

Wireless 3D Point Cloud Delivery Using Deep Graph Neural Networks

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

TR2021-066 June 04, 2021

Abstract

In typical point cloud delivery, a sender uses octreebased and graph-based digital video compression to send three dimensional (3D) points and color attributes. However, the digital-based schemes have an issue called the cliff effect, where the 3D reconstruction quality will be a step function in terms of wireless channel quality. To prevent the cliff effect subject to channel quality fluctuation, we have proposed a wireless point cloud delivery called HoloCast inspired by soft delivery. Although the HoloCast realizes graceful quality improvement according to instantaneous wireless channel quality, it requires large communication overheads. In this paper, we propose a novel scheme for soft point cloud delivery to simultaneously realize better 3D reconstruction quality and lower communication overheads. The proposed scheme introduces an end-to-end deep learning framework based on graph neural network (GNN) to reconstruct high-quality point clouds from its distorted observation under wireless fading channels. We demonstrate that the proposed GNN-based scheme can reconstruct a clean 3D point cloud with low overheads by removing fading and noise effects.

IEEE International Conference on Communications (ICC)

Wireless 3D Point Cloud Delivery Using Deep Graph Neural Networks

Takuya Fujihashi*, Toshiaki Koike-Akino[†], Siheng Chen^{†‡}, Takashi Watanabe*

*Graduate School of Information Science and Technology, Osaka University, Japan

[†]Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA

[‡]Shanghai Jiao Tong University, Shanghai, Minhang District 200240, China

Abstract—In typical point cloud delivery, a sender uses octree-based and graph-based digital video compression to send three-dimensional (3D) points and color attributes. However, the digital-based schemes have an issue called the cliff effect, where the 3D reconstruction quality will be a step function in terms of wireless channel quality. To prevent the cliff effect subject to channel quality fluctuation, we have proposed a wireless point cloud delivery called HoloCast inspired by soft delivery. Although the HoloCast realizes graceful quality improvement according to instantaneous wireless channel quality, it requires large communication overheads. In this paper, we propose a novel scheme for soft point cloud delivery to simultaneously realize better 3D reconstruction quality and lower communication overheads. The proposed scheme introduces an end-to-end deep learning framework based on graph neural network (GNN) to reconstruct high-quality point clouds from its distorted observation under wireless fading channels. We demonstrate that the proposed GNN-based scheme can reconstruct a clean 3D point cloud with low overheads by removing fading and noise effects.

Index Terms—Point cloud, deep graph neural network

I. INTRODUCTION

Three-dimensional (3D) holographic displays [1], [2] have emerged as attractive interface techniques for reconstructing 3D perceptual scenes that provide full parallax and depth information for human eyes. 3D holographic display can be widely used for many applications: entertainment, virtual training, and medical imaging. Specifically, such 3D holographic visualizations will play a more important role in the post-Coronavirus (COVID-19) society because the 3D data can realize high-presence in remote conferencing [3] and healthcare [4]. For example, holographic data of doctors and medical imaging provide more interactive verbal guidance in tele-surgery [5].

Point cloud [6] is one of data formats to represent 3D scenes/objects on the holographic display [7]. Our study focuses on one of future wireless multimedia delivery systems, i.e., wireless point cloud delivery systems, which send 3D point data to a remote display over wireless links to reproduce the corresponding 3D scenes/objects on the remote display. In contrast to conventional two-dimensional (2D) images, 3D points in point cloud data are massive, non-ordered, and non-uniformly distributed in space. One of major issues in point cloud delivery is how to compress such numerous and irregular structures of 3D points while keeping original 3D scenes/objects. For example, when the number of 3D points in the reference point cloud is 800,000, the amount of traffic



(a) Holographic Dis- (b) 3D Modeling [13] (c) AR [14]
play [12]

Fig. 1. Examples of 3D point applications.

without any compression will be approximately 38 Mbits [8]. A large traffic causes low 3D reconstruction quality in point cloud delivery over limited data-rate links in wireless communications.

For point cloud compression over wireless links, conventional schemes typically rely on digital encoding such as point cloud library (PCL) [9], [10], octree-based coding [11], and transformation-based coding schemes. Specifically, a sender takes quantization and entropy coding for the 3D point cloud to generate the compressed bitstream. Here, the compression rate of the bitstream is adaptively selected according to link capacity of wireless channels. The compressed bitstream is then transmitted over the channels by using channel coding and digital modulation schemes. Successful high-quality delivery of point clouds over wireless links can realize high presence in video applications such as virtual reality (VR) and augmented reality (AR) on remote display as shown in Fig. 1.

