It also provides the worker with a hands-free transportable device to be placed anywhere in his or her environment. Active tracking of the projection space is used in order to reliably correlate between the camera's view and the projector space.

Author Keywords
Augmented Reality; Mobile projector; Remote assistance.

ACM Classification Keywords
H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – Computer-supported cooperative work;

INTRODUCTION

A remote assistance task around a physical object is characterized by the presence of one or more objects that a person (the worker) focuses on and interacts with to complete the task. The worker lacks some knowledge pertaining to the object or the operation of the task, and a emote expert (the helper) aids the worker to diagnose the problem and provide instructions on how to perform unfamiliar operations. Often, the worker uses other objects such as tools to interact with the focus object. Examples of such tasks may include a customer calling a call center to help troubleshoot a printer, a novice technician calling an expert for guidance in repairing a complex device (e.g., an aircraft engine), or a medical expert guiding a local doctor through a complex procedure.

Studies have shown that a shared visual representation of the workers' space, including the task object, tools being used, and a view of the workers' actions is highly useful in supporting such tasks [2]. Still, video-mediated dialogs were found less efficient than side-by-side ones [3]. This was explained by the fact that the worker has difficulties establishing what visual information was shared, and more importantly, the helper could not use gestures to refer efficiently to task objects [2].

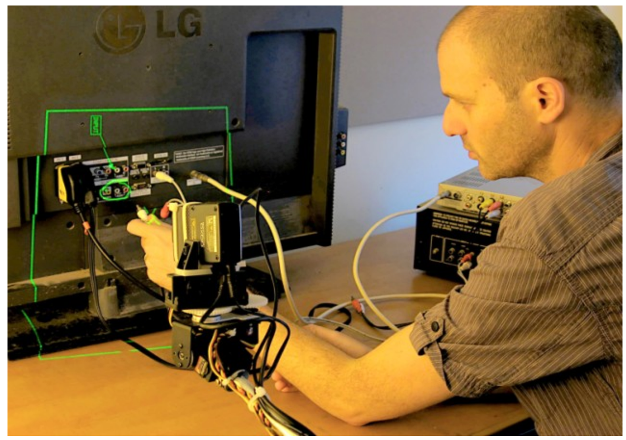


Figure 1. Using TeleAdvisor to remotely assist a person in executing a complex wiring task

A promising method for gesturing combines the gesture representation with the physical objects themselves, mostly using Augmented Reality (AR) tools. Kirk et al. [4, 5] projected the gestures onto the workers' workspace using top-mounted projectors. Top-mounted projectors have a large projection area and can support high-resolution projection, however, they are stationary and are limited to a prearranged work area, on which the projector was mounted prior to the task. Many scenarios require more flexibility, and cannot set the workspace ahead of time. With the advent of pico projectors, it is now possible to create a transportable appliance that the worker would be able to easily situate anywhere in his or her work environment.

We present TeleAdvisor, a remote assistance tool that uses a small video camera alongside a pico projector to capture the worker's environment and project on top of it (Figure 1). TeleAdvisor is designed to be transportable and hands-free, transmitting the captured video and helpers' gestures over the network, so the worker would be able to situate it anywhere needed and use both hands to interact with the object. Furthermore, the camera and projector are mounted on top of a robotic arm to enable the helper fine-level control of the

CHI'12, May 5–10, 2012, Austin, Texas, USA.

view and projection area. Finally, TeleAdvisor is equipped with an enhanced set of annotation tools which include the ability to project a pointer, freehand sketches, text and images. The combination of transportability, hands-free unobtrusive usage, and zero training of the worker, provides a solid basis for the deployment of a versatile ubiquitous appliance that can be used in all types of scenarios, such as working on the back of a washing machine or under a car. In this paper, we describe the technical challenges that stem from the dynamic field of view of TeleAdvisor and discuss the design decisions we have made. We also describe an initial evaluation that we conducted in order to examine the viability of the system.

RELATED WORK

Many studies have examined different mechanisms to allow a remote helper to assist or guide a worker through various tasks involving the use of physical objects. These studies have shown that it is important to enable the helper to view the remote workspace in order to improve coordination and maintain a shared context of the task [3, 10]. Other studies have suggested to further improve the communication between the worker and helper by communicating helpers' gestures in the form of pointers [6], sketches [2] or full images of the helpers' hands [4] to the worker. Early technologies overlaid these gestures on top of the worker's workspace video feed to be sent back and shown to the worker on a display located in the workspace [2, 6]. However, this may cause a fractured ecology in which the worker needs to split attention between the task and the external display [9].

