MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Soft 2D-to-3D Delivery Using Deep Graph Neural Networks for Holographic-Type Communication

Fujihashi, Takuya; Koike-Akino, Toshiaki; Watanabe, Takashi

TR2023-031 May 06, 2023

Abstract

Holographic-type communication, i.e., three- dimensional (3D) content delivery, will be a crucial application for modern wireless and mobile networks. In this paper, we propose a novel soft delivery scheme to realize efficient 3D content delivery. Specifically, the proposed scheme sends a single 2D image over error-prone wireless channels using discrete cosine transform followed by near-analog modulation. At the receiver, a 2D-to-3D decoder based on graph neural networks (GNN) reconstructs the corresponding 3D point cloud and mesh from the received 2D image. We verify that the proposed soft 2D-to-3D delivery scheme can reconstruct clean 3D data gracefully from the soft-delivered 2D image even in the presence of fading and noise distortion. In addition, the proposed scheme can generate higher-quality 3D data compared with direct 3D content delivery schemes.

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2023

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Soft 2D-to-3D Delivery Using Deep Graph Neural Networks for Holographic-Type Communication

Takuya Fujihashi*, Toshiaki Koike-Akino†, Takashi Watanabe*
*Graduate School of Information Science and Technology, Osaka University, Japan
†Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA

***Abstract*—Holographic-type communication, i.e., three-dimensional (3D) content delivery, will be a crucial application for modern wireless and mobile networks. In this paper, we propose a novel soft delivery scheme to realize efficient 3D content delivery. Specifically, the proposed scheme sends a single 2D image over error-prone wireless channels using discrete cosine transform followed by near-analog modulation. At the receiver, a 2D-to-3D decoder based on graph neural networks (GNN) reconstructs the corresponding 3D point cloud and mesh from the received 2D image. We verify that the proposed soft 2D-to-3D delivery scheme can reconstruct clean 3D data gracefully from the soft-delivered 2D image even in the presence of fading and noise distortion. In addition, the proposed scheme can generate higher-quality 3D data compared with direct 3D content delivery schemes.**

***Index Terms*—3D mesh, deep graph neural network (GNN)**

I. Introduction

Recent advances in wide-band wireless technology, such as the fifth-generation (5G) networks, have inspired many emerging applications. Holographic-type communication [1], [2], i.e., three-dimensional (3D) content delivery, is a crucial application for the next generation of wireless and mobile networks to provide highly immersive experiences for users through extended reality (XR) and holographic devices [3], [4].

There are two main representations to display the 3D content on XR devices: point cloud and mesh formats [5]. A point cloud consists of many points in the 3D space. Each point typically has 3D coordinates, and color attributes to represent the shape and color patterns of the 3D content. A mesh utilizes faces and textures to represent the shape and color patterns. A key challenge in 3D content delivery over wireless channels is efficiently sending high-quality 3D content within a limited bandwidth. Some compression methods [6]–[9] have been proposed for both point cloud and mesh to deliver the 3D data under the bandwidth constraint. Typically, uniform/non-uniform quantization with an arithmetic code is used to compress the 3D content into a bitstream.

However, the compressed bitstream is easily affected by fading fluctuation of wireless channels since the arithmetic coding for entropy coding has all-or-nothing behavior. Although channel coding is usually used for the bitstream to tackle the issues of fluctuation, it also has all-or-nothing behavior. When the channel signal-to-noise ratio (SNR) falls below a certain threshold, bit errors will cause fatal distortion on the reconstructed 3D data [10], [11].

To solve the abovementioned issues, the existing studies designed soft delivery [13]–[16] for directly sending the 3D content via wireless channels. HoloCast [17]–[19] and HoloCast+ [20] are pioneer works on soft 3D point cloud delivery for unstable wireless channels. The key idea of both schemes is to adopt near-analog modulation, i.e., mapping the transformed coefficients onto the transmission symbols, for the 3D content delivery. It enables graceful 3D reconstruction according to the wireless channel quality, without sudden quality degradation.

This study proposes an alternative soft delivery scheme for 3D content, i.e., 2D-to-3D soft delivery, to realize a better solution for holographic-type communication. Specifically, the transmitter only sends a single 2D image in an analog manner, and the receiver reconstructs the corresponding 3D data of the point cloud and meshes from the received 2D image. To this end, the receiver uses convolutional neural networks (CNN) and graph convolutional neural networks (GCNN) for the 2D-to-3D reconstruction. The CNN and GCNN are suited for regular-structured 2D pixels and irregular-structured 3D point cloud signals, respectively [12], [21].

