https://doi.org/10.2352/ISSN.2470-1173.2019.16.3DMP-007

Holo Reality: Real-time, Low-bandwidth 3D Range Video Communications on Consumer Mobile Devices with Application to Augmented Reality

Tyler Bell[1,*] ***and Song Zhang***[2]

[1] ***Department of Electrical and Computer Engineering, University of Iowa; Iowa City, Iowa 52242, USA***
[2] ***School of Mechanical Engineering, Purdue University; West Lafayette, Indiana 47907, USA***
[*] ***tyler-bell@uiowa.edu***

Abstract

An increasing number of mobile devices are equipped to acquire 3D range data along with color texture (e.g., iPhone X). As these devices are adopted, more people will have direct access to 3D imaging devices, bringing advanced applications, such as mobile 3D video calls and remote 3D telemedicine, within reach. This paper introduces Holo Reality*, a novel platform that enables real-time, wireless 3D video communications to and from today's mobile (e.g., iPhone, iPad) devices. The major contributions are (1) a modular platform for performing 3D video acquisition, encoding, compression, transmission, decompression, and visualization entirely on consumer mobile devices and (2) a demonstration system that successfully delivered 3D video content from one mobile device to another, in real-time, over standard wireless networks. Our demonstration system uses augmented reality to visualize received 3D video content within the user's natural environment, highlighting the platform's potential to enable advanced applications for telepresence and telecollaboration. This technology also has the potential to realize new applications within areas such as mechatronics and telemedicine.*

Introduction

Over the past several decades, much progress has been made in the area of 3D range imaging. Modern systems have the capability to capture high-resolution, high-accuracy 3D geometry data at real-time, or faster, speeds [1]. Due to these capabilities, 3D imaging systems have been employed in numerous applications across a variety of disciplines, including manufacturing, entertainment, robotics and homeland security. Somewhat overlapping with the advancements made in 3D imaging, the areas of 3D telecommunications and 3D telepresence have also been active areas of research over the past several decades [2, 3, 4, 5, 6, 7].

One current state-of-the-art works in this area is the *Holoportation* system [8]. The Holoportation system used eight camera *pods*, comprised of three cameras each, to capture the data needed to reconstruct a color-textured 3D mesh of a user in real-time. The 3D mesh was transmitted at 30 Hz to a remote user for simultaneous visualization within augmented reality headsets. Although this system was able to impressively reconstruct 3D geometry and color texture of entire bodies and objects, it required large amounts of high-end hardware: 24 cameras, 5 high-powered PCs, 10 high-powered graphics processing units (GPUs), and a connection speed of 1-2 gigabits per second (Gbps).

Recently, Bell et al. proposed the *Holostream* platform for high-quality 3D video encoding and streaming [9]. This platform used principles of triangulation to reconstruct 3D geometry [10]. Specifically, a custom-built structured light scanner was used to reconstruct accurate 3D video data and color texture information at 30 Hz. Captured 3D range geometry was encoded by [11] into regular 24-bit RGB images before being transmitted within an H.264 video stream. Depending on the desired quality, bitrates of 4.8-14 megabits per second (Mbps) were needed to transmit the compressed 3D video stream, which could be performed wirelessly. Remote users then received this stream, decoded its individual frames back into 3D geometry [11, 12], and visualized the reconstructed 3D range video.

Overall, in addition to efficiently transmitting 3D video data in real-time, 3D video communication systems must address the challenge of making such technology accessible to more fields of research and application areas. The Holoportation system, for instance, required many components of high-end hardware to capture, reconstruct, transmit, and visualize 3D data. Similarly, Holostream required a specialized structured light scanner, a high-powered PC, and a high-powered GPU to reconstruct high-quality 3D data. The requirements for each of these systems may be too demanding (both technically and financially) for others to adopt. A growing number of mobile devices, however, have the ability to acquire 3D range data along with color texture data (e.g., iPhone X). If such devices could be used directly for 3D video communications, the challenge of obtaining and integrating specialized, often expensive, hardware components can be overcome. Further, as every day users adopt 3D-enabled mobile devices, a viable and efficient 3D video communications platform could enable advanced applications, such as 3D video calls and 3D telemedicine.