To address this, several studies have looked on using Augmented Reality (AR) solutions for helping remote users perform various tasks. Bauer et al. [1] used a telepointer for supporting field workers undertake technical activities with the aid of a remote expert. GestureCam [6] implements a solution that comprises of camera and a Head Mounted Display (HMD) that are worn by the worker. The camera's image is sent to the helper, whose gestures in front of a display are captured and sent back to the workers' HMD. In addition, the helper can control an actuator with a second camera mounted to top that provides him or her with a static view of the object. A laser pointer is mounted on the actuator to enable remote pointing. O'Neill et al [9] examined a mixed reality solution in which the object representation is 3D model shared by the helper and worker and connected to the physical object itself.

Finally, other studies have used a camera in conjunction with a mobile projector for other tasks [7, 8]. The Sixth sense [8] combined a mobile projector and camera to create a wearable gestural appliance that augments the physical world with digital information and uses hand gestures to interact with that information.

TELEADVISOR

TeleAdvisor includes two remotely located parts: the TeleAdvisor device and the Controller interface. The TeleAdvisor device part (Figure 2) is comprised of a robotic arm on which are mounted a camera and pico projector. The camera captures visual data which is transmitted to the Controller. The projector projects information sent from the Controller on top of objects in the worker's environment. The worker can easily situate TeleAdvisor in his or her environment, directing the camera and projector to the point of need, and can carry on a voice conversation with the helper while having both hands free to work on the object.

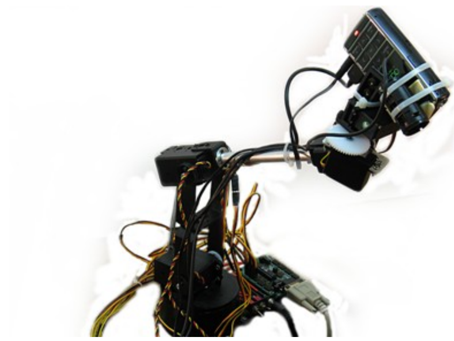


Figure 2. The TeleAdvisor end-point device prototype.

Controller's interface

The Controller interface (Figure 3) is located remotely and its purpose it to enable the helper to view the worker's workspace via the live stream of the TeleAdvisor's camera, and to gesture onto the worker's workspace. In addition, the helper can use the Controller to move the robotic arm in order to change the point of view and projection area. Gestures are made by annotations and graphics that are overlaid on the video stream using the Controller interface. The annotations are sent back to the projector via the network and projected on top of the objects at the worker's location. In addition to using a pointer or free sketched ink in various colors and sizes, the helper can choose a set image (or icon) to be pinpointed to a specific location. For example, the helper can attach a pre-set "rotate clockwise" icon to a screw in the user's environment. The helper can also add a "bubble" in which it is possible to enter text and attach it to an object. For example, the helper can mark a

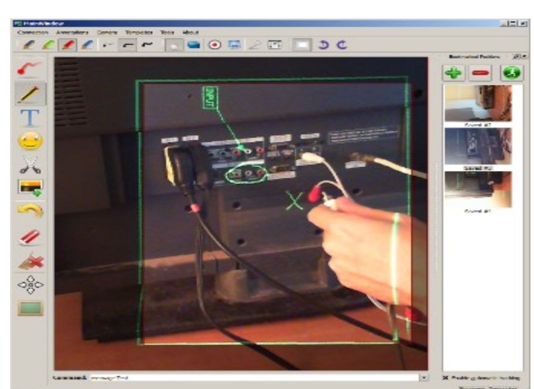


Figure 3. Controller interface (See Figure 1 to view corresponding workspace).

certain object with a bubble with the words "don't touch this".

Stroke annotations that are overlaid on top of the helpers' view are erased after a pen-up event. After that, the real projected annotations are seen by the helper through the video captured view of the worker's workspace. We considered enhancing the helpers' view with the annotation drawings, but after several experimentations decided against it. The main reason was that the annotations did not always align pixel-to-pixel with the real-world projected annotation causing visual clutter and misunderstanding. In addition, we felt it diminished helpers' understanding of what exactly the worker sees and how the annotations are perceived by the worker.