There are three major advantages for our 2D-to-3D content delivery compared with direct 3D content delivery. Firstly, the user can reconstruct clean 3D data even with deep fading and severe noise in 2D image delivery. Secondly, a decorrelation in the 2D image delivery benefits the reconstruction quality of the 3D content. Lastly, 2D-to-3D delivery can significantly reduce the data size. The existing works [12], [21] consider the 2D-to-3D reconstruction from an error-free 2D image, i.e., content delivery over wired networks.

Our contribution is three-fold as follows:

- we demonstrate that the proposed scheme can compensate for wireless channel distortion;
- we show that the trained model in the proposed scheme is resilient to channel fluctuation and bandwidth variations between training and testing, and
- we verify that the proposed 2D-to-3D content delivery can outperform the existing 3D content delivery.

II. Proposed 2D-to-3D Soft Delivery

Fig. 1 shows the overview of the proposed soft 2D-to-3D content delivery scheme, which consists of a 2D image encoder, 2D image decoder, and 2D-to-3D decoder. The 2D image encoder takes 2D-discrete cosine transform (2D-DCT)

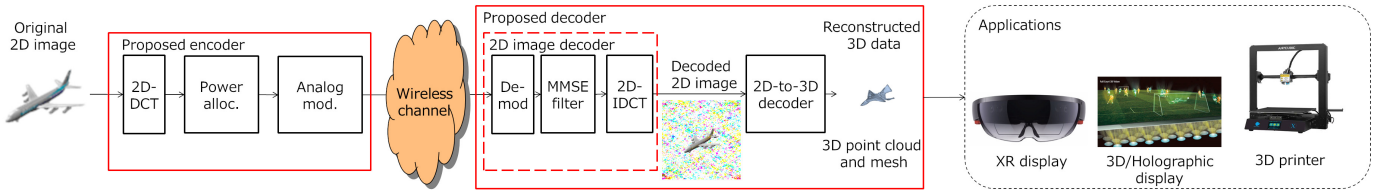


Fig. 1. End-to-end soft 2D-to-3D content delivery systems for immersive applications.

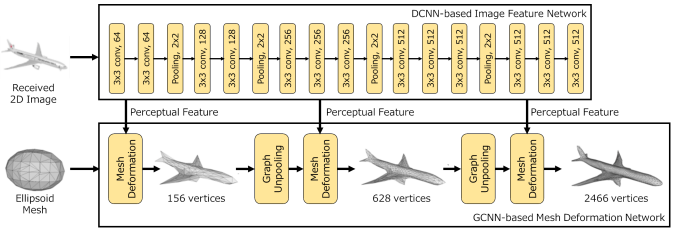


Fig. 2. Network architecture of our 2D-to-3D decoder based on pixel2mesh [12].

to transform the 2D image into the corresponding DCT coefficients. The encoder assigns a transmission power to map each DCT coefficient onto a transmission symbol. A channel fading with an additive noise impairs the transmission symbols during wireless transmission. The impaired DCT coefficients are denoised by a minimum mean-square error (MMSE) filter and transformed into the pixel-domain by inverse 2D-DCT (2D-IDCT). The reconstructed 2D image is finally fed into the 2D-to-3D decoder to obtain the 3D point cloud and mesh.

A. 2D Image Encoder

The 2D image encoder performs the 2D-DCT operation on the original 2D image to obtain the DCT coefficients. Each DCT coefficient is scaled before the near-analog modulation to achieve higher reconstruction quality of the 2D image.

Let x_i denote the ith analog-modulated symbol in either I (in-phase) or Q (quadrature) components. Each analog-modulated symbol is amplified by a scale factor g_i for noise reduction as: $x_i = g_i \cdot s_i$, where s_i is the ith DCT coefficient.

The encoder performs a near-optimal power control by adjusting g_i for each DCT coefficient. We consider the best g_i is obtained by minimizing the MSE under the power constraint with a total power budget P. For this case, the near-optimal solution is expressed as:

$$g_i = \lambda_i^{-1/4} \sqrt{\frac{P}{\sum_j \sqrt{\lambda_j}}}, \tag{1}$$

where $\lambda_i = s_i^2$ is the power of the ith DCT coefficient.