This paper proposes *Holo Reality*, a novel and modular platform for performing 3D video acquisition, encoding, compression, decompression, and visualization entirely on consumer mobile devices. Details of each module in the platform will be introduced and described. Details of our demonstration system—that successfully delivered 3D video content with color texture data from one mobile device to another, in real-time, across a standard wireless network—will also be given. Finally, we will show our system's capability to use an augmented reality environment to visualize received 3D video content in real-time within a user's physical space, highlighting the platform's potential to enable advanced applications for telepresence and telecollaboration.

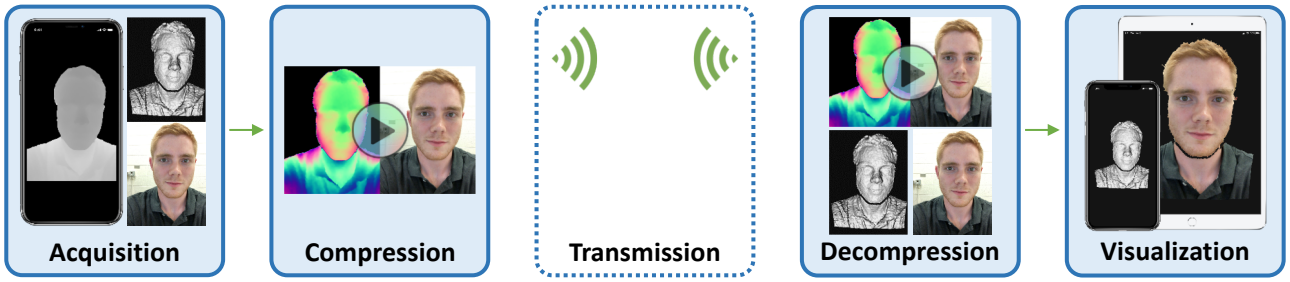


Figure 1. Holo Reality*: the proposed real-time, mobile 3D video communications platform. The* Acquisition Module *uses a mobile device's (e.g., iPhone X) on-board sensors to acquire 3D range geometry and color texture in real-time. The* Compression Module *then encodes this data into regular 24-bit RGB images that are then further compressed with H.264 video compression. The* Transmission Module *delivers the H.264 video stream wirelessly to a receiving device via WebRTC. The receiving device's* Decompression Module *receives the H.264 stream and decodes each frame back into 3D geometry. Finally, the* Visualization Module *renders the 3D video data for the remote user to interact with in real-time, optionally within their own physical environment using augmented reality.*

System Overview

This section will introduce the modules of the Holo Reality platform for mobile 3D video communications. Figure 1 provides an overview of the platform. As will be described, the platform utilizes a 3D range geometry encoding method, along with 2D video encoding, to achieve a 3D data stream that can be transmitted over low-bandwidth, wireless connections. Further, each module of the proposed platform runs entirely on a mobile device, without the need for additional or specialized hardware. These features of the proposed platform help to make 3D video transmission more directly accessible and could potentially enable advanced applications such as 3D video calls and 3D telemedicine.

Acquisition Module

The Acquisition Module for our platform consists of an Apple iPhone X's front-facing camera components. Among the device's front-facing sensors is an infrared camera, a dot projector, and a color camera. The infrared camera and dot projector comprise a structured light imaging device. The device uses its dot projector to encode the scene or the object to be captured. The infrared camera then captures the dot encoding and can recover a 3D range geometry frame from it. The iPhone X's color camera can be used to simultaneously acquire color texture information.

Using the iOS SDK on the iPhone X, the Acquisition Module is able to obtain synchronized depth maps, Z, and color texture images, C, in real-time. It is important for subsequent modules that the acquired depth maps and color texture images are rectified so that a pixel $Z(i,j)$ in the depth map aligns with a pixel $C(i,j)$ of the color texture frame. Figure 2 shows an example of a depth map, Z, a color texture image, C, a reconstructed 3D geometry frame, and a color texture mapped 3D geometry frame as captured by the Acquisition Module running on an iPhone X.

Compression Module

Based on the data resolution and frame rate, connection speeds of 1-2 Gbps are often needed to transmit 3D geometry video in real-time [7, 8, 9]. These speeds are very demanding and can often only be achieved over a wired network connection. In order to transmit the 3D and color video data to and from mobile devices wirelessly, methods of compressing the data are needed.