However, the conventional point cloud delivery suffers from the following problems due to the unreliable wireless channel and nonlinear point cloud compression. First, the encoded bitstream is highly vulnerable for bit errors [15]. When the channel signal-to-noise ratio (SNR) falls below a certain value, possible errors occurred in the bitstream during communications will disable point cloud restoration. This phenomenon is called the cliff effect [16]. Second, the 3D reconstruction quality will saturate even when the wireless channel quality is improved unless an adaptive rate control of source and channel coding is performed in real-time according to the rapid fading channels. This is called the leveling effect. Third, nonlinear quantization is a lossy process and its distortion cannot be recovered at the receiver. Finally, voxel-domain point cloud encoders [9], [10] have limited coding efficiency since it does not yield good energy compaction. Although conventional

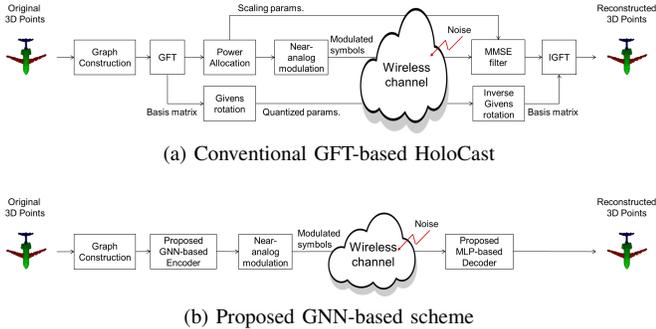


Fig. 2. Schematics of conventional and proposed schemes for wireless 3D point cloud delivery.

transform techniques, such as discrete cosine transform (DCT), can be used for point cloud data, they do not fully exploit the underlying irregular geometry of the 3D points.

To solve the above-mentioned issues, we have proposed HoloCast [17], [18] inspired by soft delivery [19]–[21] to realize graceful 3D reconstruction quality improvement with the improvement of wireless channel quality. Fig. 2(a) shows the overview of HoloCast. The key ideas of HoloCast are 1) skipping digital operations, i.e., quantization and entropy/channel coding, analogous to soft delivery schemes, e.g., SoftCast [19] and 2) introducing graph signal processing (GSP) [22] for better energy compaction. To this end, HoloCast regards the 3D points as vertices in a graph and takes graph Fourier transform (GFT) [23] to exploit the correlations between the adjacent graph signals, and directly sends the GFT coefficients by using near-analog modulation. Our HoloCast prevents cliff and leveling effects by skipping nonlinear digital operations and thus realizes better 3D reconstruction quality compared with digital-based schemes even in the presence of wireless channel fluctuation. The GFT-based soft delivery achieves better energy compaction in 3D point cloud compared with the conventional DCT-based soft delivery [19] and brings better 3D scene reconstruction irrespective of wireless channel quality. However, soft delivery schemes including SoftCast and HoloCast need a large communication overhead for signal decoding. Although overhead reduction techniques [18], [24] have been proposed for soft delivery schemes, they still need a large overhead and they are designed for additive white Gaussian noise (AWGN) wireless channels.

The main objective of our study is to propose a novel soft delivery scheme for 3D point clouds by introducing a new framework known as graph neural networks (GNN) [25] to simultaneously realize better 3D reconstruction quality and low communication overheads. GNN is a novel model for graph representation learning, which allows analyzing the irregular geometric structure of graph data. We focus on an end-to-end (E2E) deep learning, i.e., GNN-based autoencoders (GAE) [26]–[29], to encode 3D point clouds into a compressed representation. One of the benefits in the GAE is to allow the graph signal reconstruction from the limited number of latent variables without requiring additional metadata.

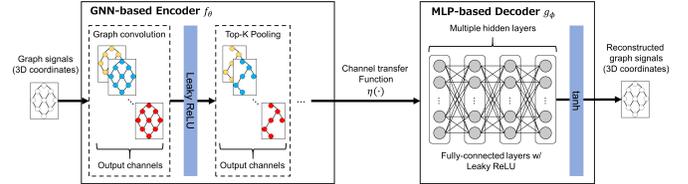


Fig. 3. Proposed GNN-based end-to-end encoder and decoder for wireless 3D point cloud delivery.

