Advances in Immersive Communication: (1) Telephone, (2) Television, (3) Teleportation

PHILIP A. CHOU, Microsoft Research

The last great advances in immersive communication were the invention of the telephone over 137 years ago and the invention of the video telephone (né television) over 86 years ago. However, a perfect storm is brewing for the next advance in immersive communication, thanks to the convergence of massive amounts of computation, bandwidth, resolution, new sensors, and new displays. It could well be the Multimedia community that turns this brew into the next great advance in immersive communication, something akin to teleportation.

Categories and Subject Descriptors: H.4.3 [**Information Systems Applications**]: Communications Applications—*Computer conferencing, teleconferencing, and videoconferencing*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; I.4.0 [**Image Processing and Computer Vision**]: General

General Terms: Human Factors

Additional Key Words and Phrases: Immersion, telepresence, virtual reality, depth cameras

ACM Reference Format:

Chou P. A. 2013. Advances in immersive communication: (1) Telephone, (2) television, (3) transportation. ACM Trans. Multimedia Comput. Commun. Appl. 9, 1s, Article 41 (October 2013), 4 pages.
DOI: http://dx.doi.org/10.1145/2492704

Immersive communication is the real-time exchange of the natural social signals between people who are geographically separated in ways that suspend their disbelief in being together [Apostolopoulos et al. 2012]. The first major advance in immersive communication was the invention of the telephone in 1876. Within a few years, the popular media anticipated that the next advance would be visual [du Maurier 1878]. However, it took 50 more years and decades of work before the necessary technologies were in place to offer the perception of moving imagery. In fact, in 1900, the term *television* was coined in anticipation of the invention of the visual counterpart to the telephone, though it was not until 1926 that real-time grayscale image capture and rendering above a few frames per second was demonstrated.

Today, 86 years after the invention of the television, use of real-time video communication is rapidly rising, with services such as Skype carrying a billion minutes of video calling per day and growing 40% annually [Steele 2013]. Nevertheless, despite the existence of simple and inexpensive access to video communication, people still travel to meet each other and meet face to face whenever possible. This is because, as social animals, we humans convey and derive meaning from physical proximity, posture,

Author's email: pachou@miscrosoft.com.

DOI: http://dx.doi.org/10.1145/2492704

gaze, and gesture, as well as numerous environmental aspects. Sharing a meal or attending a party by video remains unfulfilling. We continue to commute to work, travel to conferences, and go on family vacations in person. This is proof that there is an unfulfilled need for further advances in immersive communication.

Highly immersive communication has long been anticipated in the popular culture through works of science fiction, such as *Star Trek*, *Star Wars*, *The Matrix*, and *Avatar*. These works envision that peoples' bodies could be transported to a common physical or virtual location, where interaction could take place—something akin to teleportation. Of course, even if teleportation were to exist, not every future communication would take place through such a highly immersive medium. Highly immersive communication can reveal more information than we wish to share in many cases. For this reason, even text messaging could remain a preferred means of communication sometimes. Nevertheless, the introduction of a practical means of highly immersive teleportation would likely be a momentous advance having impact as dramatic as the introduction of the telephone and the television in their day.

There are several possible approaches to teleportation. Physicists tell us that *classical teleportation* could be achieved by cooling a person to absolute zero (thereby shrinking him into a block of Bose-Einstein condensate); transmitting the emitted photons; shining the received photons onto another block of Bose-Einstein condensate; and *voila*, reconstituting the person in another location [Kaku 2008]. Of course there may be unforeseen consequences if the person is not cooled to absolute zero, leaving some of his state behind.

A more practical method for the foreseeable future is *physical* (or *mechanical*) *teleportation*, in which the teleported person controls a physical (or mechanical) avatar that represents him in a remote location. This is the approach originally envisioned by Minksy [1980] and popularized in the movies *Avatar* and *Surrogate*. Simple versions are now being commercialized in the form of Telepresence robots [Guizzo 2010a], while more humanoid versions are being investigated in the lab [Guizzo 2010b]. Though there are many virtues of this approach, in this essay, we focus on yet another approach—virtual (or electronic) teleportation—favored for its flexibility and potential fidelity.