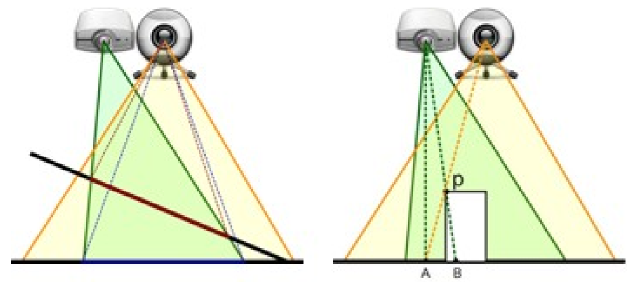
Helper's control of the point of view

A fixed-view camera can provide a good overview image but is fixed and cannot view detail, while a camera mounted on the worker's head constraints the helper's view to what the worker is focused on and can be jittery [6]. Ranjan et al. [10] propose using motion tracking technology to track the workers' hands and move the camera accordingly. Our approach leaves control in the helper's hand by enabling him or her to move the camera's view and to electronically zoom in and out in order to view in more detail.

The TeleAdvisor robotic arm allows for five degrees of freedom (DOF) rotation and azimuth angle. It consists of a base unit, a first arm connected to a second arm with an elbow joint, and a wrist joint at the end of the second arm with a clamp unit that holds the combined video camera and a projector assembly (See Figure 2). Although the robotic arm allows five DOF, its movements are mapped into two dimensions on the Controller interface. We decided to narrow down the movement of the arm to two dimensions in order to match the helper's 2D view, and thus reduce the helpers' cognitive load. From our experimentations, 2 DOF with the addition of electronic zoom were enough to reach a wide area of workspace including views of several planes. Finally, we implemented *positional bookmarks*. The helper at any time can bookmark a view. Later, selecting this bookmark will move the arm back to the exact same position.

Active tracking of the projection space

TeleAdvisor is designed to be transportable. Furthermore, the view position of TeleAdvisor's camera is constantly moving in response to the helper's control. This introduces a challenge to reliably derive the correlation between the camera and the projector space which is needed in order to know where exactly to project the annotation. A parallax is created between the camera and the projector due to the distance between them. This may cause the projected annotation to be inaccurate. While in a stationary solution, this distortion can be calibrated once and fixed accordingly, this is not possible in a mobile solution where the distance between the surface and the camera/projector may be different every time. Moreover, the surface angle compared



(a) **(b)**

Figure 4. (a) Because the distance between the surface and the camera/projector varies in distance and angle, the need for active tracking of the projector area arises. (b) Distortion problem caused by height of object

to the device might not be straight causing even more distortions. Figure 4a illustrates this problem showing the differences in parallax between a straight surface (in blue) and a tilted surface (in red).

To address this, we actively track the effective projector area and accommodate the changes accordingly. The projector continuously projects a green rectangle on the boundary of the projection area, and the system tracks it. The green rectangle also provides the worker with knowledge of the annotation and viewing space of the helper, which is important for the reciprocal communication between helper and worker [2]. For the tracking we used OpenCV primitives to build a robust and fault-tolerant detection algorithm. To alleviate laser flicker due to camera-projection scan asynchronization, we average three image frames back. Several filters are applied to remove extra details from the image and improve fault-tolerance. First, the noise edge-like lines are detected and masked out of the green channel. Then, only lines that are at least half of the frame are preserved. Second, space separation criterion is examined - four corner points should be sufficiently separated in space. Third, a median filter is applied to the individual history of each candidate corner point to avoid glitches. After the frame boundaries are detected, the reverse transformation is calculated and is applied to the live view of the helper. The annotations added by the helper are applied to the forward transformation and appear in the correct place when projected to the worker.

Another type of distortion can be caused when projecting on 3D objects because of the height of the object (Figure 4b). The correction described earlier is efficient for planar objects but can introduce distortions when the target object (p) to project on is significantly non planar with respect to the correction plane (black line). The correction procedure will result in highlighting point A instead of P. Currently, we do not have an automatic solution for this problem. However, in our experimentations, we saw that the distortion caused by the height of the objects do not exceed a few millimeters. TeleAdvisor allows the helper to move an existing annotation a few pixels in each direction to correct these kinds of distortions. Future solutions may examine the use of depth perception cameras to include the depth of the object.

INITIAL EVALUATION

We conducted an initial user evaluation of TeleAdvisor to examine the use of the software and its features and assess user acceptance. The purpose of the evaluation was to test TeleAdvisor and get feedback on the usability and usefulness of such a tool. We therefore performed a qualitative study rather than a comparative one.

