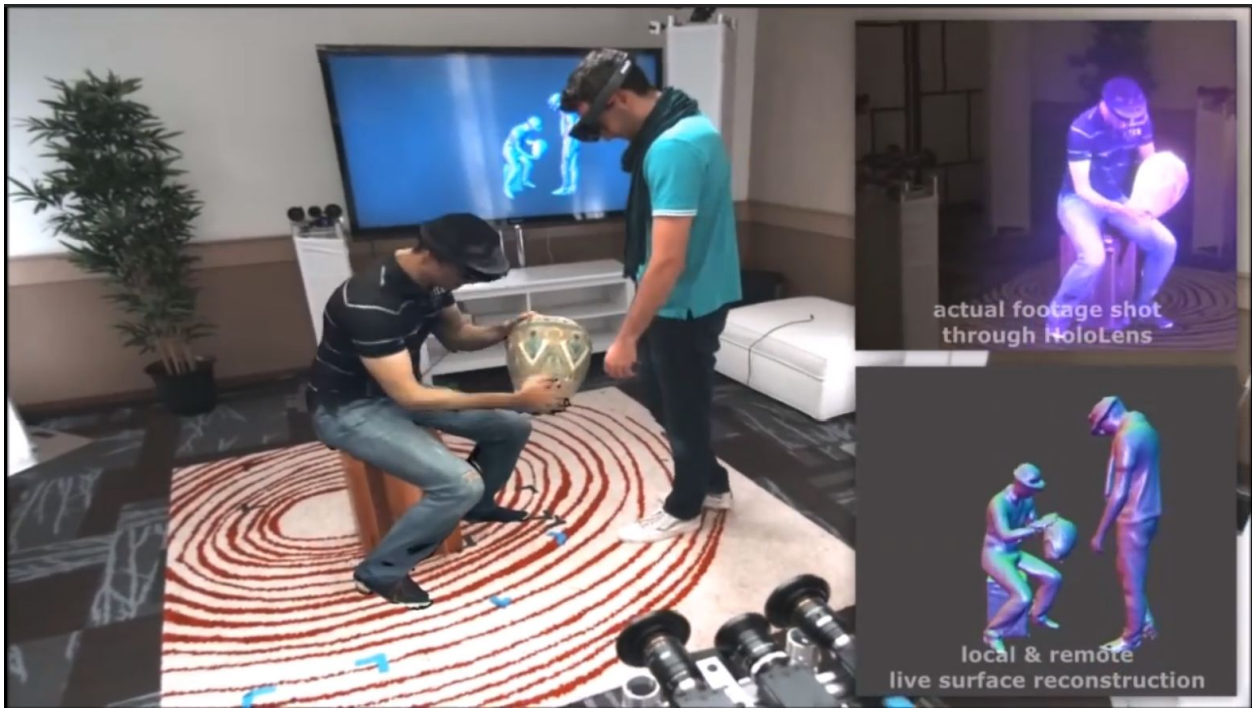
Peter Hanula

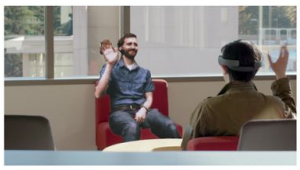
Holoportation is the future! ...or is it...? Although Holoportation seems to address current limitations with 3D capture technology and mixed reality while offering solutions, a lot of work is still needed to make this system an everyday reality for our societies. Furthermore, will communicating and interacting with remote users really become as natural as face-to-face communication? More importantly, is that what we really want or need to achieve? What percentage of people actually want to communicate this way? These are some questions to consider as we try to provide a critical analysis of Holoportation.



Holoportation tries to address technical and social limitations, where both seem dependent on each other, and I find this quite interesting. There are two main components in making this system work. The first is head-worn displays and head tracking, and this component is pretty much solved with today's plethora of different HMDs. The second component is real-time 3D scanning of humans, and this is the main focus of their research. There are several reasons why 3D reconstruction is hard. Off the shelf sensors (like the Xbox Kinect) are low quality. Parametric models, where the person is pre-scanned and movement is taken from head and controller poses, are never quite rich enough of an experience, often leading to the uncanny valley. Humans are complex to model, and the use of avatars lack human connection. You can

use studio capture, but then you will need a large space, controlled lighting, and a huge amount of cameras and processing. For reference, a typical full studio capture contains around 150 cameras and it takes 30 minutes to process one frame.