In *virtual* (or *electronic*) *teleportation*, the teleported person and possibly other physical objects are represented by a virtual (or electronic) avatar and possibly other virtual objects in a virtual environment or mixed reality location. This is the approach popularized by fiction, such *Star Trek*, *Star Wars*, and *The Matrix*, and explored by the National Tele-Immersion Initiative [Lanier et al. 1997–2000], UNC Office of the Future [Fuchs et al. 1998–2009], UIUC/Berkeley TEEVE [Nahrstedt et al. 2005–2012], Microsoft Viewport [Zhang et al. 2013], and others.

One vision of virtual teleportation is being able to drop into any location on the planet (in this world or any other world, real or virtual) as easily as we may drill down to street level today on the Internet in Google or Bing maps, except that instead of simply arriving at street level to look at the entrance to a building, we may enter it, see the inside teaming with people both real and virtual, and interact with them, as if we were there.

It is the thesis of this editorial that the right technologies are rapidly coming together to make this and similar visions of virtual teleportation practical. Against the backdrop of astounding increases in computation, bandwidth, and resolution over the last few decades and the pervasiveness of the Internet, recent advances in sensing and display are completing the puzzle, complementing key existing pieces in graphics, vision, networking, and audio processing. Though there are numerous challenges to fully realizing the vision of virtual teleportation in a practical way, important components of it are rapidly developing. The two most important components, as in classical teleportation (or any communication scenario, for that matter), are *sensing/analysis* (input) and *synthesis/rendering* (output).

However, a third component unique to virtual teleportation—*low delay tracking*—also plays a key role, spanning sensing and rendering. Key elements of these are all recently coming together.

Sensing/Analysis (Input)

On the input side of virtual teleportation, it is necessary for the person who is being teleported to be sensed (or captured) and then for the captured signals to be processed or analyzed into a representation suitable for subsequent compression, transmission, decompression, synthesis, and rendering in 3D. The ability to reproduce the person in 3D at his or her destination is perhaps the primary feature distinguishing virtual teleportation from simple video communication. The recent emergence of inexpensive depth cameras, such as the Microsoft Kinect, has made 3D capture suddenly so much easier. Depth cameras of all varieties (structured light, heterodyne time-of-flight, pulsed/gated time-of-flight, and traditional stereoscopic) are now riding the same wave of Moore's law enjoyed by traditional color cameras and are making rapid advances in resolution, accuracy, power, size, and cost. With proper synchronization and multiplexing to avoid interference issues, multiple depth and color cameras will soon be able to capture high quality 3D representations of people for virtual teleportation.

Synthesis/Rendering (Output)

On the output side is synthesis and rendering, or display. The effect popularized in *Star Wars,* of Princess Leia appearing in mid-air, is still difficult to achieve, as popularly conceptualized. Though there have been proposals for ways to build such a display [Perlin 2001], for the foreseeable future, it will be more practical to approximate it. Fortunately there are several ways, all becoming quite viable. One way is with near eye displays. To see the virtual avatars around us, we can don optical see-through glasses. The avatars then become augmentations of our reality. Stereoscopic optical see-through glasses are making rapid advances in weight, resolution, price, style, field of view, and power. Another solution is video see-through, a technology perhaps best matched with phones and tablets, which could be used as screens through which the virtual avatars could be seen. Large fixed displays are yet another possible solution where rapid progress is being made, thanks largely to the consumer electronics industry. The main difficulty with large displays is that they must be shared by multiple people, and currently it is hard for each person to have his or her own unique view. However, different views can be multiplexed, and for small numbers of viewers, there are currently several possible multiplexing solutions ranging from temporal multiplexing with active shutter glasses to spatial multiplexing with multiple rear projectors. For a larger number of viewers, though not yet commercially feasible, light field or holoscopic displays [Aggoun et al. 2013] are good bets given the astounding rate at which display resolution is increasing. Light field displays comprise a 2D array of micro displays, each behind a micro lens, configured to show a different set of pixels to viewers in different locations.

