3D Telepresence for Reducing Transportation Costs

Gregorij Kurillo
Allen Y. Yang
Ruzena Bajcsy

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2016-168
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-168.html

November 29, 2016

3D Telepresence for Reducing Transportation Costs[1]

Gregorij Kurillo, Allen Y. Yang, Ruzena Bajcsy
University of California, Berkeley
(gregorij@eecs.berkeley.edu, yang@eecs.berkeley.edu, bajcsy@eecs.berkeley.edu)

1. Introduction and Background

In this white paper we focus on the use of teleimmersion and augmented reality to achieve high level of presence while addressing the three key objectives of reducing energy in transportation, namely, *communication*, *labor*, and *experience*. We believe the recent confluence of teleimmersion and augmented reality technologies can not only reduce the cost related to transportation, but also has the potential to meaningfully increase productivity in remote collaboration and at the same time reduce performance errors related to these activities.

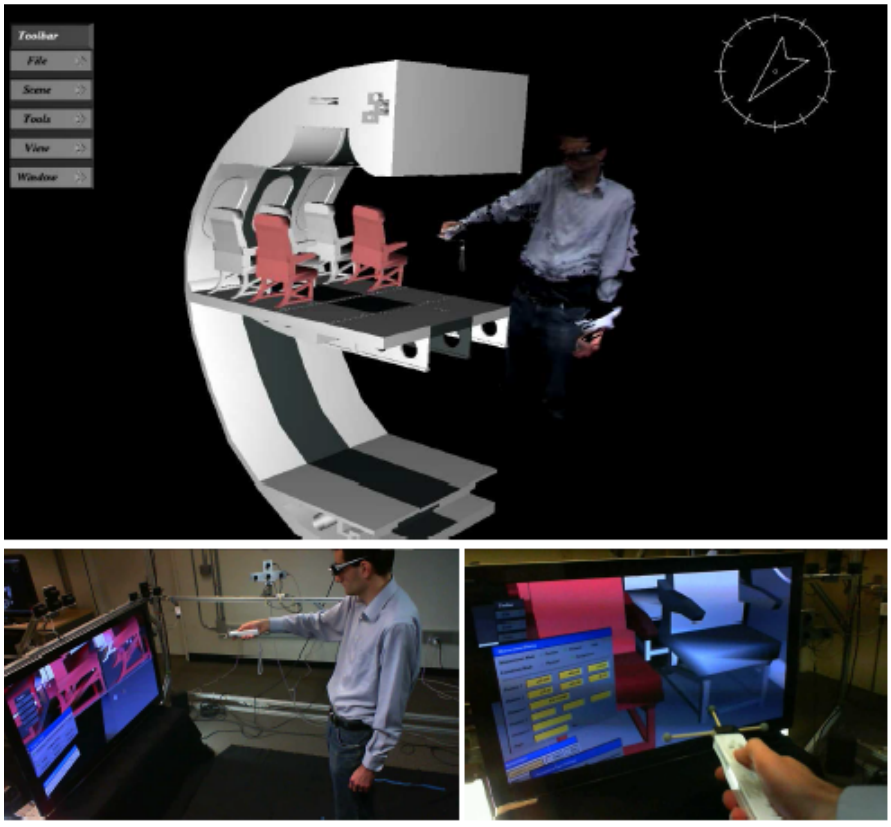

Figure 1: Teleimmersion experience designed at UC Berkeley: using 3D cameras a dynamic avatar is created in real time and projected at the remote location into a shared virtual environment to facilitate experience similar to face-to-face interaction. The users can take advantage of 3D interaction and display technologies to collaborate with their remote partners.

[1] This white paper was originally written for ARPA-E RFI DE-FOA-0001424

In recent years, the use of 2D telepresence via video conferencing has increased significantly with improved network connectivity, lower cost of camera sensors, and the ubiquity of connected smart devices, such as smartphones and tablets. This technology now facilitates tele-commuting for considerable number of people, remote meetings between geographically distributed teams, connectivity between family members, tele-medicine, and many other applications. Although the quality of experience (QoE) with regard to visual and audio transmission has significantly improved over the years, the 2D video technology has several inherent drawbacks that cannot be mitigated easily. These include partial loss of non-verbal cues such as gestures and eye contact, which have been shown to increase trust, collaboration and productivity (Fry & Smith, 1975; Doherty-Snnedon, 1997). There have been several attempts in the research and industry to improve the eye contact during video conferencing (Bohannon et al., 2012), however none of the approaches have yet been widely adopted. In addition to the loss of non-verbal cues when using traditional video conferencing technologies, there is disconnect between the users and the content (e.g. documents, models, diagrams). In 2D video conferencing sessions, remotely connected users are typically presented in separate windows on the screen or overlaid with the content. However a user cannot directly perceive or naturally initiate interactions between the users and the digital content. Although this visualization approach has been successfully used in many different applications (e.g. Skype, NetMeeting etc.), it provides very low level of presence to the users. Although some of these issues were mitigated for a remote whiteboard application by clever combination of RGBD camera and digital whiteboard (Higuchi et al, 2015), in general applications, in particular where users have to interact with 3D data or perform a task in real-world, the existing 2D teleconferencing paradigm remains highly limited.