Fig. 2(b) shows the overview of the proposed scheme, where the GNN-based encoder transforms 3D points into several latent variables, and then the variables are directly mapped to transmission signals by means of near-analog modulation without relying on digital modulation schemes. The latent variables, which are distorted through wireless fading channels, are fed into another GNN-based decoder to reconstruct clean 3D points. Our GAE consists of a sequence of graph convolution and pooling operations with nonlinear activation functions for the extraction of important features while fully-connected layers with nonlinear activation functions for the 3D points reconstruction from the extracted features.

The contribution of this paper is three-fold:

- We verify that the proposed GAE realizes better 3D reconstruction quality compared with the conventional soft delivery schemes including SoftCast and HoloCast over fading channels;
- We confirm that the proposed GNN-based encoder can reduce the amount of communication overhead by about one order of magnitude;
- We demonstrate that adaptive channel precoding offers further quality improvement by means of the diversity gain of the rapid fading channels.

II. GNN-BASED SOFT POINT CLOUD DELIVERY

Fig. 3 shows the proposed E2E point cloud encoder and decoder. It can prevent the cliff/leveling effects in 3D scene reconstruction, gracefully improve reconstruction quality along with channel quality, and reduce the amount of overhead.

Encoder: The encoder part regards the 3D point cloud as a graph signal using a weighted and undirected graph $\mathcal{G} = (\mathbf{V}, \mathcal{E}, \mathbf{W})$. Here, \mathbf{V} and \mathcal{E} are the vertex and edge sets of \mathcal{G} , respectively. \mathbf{W} is an adjacency matrix with positive edge weights and the (i, j) th entry of \mathbf{W} represents the weight of an edge between vertices i and j .

In this paper, we consider the 3D coordinates of the point cloud $\mathbf{p} = [x, y, z]^T \in \mathbb{R}^{3 \times N}$ as the vertices in the graph where N is the number of vertices. A K -nearest-neighbor graph is used for making the connection between the vertices. In this graph, each 3D point connects to its K closest 3D points. We use a binary adjacency matrix whose entry is either 1 or 0 to indicate connectivity.

The encoder maps the 3D coordinate attributes \mathbf{p} to m -dimensional and L -channel real-valued latent variables $\mathbf{z} \in \mathbb{R}^{m \times L}$ by means of an encoding function f_θ . The encoding

function f_θ is parameterized using graph convolutional neural networks (GCNN) with weights θ . The encoder consists of a series of graph convolution followed by leaky rectified linear unit (ReLU) activation function, Top-K pooling [30], and a normalization layer. The graph convolution layers extract the graph signal features and the nonlinear activation function allows to learn a non-linear mapping from the source signal to the coded signal. Top- K pooling layer chooses the largest K values from each channel to remain important features. The output of the last graph convolution layer is normalized such that $\|z\|^2 = mLP$, where P denotes the average transmission power.

Wireless Link: The coded variables z are sent over the communication channel by directly mapping to in-phase and quadrature (I-Q) symbols x for near-analog modulation. The wireless channel, denoted by η , introduces stochastic distortion to the transmission symbols. To optimize the proposed scheme under wireless communications, the channel transfer function η must be incorporated into the E2E GAE. We consider Rayleigh fading as a reasonable channel model. In Rayleigh fading, each analog-modulated symbol at the receiver can be modeled as follows: $y_i = h_i x_i + n_i$, where y_i is the i th received symbol, x_i is the i th analog-modulated symbol, h_i is the i th multiplicative fading coefficient, and n_i is an AWGN with an average noise variance of σ^2 . The fading coefficients in the Rayleigh fading channels follow the complex Gaussian distribution, i.e., $h_i \sim \mathcal{CN}(0, 1)$ where \sim means “distributed as” and $\mathcal{CN}(a, b)$ is a complex Gaussian distribution with a mean of a and a variance of b .

To reduce the impact of fading effects, we consider two equalization techniques at the sender and receiver, i.e., pre-equalization and post-equalization, for a channel transfer function η . The pre-equalization can be realized at the sender side by sending pre-equalized transmission symbol x_i to the receiver as $x_i = w_i z_i$ where w_i is a pre-equalizer weight. Although there are many variants of pre-equalizer, we assume a simple coherent pre-equalization: $w_i = h_i^*/|h_i|$ where $[\cdot]^*$ denotes the conjugate operation. In this case, the channel transfer function will be: $\eta_{\text{preeq}}(z_i) = |h_i|z_i + n_i$. The post-equalization can be realized at the receiver by taking an inverse operation of the fading attenuation. Specifically, the receiver takes the zero-forcing post-equalization such that $\hat{y}_i = y_i/h_i$, given the estimated fading coefficient h_i . In this case, the channel transfer function will be: $\eta_{\text{posteq}}(z_i) = z_i + n_i/h_i$.

