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

HoloCast+: Hybrid Digital-Analog Transmission for Graceful Point Cloud Delivery with Graph Fourier Transform

Fujihashi, Takuya; Koike-Akino, Toshiaki; Watanabe, Takashi; Orlik, Philip V.

TR2021-043 May 11, 2021

Abstract

Point cloud is an emerging data format useful for various applications such as holographic display, autonomous vehicle, and augmented reality. Conventionally, communications of point cloud data have relied on digital compression and digital modulation for three-dimensional (3D) data streaming. However, such digital-based delivery schemes have fundamental issues called cliff and leveling effects, where the 3D reconstruction quality is a step function in terms of wireless channel quality. We propose a novel scheme of point cloud delivery, called HoloCast+, to overcome cliff and leveling effects. Specifically, our method utilizes hybrid digital-analog coding, integrating digital compression and analog coding based on graph Fourier transform (GFT), to gracefully improve 3D reconstruction quality with the improvement of channel quality. We demonstrate that HoloCast+ offers better 3D reconstruction quality in terms of the symmetric mean square error (sMSE) by up to 18.3 dB and 10.5 dB, respectively, compared to conventional digital-based and analog-based delivery methods in wireless fading environments.

IEEE Transactions on Multimedia

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

HoloCast+: Hybrid Digital-Analog Transmission for Graceful Point Cloud Delivery with Graph Fourier Transform

Takuya Fujihashi, *Member, IEEE,* Toshiaki Koike-Akino, *Senior Member, IEEE,* Takashi Watanabe, *Member, IEEE,* and Philip V. Orlik, *Senior Member, IEEE*

***Abstract*—Point cloud is an emerging data format useful for various applications such has holographic display, autonomous vehicle, and augmented reality. Conventionally, communications of point cloud data have relied on digital compression and digital modulation for three-dimensional (3D) data streaming. However, such digital-based delivery schemes have fundamental issues called cliff and leveling effects, where the 3D reconstruction quality is a step function in terms of wireless channel quality. We propose a novel scheme of point cloud delivery, called HoloCast+, to overcome cliff and leveling effects. Specifically, our method utilizes hybrid digital-analog coding, integrating digital compression and analog coding based on graph Fourier transform (GFT), to gracefully improve 3D reconstruction quality with the improvement of channel quality. We demonstrate that HoloCast+ offers better 3D reconstruction quality in terms of the symmetric mean square error (sMSE) by up to 18.3 dB and 10.5 dB, respectively, compared to conventional digital-based and analog-based delivery methods in wireless fading environments.**

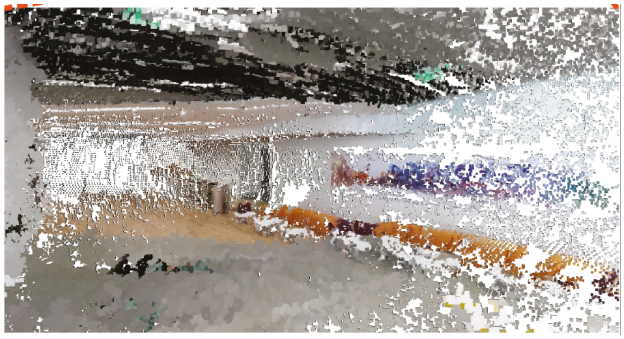
***Index Terms*—Point Cloud, Hybrid Digital-Analog Coding, Graph Signal Processing, Wireless Transmission**

I. Introduction

Holographic displays [1], [2] have emerged as attractive interface techniques for reconstructing three-dimensional (3D) scenes and objects that provide full parallax and depth information for human eyes. The 3D holographic display can be widely used for many applications: entertainment, remote device operation, medical imaging, vehicular perception, virtual/augmented reality (VR/AR), and simulated training as shown in Fig. 1. Point cloud [3] is one of data structures to represent 3D scenes and objects for such holographic displays [4]. Point cloud consists of a set of 3D points, each of which is typically associated with 3D coordinates, i.e., (X, Y, Z), and color attributes, i.e., (R, G, B) or (Y, U, V).

In contrast to conventional 2D images, 3D points in point cloud data are not regularly aligned and are non-uniformly distributed in space. One of the major issues in point cloud delivery is how to encode and send such numerous and irregular structure of 3D points while maintaining high 3D reconstruction quality on displays. Large traffic causes low reconstruction quality over limited bandwidth links, especially, in wireless links.

(a) Light detection and ranging (LIDAR) scenario [5]



(b) Virtual/augmented reality (VR/AR) scenario [6]

Fig. 1. Examples of holographic applications using point cloud.