Some of the aforementioned issues pertaining to 2D telepresence are addressed by 3D telepresence which provides streaming and rendering of 3D video data between two or more parties. The technology offered by several companies (e.g. CISCO) provides better eye contact and increased presence as the 3D telepresence communication rooms are designed to seamlessly combine the remote displays with the real environment. Traditional 3D telepresence however is primarily limited to conference rooms and still focused on accommodating business users. The technology has been showcased in several entertainment demonstrations (e.g. Musion 3D), however to date the practical applications have been limited by the high cost of such systems as compared to the 2D telepresence/video conferencing.

The interaction and experience of remote presence can be further enhanced by the 3D teleimmersion (TI) technology (DeFanti et al. 1999) which merges 3D tele-presence with virtual reality. This technology captures geographically distributed users with 3D cameras and renders their digital avatars inside a shared virtual reality (VR) environment. By capturing the digital likeness of the users and providing them with a shared (virtual) interaction space, it is possible to preserve eye contact, gestures, body language etc., while taking advantage of the important aspects of presence in communication and collaborative interaction. The participants can thus explore digital models that are completely virtual while being able to interact with those models remotely. By taking advantage of recent advances in augmented reality (e.g., Google Glass, Atheer AiR Glasses, and Microsoft HoloLens), the remote user's digital representations could be rendered over the real world, creating a mixed reality experience.

Tele-immersion technology for remote interaction has been explored by several researchers in the past decade (see the review in Kurillo & Bajcsy, 2013), including our group at University of California, Berkeley. In our group, we have demonstrated the use of tele-immersive technology in several application areas, including remote dance choreography (with University of Illinois), remote interaction with geoscientific data (with University of California, Davis), collaborative archaeology (with University of California, Merced and with University of Tokyo, Japan), and tele-medicine applications (with University of California Davis Medical Center and with University of Basque Country, Spain). In our demonstration systems that we have built over the years, we have investigated various aspects of

interaction, networking, data compression, rendering, and others, while working with multi-disciplinary groups of users. Some of the findings were reported in Kurillo & Bajcsy (2013) and Arefin et al. (2014). In this document we outline the issues that are relevant to the adoption of teleimmersion technology for telepresence and telelabor.

2. Technological Barriers

In the literature, there have been several different teleimmersion systems presented and evaluated from performance and usability perspective (see Kurillo & Bajcsy for the review). The majority of the systems were evaluated only in a laboratory setting with limited number of users and applications. There are several new challenges when transitioning to more arbitrary environments, such as offices, hospitals, or homes. The challenges include: synchronization of multiple devices, compatibility of devices from different vendors, calibration and registration between the devices/cameras, low-fidelity and low frame-rate of 3D reconstruction, network delays due to large data packages, sensitivity to the changes in environmental conditions (e.g. illumination), complexity of running the system, high upfront cost of setting up the system and high cost of maintenance, lack of software development tools for building applications, etc. In addition, having a large system that requires a dedicated room is not always practical. *Therefore, there is a need to develop smaller scale portable or even wearable systems using off-the-shelf technology that can be easily deployed and automatically calibrated.*

The ideal system for teleimmersion would include the following components:

- Active camera(s) for accurate (on the order of mm) real-time 3D acquisition with high resolution (HD+) texturing capability and synchronization between multiple cameras;
- For VR-based applications: autostereoscopic display with head tracking or a lightweight head mounted display (HMD) with large field of view and minimum latency;
- For AR or mixed reality based applications: head mounted AR display with large field of view and robust real-world tracking capabilities;
- Gesture and hand tracking system that include robust real-time gesture recognition and accurate tracking of fingers and hands;
- Various interaction devices that can be combined with the gesture-based interaction;
- Microphone array that can capture directional sound;
- Speakers for spatial audio capabilities;
- High-bandwidth network connectivity.

Recent technical advances in display, sensor, and tracking technologies are already meeting some of these requirements and thus provide a new opportunity to rekindle teleimmersion. On the data acquisition side, there is Microsoft Kinect and several similar depth-sensing cameras that provide real-time accurate 3D reconstruction and body pose tracking. There are also several emerging display technologies that offer relatively high quality VR/AR experience within reasonable price range, such various head mounted VR displays (Oculus Rift and HTC Vive) and AR displays (Epson BT-200, Microsoft HoloLens, and Magic Leap).

