A Web-Based Remote Assistance System with Gravity-Aware 3D Hand Gesture Visualization

Chelhwon Kim
FX Palo Alto Laboratory
kim@fxpal.com

Patrick Chiu
FX Palo Alto Laboratory
chiu@fxpal.com

Yulius Tjahjadi
FX Palo Alto Laboratory
yulius@fxpal.com

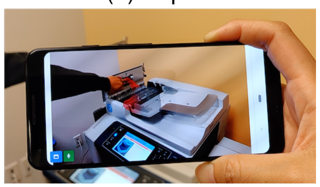
Every submission will be assigned their own unique DOI string to be included here.

Abstract
We present a remote assistance system that enables a remotely located expert to provide guidance using hand gestures to a customer who performs a physical task in a different location. The system is built on top of a web-based real-time media communication framework which allows the customer to use a commodity smartphone to send a live video feed to the expert, from which the expert can see the view of the customer's workspace and can show his/her hand gestures over the video in real-time. The expert's hand gesture is captured with a hand tracking device and visualized with a rigged 3D hand model on the live video feed. The system can be accessed via a web browser, and it does not require any app software to be installed on the customer's device. Our system supports various types of devices including smartphone, tablet, desktop PC, and smart glass. To improve the collaboration experience, the system provides a novel gravity-aware hand visualization technique.

Author Keywords
Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; See [http://acm.org/about/class/1998/]: for full list of ACM classifiers. This section is required.

(a) Expert



(b) Customer (Smartphone)

(c) Customer (Smart Glass)

Figure 1: System setup. (a) A remote expert watches a live video feed sent from a customer and move his/her hand over a Leap Motion controller to provide hand gesture based guidance. (b) A customer uses a smartphone to share the view of his/her work space and an object (printer) therein and is instructed by viewing the expert's hand movement on the smartphone's screen. (c) A setup with smart glasses (Epson Moverio BT-300).

Introduction

Current video conference technology allows people to keep collaborating even when they are in different locations. However, in a remote assistance scenario, where a remote expert helps a local customer who performs a complex physical task, the guidance conveyed by using video and audio communication media still has a high risk of misinterpreting the expert's intention and instruction, which can lead to inefficient collaboration experience and poor performance.

Researchers have found that non-verbal communication such as body language and hand gestures can significantly improve the performance of activities in remote assistance/collaboration scenarios [5, 4]. In this paper, we propose a lightweight web-based system that supports remote guidance using hand gestures from a remotely located expert to a local customer. Leveraging web-based real-time media communication technology (e.g. WebRTC), the system allows the local customer to use a commodity smartphone device [1] to share a view of the customer's workspace with the remote expert helper via a video conference, and the expert's hand gesture data can be transmitted and visualized in real-time on the shared live video stream in the web browser. See Figure 1. Since the system can be accessible via standard modern web browsers from end users by simply going to the specified web page, it does not require any prerequisite app software to be installed on user's smartphone devices. Only expert side requires a hand tracking device to track the expert's hand movement.

The effectiveness of using the hand gestural visual cue in remote assistance/collaboration scenarios has been studied well in previous research work [2, 1, 3, 6]. However, most of these systems require the installation of apps or plug-ins running on complex customized equipment from the customer's side. Those apps were developed and tested in specific operating system versions and configurations, and are not easily deployed to different types of device and setups.

Our proposed system is different from other work in terms of the following aspects.

- The convenience of our web-based system: Our system is a web-based system that does not require any prerequisite app to be installed on the customer's device, and can be accessible for use with web browsers (e.g. Chrome, Firefox, Safari).
- Easy customization: The system can be customized for use on any device connected to the internet, such as smartphone, tablet, desktop PC, and smart glass.
- Gravity-aware hand visualization for providing a better interaction experience.

In our demo, we will show our working prototype system running on a various range of devices like smartphone, tablet, and smart glasses. A novel gravity-aware hand visualization will be demonstrated.

A web based remote assistance system

Our system is built on a web real-time communication framework (WebRTC). As depicted in Figure 3, the intermediary signaling server in WebRTC establishes a peer-to-peer (P2P) connection between the customer and the expert for real-time media communication, and each end user can participate in this direct media communication session by simply browsing a web page hosted by the server using any compatible modern web browser (e.g. Chrome, Firefox, Safari etc.). Since the data communication between two peers

[1] Our system supports other types of devices such as tablet, PC, smart glass, etc. See Fig. 1 (b), (c).

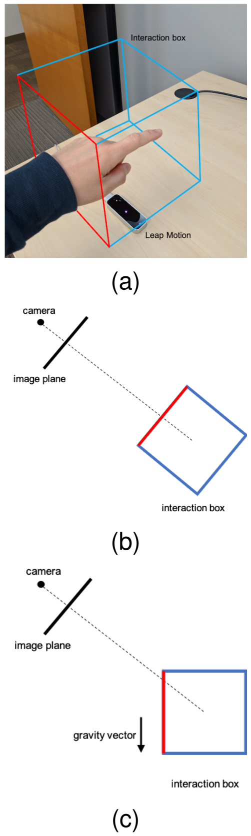


Figure 2: (a) An illustration of the interaction space over the Leap Motion controller in the expert site (b) The typical way of aligning the sensor's interaction space to the customer's camera view space (c) The gravity-aware alignment of the interaction space.

and the data access to the user's devices such as camera, microphone, or the Leap Motion controller are controlled by the WebRTC and JavaScript APIs residing in the user's web browser, it does not require any prerequisite software or plugins to be installed on their devices or browsers.