B. 2D Image Decoder

After transmission over the wireless channels, each symbol obtained at the receiver can be modeled as: $y_i = h_i x_i + n_i$, where y_i is the ith received symbol, h_i is the channel gain, and n_i is an effective noise having a variance of σ^2. The receiver extracts DCT coefficients from I and Q components, using the MMSE filter [22] as follows:

$$\hat{s}_i = \frac{h_i g_i \lambda_i}{h_i^2 g_i^2 \lambda_i + \sigma^2} \cdot y_i. \tag{2}$$

The original 2D image is reconstructed by taking 2D-IDCT for the filter output $\hat{s}_i$.

C. 2D-to-3D Decoder

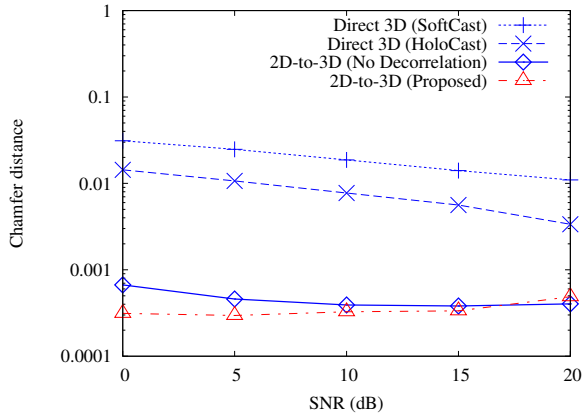
The proposed scheme then reconstructs the 3D point cloud and mesh from the reconstructed 2D image based on a deep learning framework [12] as shown in Fig. 2. It consists of a DCNN-based image feature network and a GCNN-based mesh deformation network. The image feature network consists of multiple 2D-CNN layers to extract perceptual features from the 2D image. The mesh deformation network leverages the extracted feature to deform an ellipsoid mesh into the desired 3D data progressively. The GCNN-based mesh deformation network includes three deformation blocks intersected by two graph unpooling layers. Each deformation block takes an input graph representing the current mesh model with the 3D shape feature attached to vertices and produces new vertices' locations and features. The graph unpooling layers gradually increase the number of vertices to improve the capacity to handle finer details. This architecture can learn how to deform and add details in the 3D model in a coarse-to-fine fashion.

We use a combination of Chamfer distance, surface, Laplacian regularization, and edge length regularization as a loss function to train both networks to decode accurate 3D data from the given 2D image. Specifically, the Chamfer distance constrains the 3D coordinates of the vertices, the surface loss keeps the consistency of surface normal, the Laplacian regularization maintains relative 3D coordinates between neighboring vertices during deformation, and the edge length regularization prevents outliers in the reconstructed 3D data.

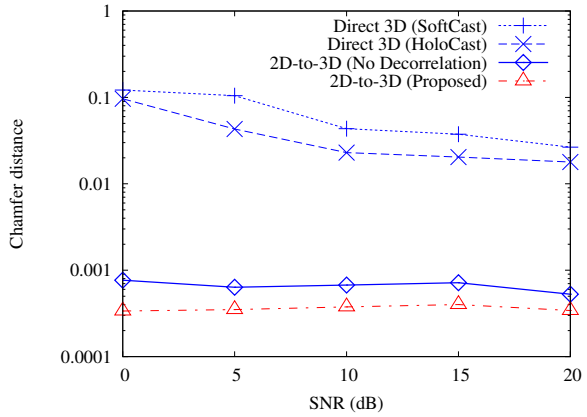
III. Performance Evaluation

A. Simulation Settings

*1) **Datasets**:* We use a benchmark dataset of ShapeNet [23] for experiments. ShapeNet contains unique 3D points from 55 categories. In our experiments, we select point clouds of the "Plane" category as an example. We use 77,663 samples of 2D images and the corresponding 3D point clouds for training and 100 samples for testing. We set the total number of training epochs and the batch size to 30 and 50, respectively. We use the training data to learn the network parameters of



(a) Chamfer distance in AWGN channels



(b) Chamfer distance in Rayleigh fading channels

Fig. 3. Chamfer distance for the different channel SNRs in AWGN and Rayleigh fading channels.