In addition to pre-/post-equalization, we also consider pre-coding method which sorts the latent variables z according to the fading level $|h_i|$ in descending order. Such sorting may facilitate to optimizing the best latent variables for GNN to achieve diversity gain.

Decoder: Upon the receipt of distorted latent variables, the decoder uses a decoding function g_ϕ , based on a multi-layer perceptron (MLP) for 3D point cloud reconstruction. The decoder consists of a series of fully-connected layers and leaky ReLU with a trainable parameter set ϕ . The MLP decoder maps the distorted latent variables \tilde{z} into an estimate \hat{p} of the

3D coordinates. The last layer uses hyperbolic tangent (tanh) activation function.

Loss Function: The proposed GNN-based encoding and decoding functions are trained to minimize a loss function:

$$(\theta, \phi) = \arg \min_{\theta, \phi} \mathbb{E}_{\Pr(\mathbf{p}, \hat{\mathbf{p}}_{\theta, \phi})} [d(\mathbf{p}, \hat{\mathbf{p}}_{\theta, \phi})], \quad (1)$$

where $\hat{\mathbf{p}}_{\theta, \phi}$ is the reconstructed 3D coordinates via E2E GAE with encoding and decoding parameter sets of θ and ϕ , $\mathbb{E}[\cdot]$ is an expectation, $d(\mathbf{p}, \hat{\mathbf{p}}_{\theta, \phi})$ is a defined distortion function between the original and reconstructed 3D coordinate attributes, $\Pr(\mathbf{p}, \hat{\mathbf{p}})$ is the joint probability distribution of the original and reconstructed 3D coordinate attributes. Since the true distribution of the input attributes is often unknown and thus the expected distortion is also unknown. To learn better weights for the minimization of the expected distortion in Rayleigh fading channels, all potential distortions due to channel fading and additive noise are synthetically analyzed by the proposed scheme in off-line learning phase. We use adaptive momentum (ADAM) optimizer for weight learning with an initial learning rate of 0.005, batch size of 10, momentum of 0.9, and momentum2 of 0.999 for 500 epochs.

III. PERFORMANCE EVALUATION

A. Simulation Settings

Datasets: We use a benchmark dataset of ShapeNet [31] for experiments. ShapeNet contains more than 50,000 unique 3D points from 55 categories. In our experiments, we select point clouds of “Airplane” category as an example. We sample 2,115 point clouds for training and 234 point clouds for testing. The training data are used for learning the network weights while the testing data are used for comparison in terms of 3D reconstruction and visual quality.

Loss Function and Quality Metric: We use the augmented Chamfer distance [28] for the loss function of our scheme and the distortion function for the 3D coordinate attributes. The augmented Chamfer distance is essentially the Hausdorff distance between two 3D point clouds. Although the original Chamfer distance is one of typical quality metrics, the augmented Chamfer distance is more robust to prevent some ill cases. The augmented Chamfer distance $d_{\text{CH}}(S, \hat{S})$ is defined as

$$\max \left\{ \frac{1}{|S|} \sum_{\mathbf{p} \in S} \min_{\hat{\mathbf{p}} \in \hat{S}} \|\mathbf{p} - \hat{\mathbf{p}}\|_2, \frac{1}{|\hat{S}|} \sum_{\hat{\mathbf{p}} \in \hat{S}} \min_{\mathbf{p} \in S} \|\mathbf{p} - \hat{\mathbf{p}}\|_2 \right\}, \quad (2)$$

where S is the input point set and \hat{S} is the reconstructed point set. The term $\min_{\hat{\mathbf{p}} \in \hat{S}} \|\mathbf{p} - \hat{\mathbf{p}}\|_2$ enforces that any 3D coordinate \mathbf{p} in the original point cloud has a matching 3D point $\hat{\mathbf{p}}$ in the reconstructed point cloud, and the term $\min_{\mathbf{p} \in S} \|\mathbf{p} - \hat{\mathbf{p}}\|_2$ enforces the matching vice versa. The max operation enforces that the distance from S to \hat{S} and the distance vice versa have to be small simultaneously.