Our Compression Module's 3D data encoding method is derived from the multiwavelength depth encoding method [13] which encodes floating-point 3D range geometry into the color channels of a regular 2D image. The Acquisition Module provides a depth map, Z, and this can be encoded into the three color channels of an output RGB image, E, as follows:

$$E_r(i,j) = 0.5 + 0.5\sin\left[(Z(i,j)\times 2\pi)/P\right], \quad (1)$$

$$E_g(i,j) = 0.5 + 0.5\cos\left[(Z(i,j)\times 2\pi)/P\right], \quad (2)$$

$$E_b(i,j) = \left[Z(i,j) - \min(Z)\right] / \left[\max(Z) - \min(Z)\right]. \quad (3)$$

Here, (i,j) are the pixel indices and P is the *fringe width*, which is represented by the depth distance encoded by each fringe stripe.

Once the 3D range data has been encoded into a 2D RGB format, it can be further compressed using 2D image compression methods. To represent the raw data of each floating-point 3D range data frame and its associated color texture image (both at a resolution of 480×640) 4.6 MB is needed. This only in-

Figure 2. *Example frame as captured by the Acquisition Module on an iPhone X. From left to right: a depth map, Z; the depth map's corresponding color texture image, C; the 3D geometry frame reconstructed from Z; the 3D geometry frame mapped with its color texture image.*

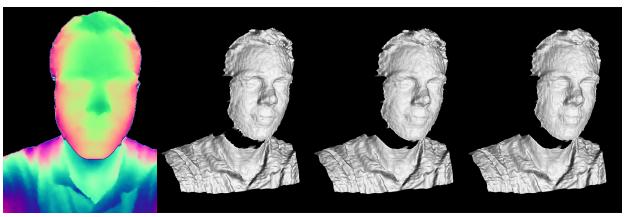


Figure 3. *Encoding and decoding at various 2D image qualities. From left to right: the encoded depth map, E; 3D reconstruction when E was compressed with lossless PNG; reconstruction when E was compressed with lossy JPEG 100; and reconstruction when E was compressed with lossy JPEG 80.*

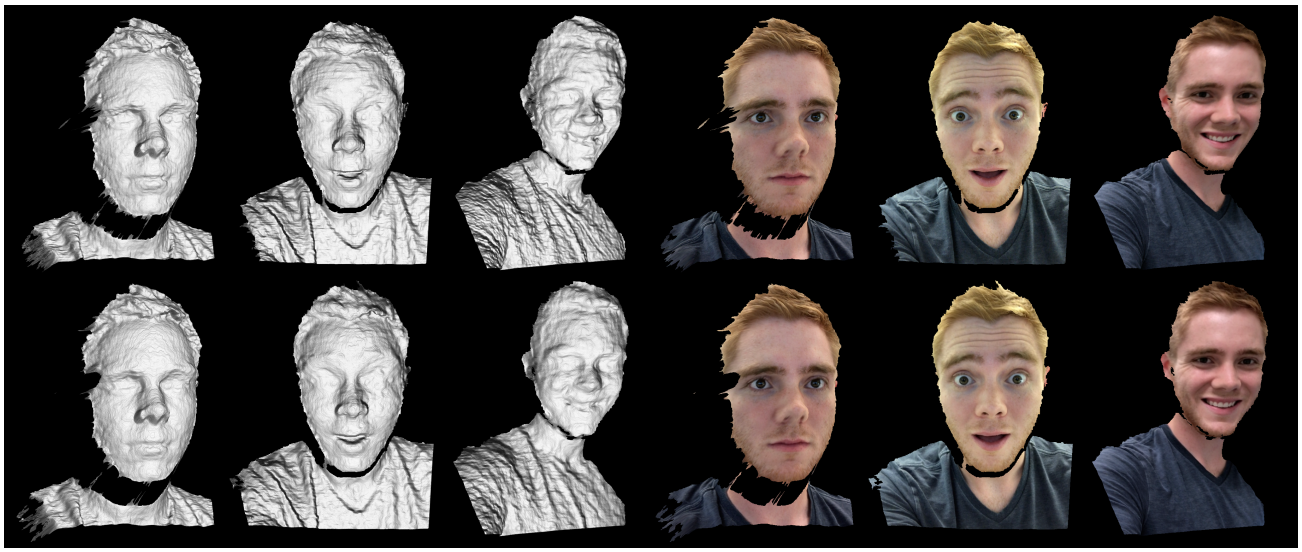


Figure 4. *Acquired 3D video data compared to decoded 3D video data. Top row: three original frames of 3D video data, rendered with and without its color texture data. Bottom row: three frames reconstructed from encoded H.264 video data received at an average bitrate of 12.1 Mbps (a 94.1:1 compression ratio).*