We recruited three participants to be helpers. Helpers were first fully trained to use the Controller interface, after which they got full training performing the tasks themselves. We recruited 10 participants (7 female) for the role of workers. Each helper worked with 3-4 workers for an experimental session. Workers did not get any initial training. Two tasks were given. The first was a Lego assembly task which is common in similar previous works [2, 4, 5]. The task consisted of connecting small pieces of Lego situated on one plane (the table), where the helper held the manual for the given model while the worker constructed the model using the helpers' instructions. The second task was designed to be more realistic. The task consisted of connecting various cables between a TV set, a DVD and an AV receiver box. The TV and DVD were set perpendicular to each other and to the table (see Figure 1) to create a realistic 3D scenario in 3 planes. The worker and helper sat in the same room enabling regular voice communication. They were separated using partitions to ensure they only had visual access to each other through TeleAdvisor.

Observations indicated that helpers were able to effectively use TeleAdvisor to complete the tasks. Similar to [5], helpers repeatedly used annotations, accompanying them with phrases such as: "put it *here*" and "move it *there*" to establish common ground using deictic references. The active tracking worked well, and almost no glitches or distortions were observed. Helpers mostly used the free sketch tool for annotation. Pointer was used a few times while text or icons were not used at all. Helpers explained that sketches were most useful for deictic gestures, highlighting a specific object or location (usually by circling it), while pointers were useful for more accurate and transient gestures. Helpers heavily relied on the movement of the robotic arm since both tasks required a wider point of view of the workspace than the camera provided. For both tasks, helpers easily managed to view the entire workspace including all three planes of the second task, supporting our approach of providing only 2DOF movements. All three helpers manually moved the viewpoint as well as used the position bookmarks to mark and return to specific areas. They all indicated in post-interviews that bookmarks were extremely useful. All workers reported that TeleAdvisor was intuitive, very useful, and saw much value in having projected annotations in such a setting. Workers easily understood the novel interaction paradigm. This is probably due to the fact that they did not need to operate anything with the device which seamlessly integrated in the environment. Workers also reported that the visible rectangle helped them be aware of the helpers' view. Most workers said that using the system was as effective as having a person giving instructions in the same room and one person even said that it was better, since it was less stressful.

CONCLUSION

Many remote assistance and remote collaboration tasks require flexibility and mobility at the workspace. We showed that it is possible to create a tool that would effectively support remote gesturing while being transportable, and robust. Future work may look at anchoring annotations to specific locations using computer vision techniques allowing the object or the device to move while keeping the annotations in the same relative location.

REFERENCES

1. Bauer, M., Kortuem, G., Segall, Z. Where are you pointing at? A study of remote collaboration in a wearable videoconference system. International Symposium on Wearable computers (1999).
2. Fussell, A. R., Setlock, L. D., Yang J., Ou J., Mauer E., and Kramer, A. D. I. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks, Human-Computer Interaction, 19: 3, 273 - 309 (2004)
3. Gergle, D., Kraut, R.E., and Fussell, S.R. Action as language in a shared visual space. *In Proceedings of. CSCW'04*, ACM Press (2004), 487-496.
4. Kirk, D., Fraser, D. S., Comparing remote gesture technologies for supporting collaborative physical tasks. *In Proc. CHI '06*, ACM Press (2006), 1191-1200.
5. Kirk, D., Rodden, T., Fraser, D. S. Turn it this way: grounding collaborative action with remote gesture. In *Proc. CHI '07*, ACM Press (2007), 1039-1048.
6. Kuzuoka, H., Kosuge, T., & Tanaka, K. GestureCam: A video communication system for sympathetic remote collaboration, *Proceedings of CSCW'94* (35-43). NY: ACM Press. (1994)
7. Linder, N, Maes, P. LuminAR: portable robotic augmented reality interface design and prototype. *In Adjunct proceedings of UIST '10.* ACM, New York, NY, USA, 395-396 (2010).
8. Mistry, P., Maes, P., Chang, L. WUW – wear Ur World: a wearable gestural interface. *In proceedings of extended abstracts of CHI'09*. ACM Press. (2009)
9. O'Neill J., Castellani S., Roulland F., Hairon N., Juliano C, Dai L. From ethnographic study to mixed reality: a remote collaborative troubleshooting system. *Proceedings of CSCW'11*, ACM Press. 225-234. (2011)
10. Ranjan, A., Birnholtz, J. P., Balakrishnan, R., Dynamic Shared Visual Spaces: Experimenting with Automatic Camera Control in a Remote Repair Task. *In Proceedings of CHI 2007*, ACM Press (2007)