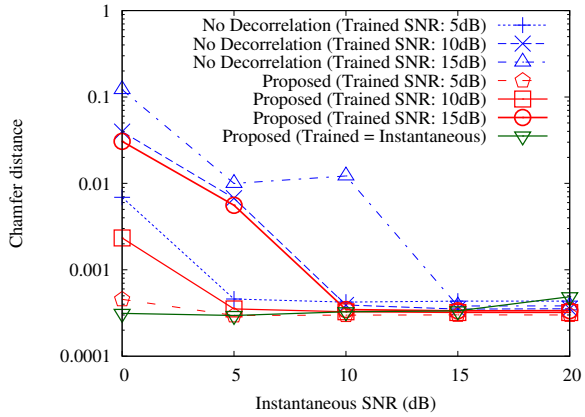


Fig. 4. Chamfer distance of the 2D-to-3D delivery schemes vs. instantaneous wireless channel SNRs in AWGN channels under the different SNRs in offline training.

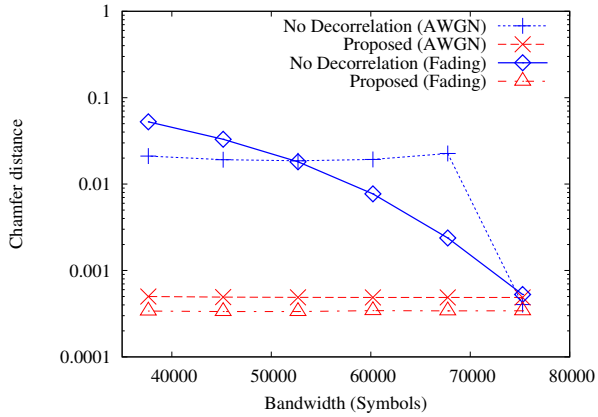


Fig. 5. Chamfer distance of the 2D-to-3D delivery schemes vs. available bandwidth in AWGN and Rayleigh fading channels for 20 dB channel SNR.

the 2D-to-3D decoder, while the testing data to evaluate 3D reconstruction and visual quality. Here, we consider a 2D image resolution of 224×224 pixels and 156 vertices with 462 edges for an initial ellipsoid mesh.

*2) **Quality Metric**:* We use the Chamfer distance [24] for performance metrics of 3D reconstruction quality. The Chamfer distance is commonly used to measure the similarity between the 3D points.

*3) **Wireless Environment**:* We consider Rayleigh fading channels with an additive noise n_i for realistic wireless environments. The additive noise n_i follows circular-symmetry complex white Gaussian distribution with a variance of σ^2, i.e., $n_i \sim \mathcal{CN}(0, \sigma^2)$. For simplicity, we assume that the channel coefficients are known at the receiver and transmitter.

*4) **2D-to-3D Decoding Training**:* We implement our 2D-to-3D decoder using Pixel2Mesh in PyTorch.The network is optimized using Adam with a weight decay of 10^{-5}. We initialize the learning rate as 3×10^{-5}.

B. 3D Reconstruction Quality in AWGN and Fading Channels

We first discuss the baseline performance over additive white Gaussian noise (AWGN) and Rayleigh fading channels. Figs. 3 (a) and (b) show the Chamfer distance over AWGN and Rayleigh fading channels as a function of wireless channel SNRs. For comparison, we consider three schemes: 2D-to-3D delivery without decorrelation; SoftCast [22]; and HoloCast [17]. The 2D-to-3D delivery without decorrelation scheme directly maps the pixel values onto the transmission symbols using the near-analog modulation and then obtains the corresponding 3D data using the proposed 2D-to-3D decoder. SoftCast and HoloCast decorrelate the 3D vertices using 1D-DCT and graph Fourier transform (GFT) [25] and map the coefficients onto the transmission symbols.

We can see that the proposed scheme achieves the best performance and maintains constantly high 3D reconstruction quality irrespective of wireless channel SNRs. In addition, direct 3D delivery schemes suffer from low reconstruction quality compared with 2D-to-3D delivery schemes.

C. Impact of Channel Quality Fluctuation

The previous evaluations considered an adaptive 2D-to-3D decoder depending on channel SNRs. Precisely, the channel SNRs at training and testing are consistently adjusted. It is

(a) Original 2D image | (b) Decoded 2D image No Decorrelation (SNR: 5 dB) MSE: -6.6 dB | (c) Decoded 2D image Proposed (SNR: 5dB) MSE: -33.8 dB | (d) Reconstructed mesh No Decorrelation (SNR: 5 dB) Chamfer distance: 0.014178 | (e) Reconstructed mesh Proposed (SNR: 5 dB) Chamfer distance: 0.000164

Fig. 6. Snapshots of the decoded 2D images and reconstructed 3D mesh in the comparative schemes in AWGN channels.