creases if additional information (e.g., normal map, mesh connectivity information) is desired. Using the data in Fig. 2 as an example, when its geometry and color data are encoded into a single RGB image and compressed using PNG, JPEG 100, and JPEG 80, file sizes of 576 KB (8.0:1 compression ratio), 300 KB (15.4:1 compression ratio), and 64 KB (71.5:1 compression ratio) can be achieved, respectively. Figure 3 shows reconstructions when the encoded image, E, is compressed at the various levels 2D image compression. In Fig. 3, the first image is the encoded image, E, and the second, third, and fourth images are 3D reconstructions from E when compressed with lossless PNG, lossy JPEG 100, and lossy JPEG 80, respectively.

Recall that the 4.6 MB needed to represent 3D range data and color texture was only for a single frame. Since our goal is to achieve real-time 3D video communications, it is desirable to transmit this data at rates of at least 30 Hz. This means that a connection speed of approximately 1.1 Gbps is needed to transmit just the raw data. Given this, in our Compression Module, each encoded frame E, in a mosaic format along with its color texture image, C, is further compressed using H.264 video compression. It is this H.264 video stream that is then used as input to the Transmission Module for sending to a remotely connected device. Since H.264 video compression is being used, the bandwidth needed to transmit the encoded video is adaptable. Typically, we targeted compressed video bitrates within the range of 5-15 Mbps, so that the resulting video could be delivered feasibly over a wireless network. At these bitrates, compression ratios versus the raw data are in the range of 73.7-221.2:1.

Figure 4 shows an example of data reconstructed from the encoded H.264 video stream. The first row shows several lossless reconstructions of 3D data acquired from the iPhone X, both with and without color texture mapping. The second row shows the 3D data reconstructed from the received lossy H.264 video stream. The second row's reconstructions were taken from an encoded H.264 video stream that was being delivered at an average bitrate of 12.1 Mbps, a 91.4:1 compression ratio versus the original data.

Transmission Module

To transmit encoded 3D geometry data within the H.264 video stream in real-time, WebRTC [14] was utilized via the QuickBlox iOS SDK [15]. WebRTC allows two devices to communicate with one another directly in a peer-to-peer fashion which helps facilitate the low-latency transmission of the encoded H.264 video that is crucial for real-time 3D video communications. Using WebRTC also enables transmission of encoded 3D video data to a wide variety of devices, including most modern web browsers which have native WebRTC support built-in.

Decompression Module

Upon obtaining encoded H.264 video from the Transmission Module, the Decompression Module is responsible for decoding the received data. For each frame of H.264 video that is received, a 3D frame can be recovered. Assuming our Decompression Module has just received encoded image, E, it is uploaded to the mobile device's GPU as a texture image of a flat plane mesh. The flat plane mesh is initialized to have the same number of vertices as pixels within the depth map, Z; this is consequently the same number of pixels that are in the encoded depth map, E. This relationship allows a single vertex, $(u,v,)$, in the mesh to correspond to a single pixel, (i,j), in the texture mapped image. Then, following the multiwavelength depth decoding procedure given in [13], for each pixel $E(i,j)$ in the encoded image, a depth coordinate, $Z'(u,v)$, can be can be recovered. Finally, $X'(u,v)$ and $Y'(u,v)$ coordinates can be derived using each recovered depth value, $Z'(u,v)$, and the device's calibration information:

$$X'(u,v) = \frac{Z'(u,v) \times (u-u_0)}{f_u}, \tag{4}$$

$$Y'(u,v) = \frac{Z'(u,v) \times (v-v_0)}{f_v}. \tag{5}$$

Here, f_u and f_v are the focal lengths along the u and v directions; u_0 and v_0 are the offsets of the principal point; and (u,v) are the vertex indices corresponding to pixel indices (i,j).