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., $n_i \sim \mathcal{CN}(0, \sigma^2)$. We consider the range of the noise power σ^2 is from 0 dB to -30 dB. We also assume that the channel coefficients are known at the receiver and/or transmitter. Detailed analysis of the error impact on the channel estimates h_i will be left as a future work.

GAE Architecture: We use PyTorch Geometric (PyG) [32] for the simulations of our GAE architecture. The encoder repeats a series of GCNConv [33] with the output channels between 12 and 48, leaky ReLU activation function, and Top-K pooling at the graph pooling ratio between 0.5 and 0.9 three times. The output of the last Top-K pooling layer is followed by a normalization layer which enforces the average power constraint. The decoder uses a series of fully-connected layer and leaky ReLU three times to reconstruct the 3D coordinate attributes from the distorted latent variables via a channel transfer function. Here, the output channels of the first and the second fully-connected layers are the same as the output channels of GCNConv while the output channels of the last fully-connected layer is 3.

B. Overhead Reduction

We firstly discuss an impact of the proposed GNN-based coding on the amount of communication overheads. Fig. 4 shows the 3D reconstruction quality over Rayleigh fading channels as a function of communication overheads at a wireless channel SNR of 20 dB. Here, the communication overhead represents the total number of transmission symbols in near-analog modulation for the proposed scheme. SoftCast additionally sends the power information of the analog-modulated symbols for decoding while HoloCast also sends GFT basis matrix for inverse GFT operations. For the comparison, all the schemes intentionally discard the number of analog-modulated symbols. We note that power information and GFT basis matrix are transmitted by binary phase-shift keying (BPSK) modulation format without channel coding. When SoftCast and HoloCast use a channel code for error protection, the communication overhead will become several-fold larger than the results in our evaluations.

We compare the proposed scheme with HoloCast [18] and SoftCast [19]. HoloCast uses octree decomposition and takes GFT for the graph signals in each octree block with the size of 1,000 to convert into spectrum domain by using the eigenvectors of the random-walk graph Laplacian matrix. HoloCast uses the Givens rotation with a quantization bit depth b between 2 and 12 to compress the eigenvectors for overhead reduction. SoftCast takes DCT-based decorrelation for 3D coordinate attributes and directly maps the power-allocated coefficients on the I-Q components. For decoding the power-allocated coefficients, it needs to send the power information of all the DCT coefficients as metadata. The power information is compressed by using a Huffman code. The proposed scheme uses GNN-based encoding and decoding for overhead reduction. Here, the proposed scheme uses precoding with a channel transfer function of post-equalization.

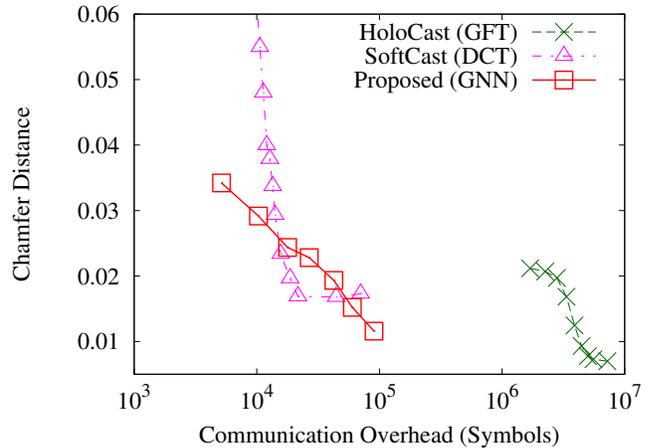


Fig. 4. 3D reconstruction quality as a function of communication overheads at a wireless SNR of 20 dB.

We can see that the proposed scheme achieves a significant overhead reduction at the same 3D reconstruction quality compared with the conventional HoloCast. For example, Chamfer distance of the proposed scheme is 0.011 at the communication overhead of 9.0×10^4 symbols. On the other hand, Chamfer distance of HoloCast is 0.009 at the communication overhead of 4.5×10^6 symbols. In this case, the proposed scheme achieves 98.0% overhead reduction to achieve the same 3D reconstruction quality. The conventional SoftCast has a limited 3D reconstruction quality irrespective of the communication overhead. Although SoftCast achieves better 3D reconstruction quality compared with the proposed scheme around the communication overhead of 2.2×10^4 symbols, the reconstruction quality sharply degrades as the communication overhead decreases. For a small communication overhead, SoftCast discards DCT coefficients in lower-frequency components. The discarded coefficients cause a significant impact on 3D reconstruction quality at the receiver.