(a) No Decorrelation (SNR: 5 dB) (b) Proposed (SNR: 5 dB)

Fig. 7. Snapshots of 3D printed models delivered at a channel SNR of 5 dB in AWGN channels.

more practical when one 2D-to-3D decoder trained at a certain SNR is non-adaptively re-used during 2D image delivery regardless of channel SNRs. We discuss the impact of the mismatch between the training and testing channel SNRs on the Chamfer distance.

Fig. 4 shows the Chamfer distance of the 2D-to-3D delivery schemes as a function of instantaneous channel SNRs in AWGN channels. Here, we compare three models trained at a channel SNR of 5, 10, and 15 dB. In both schemes, the model trained at a lower channel SNR can improve performance irrespective of the instantaneous SNRs. When the instantaneous channel SNR is lower than the trained SNR, the mismatch causes degraded 3D reconstruction quality.

D. Impact of Bandwidth Fluctuation

Fig. 5 shows the Chamfer distance of the 2D-to-3D delivery schemes as a function of the available bandwidth in AWGN and Rayleigh fading channels at a channel SNR of 20 dB. The proposed scheme keeps almost the same Chamfer distance even in the band-limited channels. On the other hand, no decorrelation scheme suddenly degrades the Chamfer distance when the available bandwidth is limited. It implies that the proposed scheme has stronger resilience to bandwidth fluctuation.

E. 2D Visual and 3D Printed Quality

Finally, we show some examples of visual snapshots of the decoded 2D image and reconstructed 3D mesh for the proposed and existing schemes. Figs. 6 (a) and (e) show original 2D images selected from the ShapeNet dataset. Figs. 6 (b) and (c) show the decoded 2D images of delivery schemes at a wireless channel SNR of 5 dB. In addition, Figs. 6 (d) and (e) show the reconstructed mesh of delivery schemes at a wireless channel SNR of 5 dB. The proposed scheme can reconstruct 3D data with clean aircraft tail and wings even in a low channel SNR.

As a proof-of-concept application, we also discuss the 3D reconstruction quality using a 3D-printed mesh model. We used a commercial 3D printer of Anycubic Mega X with polylactic acid (PLA) for fabrication. Once the 3D mesh was reconstructed through wireless transmission in the proposed 2D-to-3D delivery, the mesh was exported to Ultimaker Cura—an open-source 3D printer slicing application—for preprinting. We set Cura's printing parameters to Anycubic Mega i3 and generated a gcode file for printing the 3D mesh model.

Figs. 7 (a) and (b) show the 3D printed models, which correspond to the reconstructed mesh data selected from the ShapeNet dataset. As expected, the printed model in the proposed scheme can reproduce relatively cleaner wings.

IV. Conclusions

Our study is the first paper investigating 2D-to-3D content delivery over wireless channels. To obtain a clean 3D point cloud and mesh at the receiver in fading channels, we proposed a novel soft 2D-to-3D delivery scheme. We verified that the proposed scheme achieves constantly high 3D reconstruction quality irrespective of wireless channel SNRs. In addition, it was shown that the signal decorrelation offers a high resilience against bandwidth limitation. We further demonstrated remote 3D printing applications as proof-of-concept validation.

In our future work, we will evaluate the performance of the proposed scheme with state-of-the-art of digital-based solutions such as geometry-based point cloud compression (PCC) and video-based PCC [26] to clarify the benefits of our analog-based scheme against the digital-based solutions.

REFERENCES

[1] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. D. Turck, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, 2020.

[2] I. Akyildiz and H. Guo, "Holographic-type communication: A new challenge for the next decade," vol. 3, pp. 421–442, 09 2022.

[3] P. Su, W. Cao, J. Ma, B. Cheng, X. Liang, L. Cao, and G. Jin, "Fast computer-generated hologram generation method for three-dimensional point cloud model," *Journal of Display Technology*, vol. 12, no. 12, pp. 1688–1694, 2016.