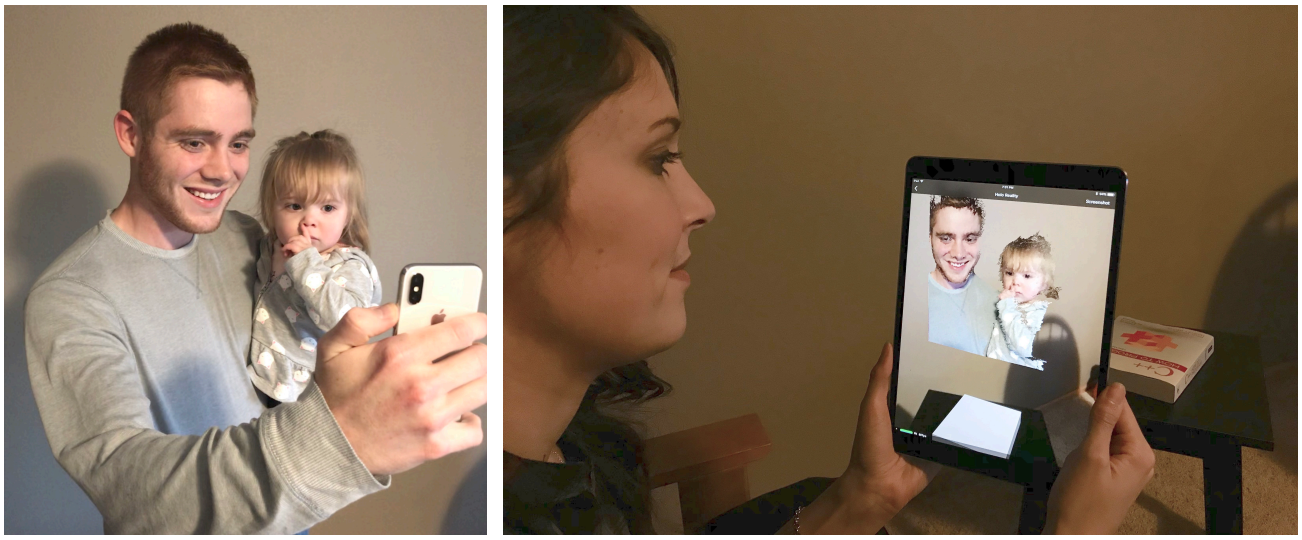


Figure 5. *The Holo Reality platform in use. Left: two people using the Holo Reality platform on an iPhone X to capture, encode, and wirelessly transmit 3D video of themselves in real-time. Right: the receiving user visualizing received 3D video content within their physical environment using augmented reality.*

After a vertex (u,v) has calculated its new coordinate, $[X'(u,v),Y'(u,v),Z'(u,v)]$, the vertex updates its position in the mesh accordingly. The geometry decoding of the Decompression Module takes place entirely on the GPU within the vertex shader, allowing for parallel processing and reconstruction of 3D data. Lastly, the color texture mapping of the reconstructed 3D geometry takes place within the mesh's fragment shader.

Visualization Module

The Visualization Module allows for the received, decompressed, and reconstructed 3D video data to be visualized and interacted with. As mentioned above, this module is implemented entirely on mobile devices. Following the method mentioned in the Decompression Module, a mesh plane is first initialized for visualization. Incoming encoded H.264 video then is texture mapped onto this plane for 3D reconstruction.

Currently, the Visualization Module is implemented using SceneKit [16] which can perform its rendering utilizing either the Metal or OpenGL graphics technology. Reconstructed 3D video content can be interacted with in real-time within the virtual graphics environment, allowing the user to rotate, zoom, and pan the received 3D video content. In general, as long as the H.264 video stream can be received for decoding, a Visualization Module can be implemented for a wide variety of devices (e.g., desktop, web, mobile, virtual reality, augmented reality). To demonstrate this, and to highlight our platform's potential to enable advanced applications within telepresence and telecollaboration, our Visualization Module also can display reconstructed 3D video content using augmented reality. Details and an example of this visualization will be given in the following section.

Results

We developed a demonstration system to test the performance of our proposed Holo Reality platform. As mentioned above, all of the modules were implemented on consumer mobile devices. In our demonstration system, a single Apple iPhone X (Model: A1865) was used to run the Acquisition, Compression, and Transmission Modules. The device's front-facing TrueDepth camera was used to provide depth maps of 480×640 at 30 Hz, which provides up to 307,200 unique 3D coordinates per frame. In addition, the front-facing color camera of the device was used to provide color texture frames with the same resolution of 480×640. Both of these frames were synchronously acquired at 30 Hz so that the most recently obtained depth map corresponded to the most recently obtained color texture image. The frames were then encoded and compressed as described above before being wirelessly transmitted using WebRTC.