Here is what Microsoft wants to do in regards to real-time 3D scanning of humans. They want the system to be supported in normal environments (with less infrastructure) with 24 cameras and be real-time (30 FPS). To achieve this, two major problems needed to be solved. First, a new custom depth camera needed to be created, capturing IR images. Second, a temporal reconstruction algorithm was created that allows the model to be fused temporally over time. A third, but not talked about problem, is the rendering that happens on the remote side. These problems were addressed and the demo shows it. In their demo, the system doesn't discriminate between what is captured in the scene, so you can actually see the user pick up an object on the remote side and have it scan instantaneously, and then show it remotely to his colleague. Furthermore, all the reconstructions are of high quality and there is minimal temporal and spatial artifacts. The research team stresses that temporal consistency is really important for the experience to feel human and natural. I also like the hologram effect added to the model, creating consistency over a range of diverse environments, especially with lighting.

Up to this point, my analysis was mostly based off research oriented videos or text presented by the Microsoft research team (example paper), but the consumer demo video paints a slightly different picture. While the research is very promising and proof that we are closer to achieving Holoportation, the demo video (linked above) kind of seems gimmicky and like a white lie. Other words to describe this would be "marketed" and "scripted". For example, to portray a fully immersive environment and heighten the effect and sense of interaction, both the local and remote scenes had to be preconfigured to get a nice mapping between the two. Specifically, the chair and coffee table are in both scenes, perfectly mapped and in the

exact relative positions. This is something that is unlikely in a real user scenario. Furthermore, they typically only show one person with a headset (typically the person in the same room as the camera that shot the video), and the other person (the hologram) without a headset. This is a very clever trick to make the hologram appear as human as possible, since our faces are key in making us human-like. In reality, all parties utilizing the system would need headsets to share the same experience. Elaborating on this, the little girl can hear and speak to her father, but can't see him without a headset. She even almost runs through him at one point. To the audience watching the video, what is more human and has a greater effect, a hologram girl with a giant headset covering her entire face, or a hologram girl without a headset telling her father she loves him? I will talk about this further down below. On a general note, it's nice that you can interact with the system using multiple devices (video see-through AR, Hololens AR, Oculus Rift VR). They also mention you only need two cameras, but there are no demos with just two cameras. With two cameras, it is not certain you will get a nice 3D full 360 hologram. One camera would have to be directly in front of you, and another directly behind you. Any movement away from this line and part of the hologram can't be computed accurately. Furthermore, they cut bandwidth by as much as 95%, but how much processing power is needed to do that in real time? That was a huge computer in the back of the car in the mobile demo video. Holoportation, while having lots of potential, still needs work to achieve what they are describing in the videos.

There are several pivotal challenges in terms of making Holoportation a reality. First is field of view, which is pretty much solved through VR headsets, but not AR. Field of view is needed to see the entire person you are communicating with, and not just their top half. Second, this system still requires heavy use of infrastructure. Third, as briefly mentioned above, is “headset removal” in the virtual representations. If everyone needs headsets and they are trying to collaborate especially if some people are co-located, this isn’t a very appealing experience. People want to have a more human-like experience, but everyone sees computers on each others faces; Sounds counter intuitive. There is ongoing research in this area, for example eye tracking and in-painting replacing headsets virtually. In the case above, the little girl can’t see her father, but he can see her because he is wearing a Hololens. There is a disconnection between the two, and not everyone will have the same experience. However, it was really interesting to see the father being cautious when his daughter climbed on the furniture. He put out his hand as to safeguard her from falling over, but in reality that wouldn’t help because she is virtual and physically located in a different room. It is still intriguing to see how the human mind responds to such scenarios. The fourth challenge is compression. Raw data from the cameras equal about 23 Gb/s. Current Holoportation compression comes down to about 1 Gb/s. For reference, HD video takes about 10 Mb/s. Future work also involves exploiting temporally consistency, a texture atlas, and compressing meshes over time.

Resources

1) https://www.strangerdimensions.com/2013/11/25/10-creepy-examples-uncanny-valley/
2) https://www.microsoft.com/en-us/research/project/holoportation-3/
3) https://youtu.be/7d59O6cfaM0
4) https://youtu.be/nTkFO2xNkIk
5) https://youtu.be/KeuGwhv6yN4
6) http://delivery.acm.org/10.1145/2990000/2984517/p741-orts-escolano.pdf?ip=131.193.76.54&id=2984517&acc=ACTIVE%20SERVICE&key=B63ACEF81C6334F5%2EAACB7351D18CAF98%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=828039385&CFTOKEN=91671568&__acm__=1510360876_c8da085c6e20d089f8a4a7f0bf3fd18a
7) https://youtu.be/a3BEbki6x9c
8) https://youtu.be/o00mn1XbCIg