In addition to the hardware requirements, there are several other research and technical challenges that need to be addressed in order to improve the quality of experience with the teleimmersion:

- Automatic, user-friendly calibration of multiple cameras;
- Synchronization of multiple active cameras to reduce interference;
- Automatic registration of 3D cameras with display and tracking systems;
- Development of lossless or near-lossless compression methods to efficiently transmit RGB+D data for full 3D reconstruction (e.g. textured mesh);

- Generalized framework that integrates devices from various vendors;
- Easy to use API for building teleimmersion applications.

3. User Adoption

The adoption and use of 2D tele-presence via video-conferencing has been relatively high due to its simplicity and high quality of the video streaming in the recent years. Similar advances need to happen in the 3D tele-presence and teleimmersion.

The VR technology is becoming increasingly popular in entertainment with several display devices on the market. There is currently emerging interest in AR technology, however the development of new hardware devices has been limited in this area. One of the upcoming technologies for AR is the Microsoft HoloLens and less advertised AR technology in development by Magic Leap.

The early adoption in entertainment will facilitate better understanding and availability of these technologies to average users, providing an opportunity to use these systems also in professional activities. Primary focus of the industry has been on providing the experience of real-time presence in virtual worlds in a local setting. Transitioning to the remote experience is currently limited due to the low quality of captured 3D data versus computer generated 3D data (e.g. graphics models), inherent lag of the networking, and lack of applications that would demonstrate the multi-user collaborative aspect of VR. People have been reluctant to use avatars as a form of remote communication due to their limitations in replicating the body language and establishing trust during communication. Teleimmersion technology on the other hand features realistic digital representations of users (i.e., real-time avatars) that are generated through computer vision techniques. These can be further combined with various modeling approaches. The teleimmersion technology can thus through its 3D telepresence capabilities improve the experience of presence by projecting digitized versions of remote users into the virtual or mixed reality.

Based on our experience within various applications of teleimmersion, the quality of 3D visual models of users still need to improve considerably to achieve the levels similar to the fidelity of 2D video conferencing. With high quality 3D visuals, the micro-expressions, eye gaze and various aspects of body language will be possible to convey. However, the visual fidelity needed depends on the application. In applications that focus on recognition of body gestures (e.g. training in assembly operations), the facial fidelity may not be as important. On the other hand, the applications that focus on communication and personal interaction should convey the digital representation of remote users as accurately as possible. Current technology advances in 3D data acquisition are finally closing in on this goal. Microsoft Kinect v2 sensor for XBOX One for example achieves accuracy of real-time 3D data acquisition in the order of 2mm with high-definition texture acquisition (Yang et al., 2015). Such high resolution can provide sufficient quality for detection of facial expression for depth data. High resolution of digital models however poses a challenge to efficient data transmission over the current networks. Hence, there is a need for better compression techniques that take into consideration various levels of detail needed in different applications.

Another important aspect for user adoption is the display quality. Although VR head mounted devices provide a convenient alternative to the existing expensive CAVE systems, such devices also create a disconnect from the real world and often induce motion sickness. An alternative are high quality wearable head mounted AR displays, which overlay the digital rendering over the real world. This type of displays offer new opportunities for tele-presence and teleimmersion as the remote users can be rendered as if they are located in the same physical (real) space.

4. Utility

The teleimmersion technology provides increased level of presence of the remote users, whether it is bundled with virtual or augmented reality. Aside from enhancing the communication that is currently performed with 2D video conferencing, there is a potential to use this technology in several other areas that would reduce the need for travel. The technology could be applied in remote collaboration where multidisciplinary teams have to work closely with 3D (and 2D) data or models, such as in design, architecture, manufacturing, medicine, physical sciences, astronomy, education, digital humanities etc. The technology could also be applied for tele-medicine. Although video-based tele-medicine is starting to take off, the value of having three-dimensional visualization of remotely located patients can provide additional cues that may be missed from 2D views. This is especially important in the future applications of remote physical rehabilitation. Currently, many of the patients from rural areas have to be transported for their regular checkups or physical therapy. Some of the visits could be performed via tele-medicine while taking advantage of the 3D information provided by this technology.

Furthermore, the teleimmersion technology could be used to create remote presence in real environments via augmented reality displays. This would be especially applicable for supervisory jobs where experts would not need to travel to multiple locations for the purpose of training, supervising or assisting labor force out in the field who may not have complete expertise to solve the problem at hand. In combination with virtual reality and measurements/data obtained at the real world location, the experts could make more informed decisions and thus reduce the errors due to miscommunications between the remote teams.