C. 3D Reconstruction Quality

We now discuss an effect of wireless channel quality on the reconstructed point cloud quality. We consider two HoloCast schemes with a bit depth of 3 and 12 in the Givens rotation at an octree decomposition size of 1,000. Fig. 5 shows the 3D reconstruction quality over Rayleigh fading channels as a function of wireless channel SNRs. We observe the following results:

- The proposed scheme yields the best 3D reconstruction quality in low wireless SNR regimes.
- Although the conventional HoloCast scheme realizes better 3D reconstruction quality in high wireless SNR regimes, the required overhead is more than 10-times larger than that of the proposed method.
- The 3D reconstruction quality of SoftCast is lower than that of the proposed scheme irrespective of wireless channel SNRs.

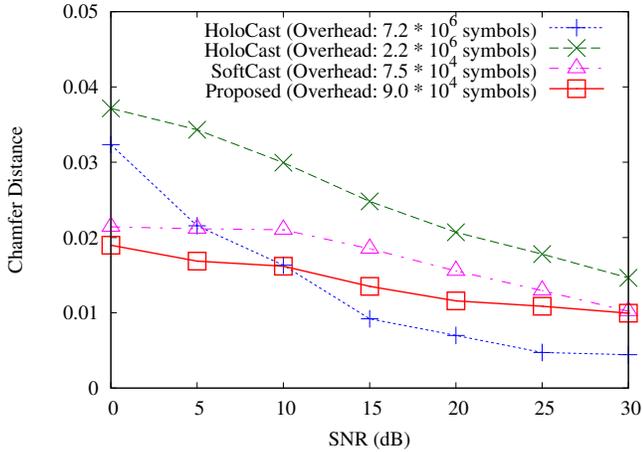


Fig. 5. 3D reconstruction quality as a function of wireless channel SNRs in Rayleigh fading channels.

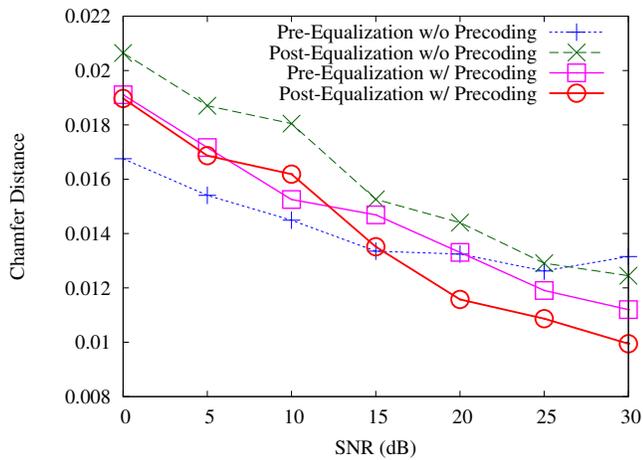


Fig. 6. 3D reconstruction quality of the proposed schemes over different channel transfer functions with/without precoding.

D. Impact of Precoding and Equalization

In this section, we evaluate the effects of precoding and equalizations, i.e., post-equalization and pre-equalization, on the 3D reconstruction quality of the proposed scheme. As mentioned in Sec. II, the channel transfer functions for the latent variables z_i in post-equalization and pre-equalization are $\eta_{\text{posteq}}(z_i) = z_i + n_i/h_i$ and $\eta_{\text{preeq}}(z_i) = |h_i|z_i + n_i$, respectively. In addition, the precoding method sorts the latent variables according to the fading coefficient in descending order for diversity gain.

Fig. 6 shows the 3D reconstruction quality of the proposed schemes over Rayleigh fading channels as a function of wireless channel SNRs for the case with different channel transfer functions of η_{preeq} and η_{posteq} with/without precoding. The evaluation results are summarized as follows:

- Precoding performs well in high wireless SNR regimes since it may achieve a higher diversity gain.