Low-Delay Location and Pose Tracking

A third technology, low-delay location and pose tracking, has surprisingly turned out to be critical for virtual teleportation. Precise, low-delay tracking, either of a moving display in relation to its environment or of a moving observer in relation to the display, or both, is critical for any of the previously mentioned displays (with the possible exception of the light field display) so that the avatars can be displayed from the proper point of view for each viewer. Fortunately, a combination of GPS, inertial, magnetic, and visual sensors (including depth cameras) are making it increasingly feasible to precisely determine the instantaneous location and pose of a person or device, whether indoor or outdoor.

Conclusion

Of course there are many other technical issues to be solved as well, including spatial audio capture and rendering, hole filling by modeling, denoising, compression, relighting, re-reverberation, correct occlusion, and so forth. Passable solutions to these problems already exist and are being improved by the Multimedia, Signal Processing, Computer Graphics, and Computer Vision research communities.

As with the development of the telephone and television, the development of tele-immersion has had its visionaries and pioneers. However, it could take decades of hard work before conditions are ripe for breakthroughs in practical systems. But after conditions are ripe, it may be only a few years, marked by rapid progress, before a worldwide rush to the patent office. It is evident that for virtual teleportation and immersive communication in general, conditions are ripening quickly. Given the urgent need to reduce carbon emissions, bring jobs to workers and workers to jobs, increase productivity and security, bring education and medical care to people in all corners of the world, and generally to bring families, friends, and communities together, there is also a moral imperative to make these breakthroughs. Let it the Multimedia community who will put the pieces together to make them happen.

REFERENCES

AGGOUN, A., TSEKLEVES, E., SWASH, M. R., ZARPALAS, D., DIMOU, A., NUNES, P., AND SOARES, L. D. 2013. Immersive 3D holoscopic video system. *IEEE MultiMedia 20*, 1, 28–37.

APOSTOLOPOULOS, J. G., CHOU, P. A., CULBERTSON, B., KALKER, T., TROTT, M. D., AND WEE, S. 2012. The road to immersive communication. *Proc. IEEE 100*, 4. *(Apr. 2012).*

DU MAURIER, G. 1978. Edison's telephonoscope (transmits light as well as sound). In *Punch's Almanack.*

FUCHS, H., ET AL. 1998–2009. Office of the future. http://www.cs.unc.edu/Research/stc/index.html. (Last accessed 4/13).

GUIZZO, E. 2010a. When my avatar went to work. *IEEE Spectrum* (Sept.).

GUIZZO, E. 2010b. Hiroshi Ishiguro: The man who made a copy of himself. *IEEE Spectrum* (Apr.).

KAKU, M. 2008. *Physics of the Impossible*. Doubleday, New York, NY.

LANIER, J., SADAGIC, A., ONDAYKO, D., AND BROWN, M. 1997–2000. About NITT. http://www.advanced.org/tele-immersion/introduction.html. (Last accessed 4/13).

MINSKY, M. 1980. Telepresence: A manifesto. *Omni Mag*. June, 44–52.

NAHRSTEDT, K., ET AL. 2005–2012. TEEVE: Tele-immersive Environment for EVErybody. http://cairo.cs.uiuc.edu/projects/teleimmersion/. (Last accessed 4/13).

PERLIN, K. 2001. Star Wars chess on the holodeck. In *Proceedings of the 1st International Conference on Semiotics for Games and New Media (COSIGN'01)*. http://www.cosignconference.org/conference/2001/. (Last accessed 4/13).

STEELE, E. 2013. Thanks for making Skype a part of your daily lives-2 billion minutes a day! Skype Blog. http://blogs.skype.com/2013/04/03/thanks-for-making-skype-a-part-of-your-daily-lives-2-billion-minutes-a-day/. (Last accessed 4/13).

ZHANG, C., CAI, Q., CHOU, P. A., ZHANG, Z., AND MARTIN-BRUALLA R. 2013. Viewport: A fully distributed immersive teleconferencing system with infrared dot pattern. *IEEE MultiMedia 20*, 1, 17–27.

Received May 2013; accepted May 2013