5. Technology-Specific Issues and Future

On the acquisition side, there is potential to improve the accuracy of the 3D reconstruction and improve the body tracking even further. As we have seen in the recent improvements of the Kinect v2 versus Kinect v1, the quality and accuracy of the 3D data has significantly improved. It is expected that even higher resolution 3D cameras will be available within the time period of 1-3 years. This would provide more accurate depth maps from which better tracking of finger movement and facial expressions could be achieved.

Accurate 3D acquisition is also needed in order to create realistic and high fidelity three-dimensional rendering of human users, either for VR or AR tele-presence. There are several challenges with using multiple cameras as mentioned before. Alternative approaches that could be developed include a combination of model based real-time animation. Realistic models could be created a priori and then manipulated based on RGB+D data and skeletal tracking. Similar technology was recently demonstrated for facial manipulation by researchers from Stanford University (New York Times, 2015). Parametric models are also more efficient to transmit over the network as opposed to streaming the entire 3D mesh. This approach would however only work for limited scenarios of one-to-one interaction.

On the display side, it is expected that the head mounted VR HMDs will somewhat improve with higher resolution of display, larger field of view and possible eye tracking capability which could further enhance the experience of presence in the virtual environment. For the VR displays, it is important to further reduce the latency and improve the accuracy of orientation and position tracking in order to bridge the disconnect between the human sensory senses and the virtual reality experience. This is the most important aspects of VR development at this time. In other words, if the motion sickness cannot be eliminated, the adoption of wearable VR displays may be limited by general population.

Significant advances can be expected in the AR wearable displays. Current AR glasses have smaller field of view as compared to the available VR head mounted displays. Some of the issues with the see-

through displays could be mitigated by the use of retinal projection. However there are currently no working systems with retinal projection yet available in the market. In addition to the display quality, more accurate and more robust tracking of environmental features is needed in order to render 3D models without jitter and robustly to occlusions.

In addition to aforementioned hardware and algorithmic limitations, there is a need to perform user studies that would evaluate the most appropriate display, interface device, and interaction mode for particular areas of remote collaboration and tele-labor applications. These would require development of standardized quality of experience measures that could subjectively and objectively evaluate how different combination of technologies improves efficiency and performance in specific type of tasks. The subjective measures could include development of various questioners while the objective measures could specify a standardized set of tests that would observe the performance speed, accuracy, movement efficiency, fatigue etc.

6. Conclusion

Recent advances in sensor technologies and proliferation of virtual reality market have created new opportunities for developing low-cost teleimmersion systems that would reduce the transportation costs and increase productivity through virtual meetings, remote collaboration, and mixed reality applications. In this RFI document we have summarized the technical and user adoption challenges with regard to the teleimmersion technology. As outlined, the primary areas that would benefit from this technology in terms of transportation cost include tele-medicine, military applications, training, and design in manufacturing.

References

- Arefin A, Huang Z, Rivas R, Shi S, Xia P, Nahrsted K, Wu W, Kurillo G, Bajcsy R (2013), Classification and analysis of 3D teleimmersive activities, *IEEE MultiMedia* 20(1):38-48.
- Bohannon LS, Herbert AM, Pelz JB, Rantanen EM (2012). Eye contact and video-mediated communication: A review, *Displays* 34(2):177-185.
- Cisco, URL: http://www.cisco.com/c/en/us/solutions/telepresence/public_telepresence.html
- DeFanti T, Sandin D, Brown M, Pape D, Anstey J, Bogucki M, Dawe G, Johnson A, Huang TS (1999). Technologies for virtual reality/tele-immersion applications: issues of research in image display and global networking. In: EC/NSF workshop on research frontiers in virtual environments and human-centered computing.
- Doherty-Sneddon G, Anderson A, O'Malley C, Langton S, Garrod S, Bruce V (1997) Face-to-face and video-mediated communication: a comparison of dialogue structure and task performance. J Exp Psychol Appl 3(2):105–125.
- Fry R, Smith G (1975). The effects of feedback and eye contact on performance of a digit-coding task. J Soc Psychol 96:145–146.
- Higuchi K, Chen Y, Chou PA, Zhang Z, Liu Z (2015). ImmerseBoard: Immersive Telepresence Experience using a Digital Whiteboard, Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), 2015.
- Kurillo G, Bajcsy R (2013), 3D teleimmersion for collaboration and interaction of geographically distributed users, *Virtual Reality* 17:29-43.
- Microsoft HoloLens (2015). URL: http://www.microsoft.com/microsoft-hololens/en-us
- Musion 3D, URL: http://www.musion3d.co.uk
- New York Times (2015), Manipulating Faces From Afar in Realtime, URL: http://www.nytimes.com/2015/10/26/science/manipulating-facial-expressions-in-live-video.html
- Yang L, Zhang L, Dong H, Alelaiwi A, El Saddik A (2015). Evaluating and improving the depth accuracy of Kinect for Windows v2, *IEEE Sensors Journal* 15(8).