The Transmission, Decompression, and Visualization Modules of our platform were implemented on an iPad Pro (Model: A1701). This device wirelessly received 3D video data, performed the decompression and 3D reconstruction, and was used to visualize the 3D video content within an augmented reality environment. As the iPad user moves the device around the real world, the real-time 3D video content is anchored within their physical space, giving the appearance that the remotely captured user has a physical presence within their environment.

Figure 5 shows photographs of the proposed Holo Reality platform in use. The left image shows two people with the iPhone X using the platform to capture, compress, and transmit 3D video data of themselves. The right image shows a third person with the iPad Pro using augmented reality to visualize the reconstructed 3D video content in real-time within their physical space. In this case, the 3D video content is anchored within the iPad user's environment using the textbook (on the table in the background) as an anchor. This gives the receiving user the ability to select and move where they want incoming 3D video to be visualized within their current space. In this demonstration, the augmented reality Visualization Module renders received 3D video at a 1:1 scale. This is done to give the receiving user a more natural and intuitive experience when interacting with the 3D video data, however, given that everything is virtual, the scale of the rendered 3D video could be smaller or larger depending on the desired effect.

Since H.264 video compression is being used, the bandwidth needed to transmit the encoded video is flexible and can easily be

adapted to fit the desired application or usage environment. For our example of an augmented reality 3D video call, we found that an average bandwidth of 10.0-12.5 Mbps provided good geometry and color texture fidelity while still being feasibly transmitted and received wirelessly by the mobile devices at low latencies (typically 0.1-0.5 seconds). In situations where reconstruction quality can be lowered, higher levels of H.264 compression can be applied to reduce the bitrate of transmitted video. Similarly, in situations where there are reliable network resources, the bitrate can be increased to achieve a higher quality geometry encoding.

Summary

This paper proposed Holo Reality, a novel platform for achieving real-time, low-bandwidth 3D video communications entirely on consumer mobile devices. Our demonstration system successfully delivered 3D video content from one mobile device to another, in real-time, over wireless networks. Our platform achieved frame rates of 30 Hz throughout, allowing for real-time, low-latency 3D video communications. This paper also highlighted the system's ability to use augmented reality for the real-time visualization of received 3D video content. Our platform's capability to provide low-bandwidth 3D video communications entirely on mobile devices, while requiring no additional or specialized hardware devices, has the potential to help realize a broad array of 3D-enabled applications, including those within areas such as telepresence, telecollaboration, and telemedicine.

References

[1] Song Zhang. Recent progresses on real-time 3d shape measurement using digital fringe projection techniques. *Optics and Lasers in Engineering*, 48(2):149 – 158, 2010. Fringe Projection Techniques.

[2] Henry Fuchs, Gary Bishop, Kevin Arthur, Leonard McMillan, Ruzena Bajcsy, Sang Lee, Hany Farid, and Takeo Kanade. Virtual space teleconferencing using a sea of cameras. In *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*, volume 26, 1994.

[3] Herman Towles, Wei chao Chen, Ruigang Yang, Sang uok Kum, Henry Fuchs Nikhil Kelshikar, Jane Mulligan, Kostas Daniilidis, Henry Fuchs, Carolina Chapel Hill, Nikhil Kelshikar Jane Mulligan, Loring Holden, Bob Zeleznik, Amela Sadagic, and Jaron Lanier. 3d tele-collaboration over internet2. In *In: International Workshop on Immersive Telepresence, Juan Les Pins*, 2002.

[4] Markus Gross, Stephan Würmlin, Martin Naef, Edouard Lamboray, Christian Spagno, Andreas Kunz, Esther Koller-Meier, Tomas Svoboda, Luc Van Gool, Silke Lang, Kai Strehlke, Andrew Vande Moere, and Oliver Staadt. Blue-c: A spatially immersive display and 3d video portal for telepresence. *ACM Trans. Graph.*, 22(3):819–827, 2003.

[5] Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec. Achieving eye contact in a one-to-many 3d video teleconferencing system. *ACM Trans. Graph.*, 28(3):64:1–64:8, 2009.

[6] A. Maimone and H. Fuchs. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 137–146, 2011.

[7] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, 2013.

[8] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 741–754, New York, NY, USA, 2016. ACM.

[9] Tyler Bell, Jan P. Allebach, and Song Zhang. Holostream: High-accuracy, high-speed 3d range video encoding and streaming across standard wireless networks. In *IS&T International Symposium on Electronic Imaging 2018*, pages 425–1–425–6, 2018.