To create and display the 3D hand model in the user's web browser in real-time, we use Three.js API that is based on WebGL. Three.js uses the JavaScript language as part of the website, and thus it does not require the need for installing any web browser plugins. The hand skeleton joints data obtained from the expert's Leap Motion device is used for rigging the 3D hand model in the Three.js scene. In order to synchronize the perspective view of the rendered 3D hand model and its pose between the customer and the expert, the hand skeleton joints data and the Three.js's perspective camera parameters (e.g. fov, view direction, or position etc.) are set at the expert side, and then transmitted to the customer side (see arrows illustrating the data flow between customer and expert in Fig. 3), where the transmitted hand skeleton joints data and the Three.js camera parameters are used for rigging and displaying the customer's own hand model in the same perspective view on the customer's web browser. Note that the 3D hand model

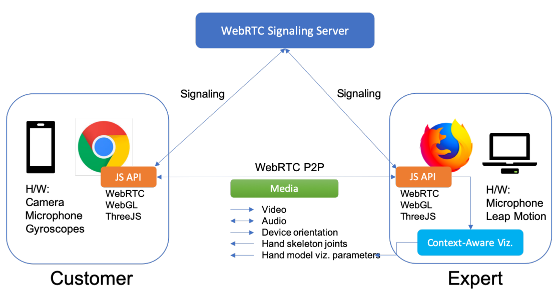


Figure 3: System overview.

data is embedded in the web page so that the system does not need to send the large 3D mesh data to the customer; instead only small amount of the hand skeleton joints and the camera parameters are transmitted to the customer.

The proposed system also utilizes the gravity direction of the customer's local work space to reorient the hand models, and this is managed by Context-Aware Visualization module in the expert side that processes the transmitted gyro sensor data from the customer, and produces the updated visualization parameters of the hand model that is transmitted and applied to the customer's hand model.

Gravity-aware hand visualization

In our system, a precise positioning of the expert's virtual hand with respect to the customer's real environment and objects of interest is important for the expert to convey accurate guidance and information on the required physical operations on the objects. To do this, a proper alignment between the expert's Leap motion device's interaction space (and thus the virtual hand therein) and the customer's camera view space is required (See Fig. 2).

In this work, we propose a novel gravity-aware alignment of the interaction space based on the gravity direction of the customer's environment to snap the interaction space onto the ground of the customer's real world, which helps the expert manipulate the virtual hand in a more intuitive and natural way as he/she would be present physically in the customer's environment.

More precisely, we estimate the direction of the gravity of the customer's work space using inertial sensors in the customer's smartphone, and use it to continuously change the orientation of the interaction space in the customer's camera view space, and thus the virtual 3D hand model therein, and its relative pose to the ground of the customer's work

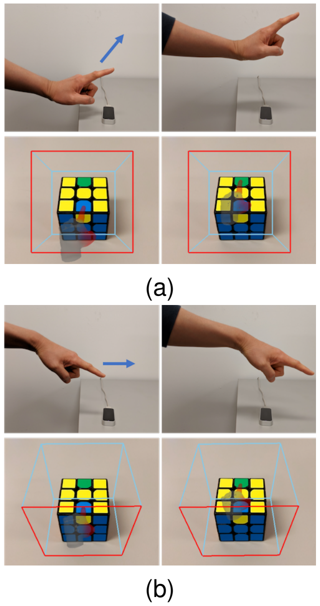


Figure 4: Visualization of the expert's pointing gesture and the interaction space without (a) and with (b) the gravity-aware alignment. In each sub-figure, Top: hand movement with pointing gesture over the Leap Motion device (from left to right column) to manipulate his virtual hand to point the blue and green top edges of the Rubik's cube in the customer's work space. Bottom: visualization of the 3D hand model and the interaction box representation superimposed on the view of the customer.

space is stabilized while the customer's hand-held camera freely moves. Figure 2 (a) shows an illustration of the expert's interaction space as a 3D box representation over the Leap Motion device, and (b) & (c) show the alignment of the interaction space to the customer's camera without/with the gravity-awareness. Figure 4 (a) and (b) shows the expert's hand movement over the Leap Motion device at the expert's site and the hand visualization in the customer's camera space without/with the gravity awareness. The arrow indicates the direction of the expert's hand movement. With the gravity-aware alignment, the movement of the expert's real hand in his physical world matches to the one of the virtual hand model in the customer's physical world.

Conclusion

We present a light-weight web-based remote assistance system, which does not require any prerequisite app to be installed, and can be accessed for use on any device connected to the internet such as smartphone, tablet, and smart glass. With our gravity-aware hand visualization, the expert can manipulate the virtual hand in a more intuitive and natural way. A demo video is in the supplemental material for review, and will be available on YouTube after publication.

REFERENCES

1. Leila Alem, Franco Tecchia, and Weidong Huang. 2011. Remote tele-assistance system for maintenance operators in mines. (2011).
2. Judith Amores, Xavier Benavides, and Pattie Maes. 2015. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1343–1348.
3. Weidong Huang and Leila Alem. 2013. HandsinAir: a wearable system for remote collaboration on physical tasks. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, 153–156.
4. Seungwon Kim, Gun Lee, Nobuchika Sakata, and Mark Billinghurst. 2014. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 83–92.
5. David S Kirk and Danaë Stanton Fraser. 2005. The effects of remote gesturing on distance instruction. In *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!* International Society of the Learning Sciences, 301–310.
6. Rajinder S Sodhi, Brett R Jones, David Forsyth, Brian P Bailey, and Giuliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 179–188.