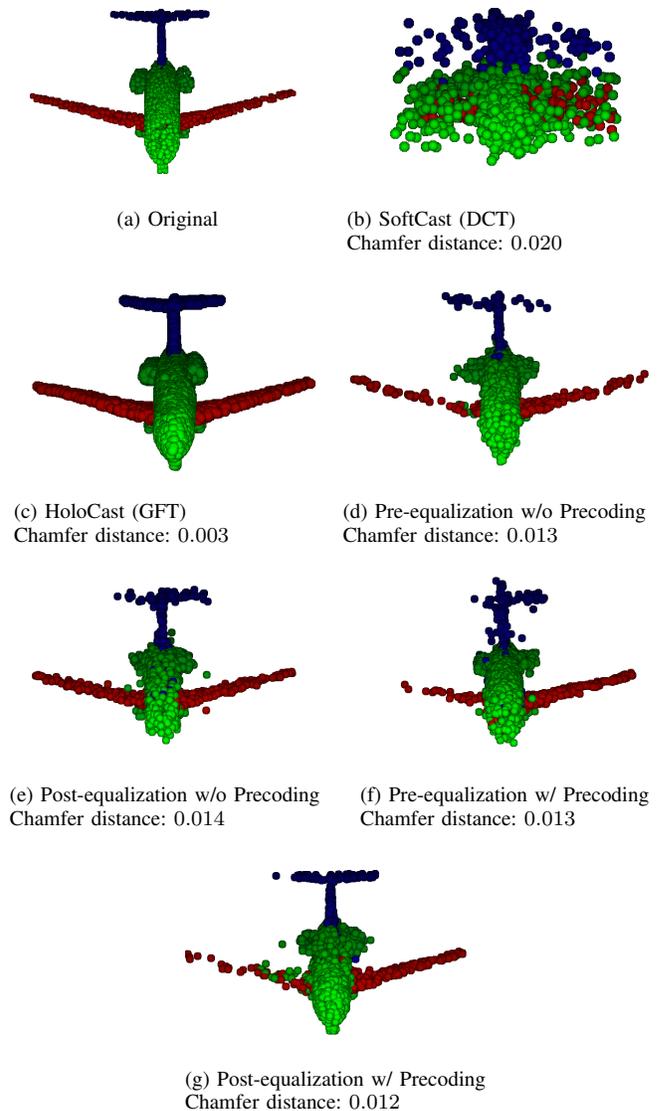


Fig. 7. Snapshot of reconstructed 3D point cloud over different channel transfer functions with/without precoding at a channel SNR of 20 dB.

- Pre-equalization works well at lower SNR regimes, whereas post-equalization does well at higher SNR regimes.
- As a consequence, pre-equalization without precoding yields the best 3D reconstruction quality at low SNR regimes below 10 dB.
- Accordingly, the post-equalization with precoding becomes the best one in the high SNR regimes above 10 dB.

We finally compare some examples of visual snapshots for SoftCast, HoloCast, and the proposed schemes over Rayleigh fading channels in Figs. 7(a) through (g) at a channel SNR of 20 dB. Here, the point cloud is selected from one point cloud from the test data in ShapeNet database. Although each proposed scheme may reconstruct the 3D shape of the aircraft, the proposed scheme with precoding may realize clear

reconstruction compared with the proposed scheme without precoding. In particular, SoftCast has an obvious degradation over other schemes. Nevertheless, the 3D shape of the aircraft tail remains noisy even with proposed methods. Note that we focused on a simplified GNN method compared with state-of-the-art techniques such as graph inception networks (GIN) [28] and FoldingNet [27] in order to demonstrate an initial proof-of-concept study of GNN-based 3D point cloud delivery. Extension to further improve 3D reconstruction quality will be considered as another follow-up work. To the best of authors' knowledge, this paper is the very first study exploiting GNN methods for wireless 3D point cloud delivery.

IV. CONCLUSIONS

We have proposed a novel scheme of soft point cloud delivery for future wireless streaming of holographic and 3D data. Specifically, the proposed scheme integrates GNN-based point cloud coding and near-analog modulation to simultaneously achieve: 1) prevention of the cliff effect, 2) prevention of the leveling effect, 3) high energy compaction, and 4) low communication overhead. In addition, the proposed E2E design of the GAE scheme accounts for random distortion due to fading channels through the use of pre-/post-equalization and precoding techniques. We demonstrated that the proposed scheme achieved a good trade-off between 3D reconstruction quality and communication overhead compared with the conventional SoftCast and HoloCast. More rigorous analysis with GIN and FoldingNet will follow as future work.