[4] P.-A. Blanche, A. Bablumian, R. Voorakaranam, C. Christenson, W. Lin, T. Gu, D. Flores, P. Wang, W.-Y. Hsieh, M. Kathaperumal, B. Rachwal, O. Siddiqui, J. Thomas, R. Norwood, M. Yamamoto, and N. Peyghambarian, "Holographic three-dimensional telepresence using large-area photorefractive polymer," *Nature*, vol. 468, pp. 80–3, 11 2010.

[5] R. Mekuria and L. Bivolarsky, "Overview of the MPEG activity on point cloud compression," in *Data Compression Conference*, 2016, p. 620.

[6] K. Mamou, T. Zaharia, and F. Prêteux, "TFAN: A low complexity 3D mesh compression algorithm," *Computer Animation and Virtual Worlds*, vol. 20, 2009.

[7] A. Lalos, I. Nikolas, E. Vlachos, and K. Moustakas, "Compressed sensing for efficient encoding of dense 3D meshes using model-based Bayesian learning," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 41–53, 2017.

[8] J. Kammerl, N. Blodow, R. B. Rusu, S. Gedikli, M. Beetz, and E. Steinbach, "Real-time compression of point cloud streams," in *IEEE International Conference on Robotics and Automation*, 2012, pp. 778–785.

[9] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, "3D is here: Point cloud library (PCL)," in *IEEE International Conference on Robotics and Automation*, 2011, pp. 1–4.

[10] H. Chui, R. Xiong, C. Luo, Z. Song, and F. Wu, "Denoising and resource allocation in uncoded video transmission," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 102–112, jul 2015.

[11] Q. Fan, D. J. Lilja, and S. S. Sapatnekar, "Adaptive-length coding of image data for low-cost approximate storage," *IEEE Transactions on Computers*, vol. 69, no. 2, pp. 239–252, 2020.

[12] N. Wang, Y. Zhang, Z. Li, Y. Fu, H. Yu, W. Liu, X. Xue, and Y.-G. Jiang, "Pixel2Mesh: 3d mesh model generation via image guided deformation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–15, 2020.

[13] X. W. Tang, X. L. Huang, F. Hu, and Q. Shi, "Human-perception-oriented pseudo analog video transmissions with deep learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9896–9909, 2020.

[14] H. Hadizadeh and I. V. Bajic, "Soft video multicasting using adaptive compressed sensing," *IEEE Transactions on Multimedia*, vol. 23, pp. 12–25, 2021.

[15] X. w. Tang, X. l. Huang, and F. Hu, "QoE-driven UAV-enabled pseudo-analog wireless video broadcast: A joint optimization of power and trajectory," *IEEE Transactions on Multimedia*, vol. 23, pp. 2398–2412, 2021.

[16] Y. Gui, H. Lu, F. Wu, and C. W. Chen, "LensCast: Robust wireless video transmission over MmWave MIMO with lens antenna array," *IEEE Transactions on Multimedia*, vol. 24, pp. 33–48, 2022.

[17] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "HoloCast: Graph signal processing for graceful point cloud delivery," in *IEEE International Conference on Communications*, 2019, pp. 1–7.

[18] ——, "Overhead reduction in graph-based point cloud delivery," in *IEEE International Conference on Communications*, 2020, pp. 1–7.

[19] S. Ueno, T. Fujihashi, T. Koike-Akino, and T. Watanabe, "Point cloud soft multicast for untethered xr users," *IEEE Transactions on Multimedia*, vol. PP, pp. 1–11, 2022.

[20] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Holocast+: Hybrid digital-analog transmission for graceful point cloud delivery with graph fourier transform," *IEEE Transactions on Multimedia*, vol. 24, pp. 2179–2191, 2022.

[21] C. Wen, Y. Zhang, Z. Li, and Y. Fu, "Pixel2Mesh++: Multi-view 3D mesh generation via deformation," in *International Conference on Computer Vision*, 2019, pp. 327–336.

[22] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *ACM Annual International Conference on Mobile Computing and Networking*, Las Vegas, NV, sep 2011, pp. 289–300.

[23] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An information-rich 3D model repository," *arXiv e-prints*, Dec. 2015.

[24] S. Chen, C. Duan, Y. Yang, D. Li, C. Feng, and D. Tian, "Deep unsupervised learning of 3D point clouds via graph topology inference and filtering," *IEEE Trans. Image Process.*, vol. 29, pp. 3183–3198, 2020.

[25] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[26] D. B. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, pp. 1–17, 2020.