[10] Joaquim Salvi, Sergio Fernandez, Tomislav Pribanic, and Xavier Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666 – 2680, 2010.

[11] Tyler Bell, Bogdan Vlahov, Jan P. Allebach, and Song Zhang. Three-dimensional range geometry compression via phase encoding. *Appl. Opt.*, 56(33):9285–9292, 2017.

[12] Yatong An, Jae-Sang Hyun, and Song Zhang. Pixel-wise absolute phase unwrapping using geometric constraints of structured light system. *Opt. Express*, 24(16):18445–18459, 2016.

[13] Tyler Bell and Song Zhang. Multiwavelength depth encoding method for 3d range geometry compression. *Appl. Opt.*, 54(36):10684–10691, 2015.

[14] WebRTC. WebRTC. `https://webrtc.org`.

[15] QuickBlox. QuickBlox iOS SDK. `https://quickblox.com/developers/IOS`.

[16] Apple. SceneKit. `https://developer.apple.com/scenekit`.

Author Biography

Prof. Tyler Bell is an Assistant Professor of Electrical and Computer Engineering at the University of Iowa. He leads the Holo Reality Lab and is a faculty member of the Public Digital Arts (PDA) cluster. Tyler received his Ph.D. from Purdue University in 2018. His current research interests include high-quality 3D video communications; high-speed, high-resolution 3D imaging; virtual reality, augmented reality; human computer interaction; and multimedia on mobile devices.

Prof. Song Zhang is an associate professor of mechanical engineering at Purdue University. He received his Ph.D. degree in mechanical engineering from Stony Brook University in 2005. His major research interests are superfast 3D optical sensing, 3D biophontic imaging, 3D geometry/video analysis, human and computer interaction, and virtual reality. Dr. Zhang has published over 100 research articles including more than 70 journal papers; edited one book; and holds 3 US patents.

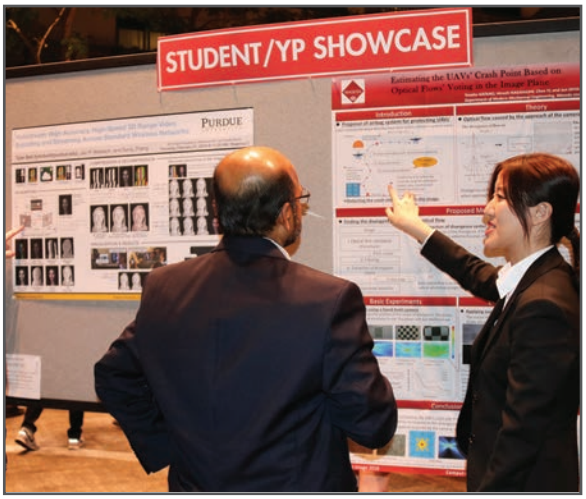
JOIN US AT THE NEXT EI!

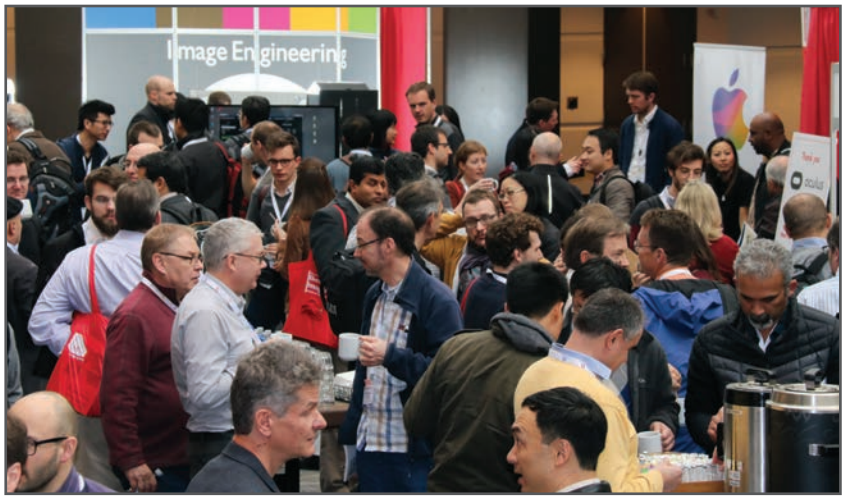
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org