ACKNOWLEDGMENT

T. Fujihashi was partly supported by JSPS KAKENHI Grant Number 20K19783.

REFERENCES

- [1] P. A. Blanche, A. Bablunian, R. Voorakaranam, C. Christenson, W. Lin, T. Gu, D. Flores, P. Wang, W. Y. Hsieh, M. Kathaperumal, B. Rachwal, O. Siddiqui, J. Thomas, R. A. Norwood, M. Yamamoto, and N. Peyghambarian, "Holographic three-dimensional telepresence using large-area photorefractive polymer," *Nature*, vol. 468, no. 7320, pp. 80–83, 2010.
- [2] H. Yu, K. Lee, J. Park, and Y. Park, "Ultrahigh-definition dynamic 3D holographic display by active control of volume speckle fields," *Nature Photonics*, vol. 11, no. 3, pp. 186–192, 2017.
- [3] M. Murphy and C. M. Sashi, "Communication, interactivity, and satisfaction in B2B relationships," *Industrial Marketing Management*, vol. 68, pp. 1–12, January 2018.
- [4] S. Nayak and R. Patgiri, "6G communication technology: A vision on intelligent healthcare," *arXiv preprint*, pp. 1–9, 2020.
- [5] A. J. Hung, J. Chen, A. Shah, and I. S. Gill, "Telementoring and telesurgery for minimally invasive procedures," *The Journal of urology*, vol. 199, no. 2, p. 355–369, February 2018.
- [6] R. Mekuria and L. Bivolarsky, "Overview of the MPEG activity on point cloud compression," in *Data Compression Conference*, 2016, p. 620.
- [7] 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.
- [8] M. Preda, "Point cloud compression in MPEG," 2017.
- [9] 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.

- [10] 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.
- [11] R. schnabel and R. Klein, "Octree-based point-cloud compression," in *Eurographics Symposium on Point-Based Graphics*, 2006, pp. 111–121.
- [12] marketrends, "Global holographic display market 2018 size, share, outlook and forecast to 2023," 2018. [Online]. Available: <https://steemit.com/electronics/@marketrends/global-holographic-display-market-2018-size-share-outlook-and-forecast-to-2023>
- [13] L. Geosystems, "BLK360 datasets," 2020. [Online]. Available: <https://shop.leica-geosystems.com/jp/blk360-dataset-downloads>
- [14] NCTech, "Arcore point cloud viewer," 2017. [Online]. Available: <https://www.youtube.com/watch?v=B-JPZ6BZpvc>
- [15] S. Pudlewski, N. Cen, Z. Guan, and T. Melodia, "Video transmission over lossy wireless networks: A cross-layer perspective," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 6–21, 2015.
- [16] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "High-quality soft video delivery with GMRF-based overhead reduction," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 473–483, Feb. 2018.
- [17] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. 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. 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.
- [20] Y. Lu, T. Fujihashi, S. Saruwatari, and T. Watanabe, "360Cast: Foveation-based wireless soft delivery for 360-degree video," in *IEEE International Conference on Communications*, 2020, pp. 1–7.
- [21] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "DNN-based overhead reduction for high-quality soft delivery," in *IEEE International Conference on Communications*, 2019, pp. 1–7.
- [22] 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.
- [23] C. Zhang, D. Florêncio, and C. Loop, "Point cloud attribute compression with graph transform," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 2066–2070.
- [24] T. Zhang and S. Mao, "Metadata reduction for soft video delivery," *IEEE Networking Letters*, vol. 1, no. 2, pp. 84–88, 2019.
- [25] C. T. Duong, T. D. Hoang, H. H. Dang, Q. V. H. Nguyen, and K. Aberer, "On node features for graph neural networks," *arXiv e-prints*, pp. 1–6, Nov. 2019.
- [26] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas, "Learning representations and generative models for 3D point clouds," *arXiv e-prints*, Jul. 2017.
- [27] Y. Yang, C. Feng, Y. Shen, and D. Tian, "Foldingnet: Point cloud auto-encoder via deep grid deformation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 206–215.
- [28] 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.
- [29] S. Chen, S. Niu, T. Lan, and B. Liu, "PCT: Large-scale 3D point cloud representations via graph inception networks with applications to autonomous driving," in *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4395–4399.
- [30] H. Gao and S. Ji, "Graph U-nets," *arXiv e-prints*, May 2019.
- [31] 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.
- [32] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch geometric," *arXiv e-prints*, Mar. 2019.
- [33] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv e-prints*, Sep. 2016.