For point cloud compression over wireless links, conventional schemes use digital-based spatial-domain encoders, such as popular Point Cloud Library (PCL) [7], [8], and graph-domain encoders. The encoders mainly consist of octree decomposition, decorrelation, quantization, and entropy coding. Specifically, a sender first decomposes 3D points into multiple 3D point sets [9] and takes quantization and entropy coding for each point set to generate the compressed bit stream for transmissions. After the compression, the transmission part sequentially uses channel coding and digital modulation schemes to reliably send the compressed bit stream to the wireless 3D display. High-quality transmissions of point clouds over wireless links can provide a good immersion for VR/AR users on mobile devices as shown in Fig. 2. Such mobile high-quality

T. Fujihashi and T. Watanabe are with Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, 565-0871 Japan. e-mail: {fujihashi.takuya, watanabe}@ist.osaka-u.ac.jp.

T. Koike-Akino and P. V. Orlik are with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. e-mail: {koike, porlik}@merl.com.

T. Fujihashi conducted this research while he was an intern at MERL.

Manuscript received September 1, 2020; revised January 29, 2021; accepted April 23, 2021.

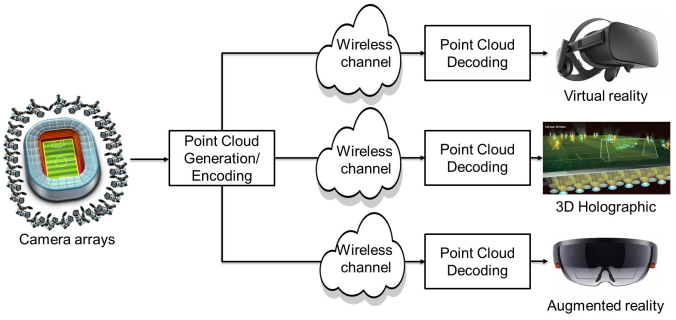


Fig. 2. Wireless point cloud delivery for immersive video applications.

VR/AR applications will bring significant benefits for post-coronavirus (COVID-19) society. For example, holographic teleconference based on the AR can realize natural and smooth online communications.

However, the existing schemes of digital-based point cloud delivery suffer from the following problems due to the wireless channel unreliability. First, the encoded bit stream is highly vulnerable for bit errors [10] occurred in wireless channels. Below a certain signal-to-noise ratio (SNR), wireless fading can cause catastrophic errors for entropy decoding of point cloud data, resulting into a significant degradation of reconstruction quality. This phenomenon is called cliff effect [11]. Second, the reconstruction quality does not improve even when the wireless channel quality is improved unless an adaptive rate control of source and channel coding is performed in a real-time manner according to the rapid fading channels. This is called leveling effect. Finally, quantization is a lossy process, and its distortion cannot be recovered at the receiver.

To overcome the above-mentioned problems, we have previously proposed HoloCast [12] for wireless point cloud delivery. HoloCast considers 3D points as the vertices in a graph with edges between nearby the vertices to deal with the irregular structure of the 3D points, motivated by graph signal processing (GSP) [13], [14]. HoloCast takes graph Fourier transform (GFT) for such graph signals to exploit the underlying correlations among the adjacent graph signals, and directly transmits linear-transformed graph signals as a quai-analog modulation over the channel. Instead of requiring for the transmitter to control the bit rate and video resolution, HoloCast enables the receiver to decode the point cloud with a bit rate and resolution commensurate with the wireless channel quality. However, in general, analog transmission schemes via linear mapping (from source signals to channel signals) are relatively inefficient. The performance of an analog transmission scheme degrades as the ratio of maximum energy to minimum energy of the source component increases, according to [15].

In this paper, we propose an extended version called HoloCast+ to achieve better 3D reconstruction quality without cliff and leveling effects. For this purpose, HoloCast+ introduces hybrid digital-analog (HDA) coding for high-quality 3D point cloud delivery. The HDA coding integrates the digital and analog point cloud compression to exploit both merits offered by analog and digital transmission schemes. Each technique plays the following roles in 3D reconstruction quality improvement.

- Digital Coding: HoloCast+ separately encodes the 3D coordinates and color components of an original point cloud and uses binary phase shift keying (BPSK) with a low-rate convolutional code for transmission. The encoder then calculates the residuals between the original and digital-encoded attributes. The residuals can reduce the ratio of maximum variance to minimum variance so that the subsequent analog coding can achieve the highest performance gain.
- Analog Coding: The residuals of both attributes are transformed into the frequency domain by using GFT to further compact the signal energy. The transformed residuals are directly mapped to Q (quadrature-phase) component to avoid interference to the digital-modulated symbols. Analog transmission ensures that the received 3D reconstruction quality will improve with the instantaneous magnitude of wireless channels.

Using public point cloud data, we verify that HoloCast+ prevents cliff and leveling effects at high SNR regimes and yields better 3D reconstruction quality irrespective of wireless channel quality. For example, the proposed method improves the reconstruction quality of the 3D coordinates by 18.3 dB and 10.5 dB on average compared with the existing octree-based digital and analog-based HoloCast schemes, respectively, across the wireless channel SNRs from -3 to 30 dB.

This paper has the following major contributions:

- We extend HoloCast with HDA delivery of 3D point cloud data.
- Our work is the first study on the integration of the graph-based analog coding for the quality enhancement of HDA point cloud delivery.
- HoloCast+ adopts blind data detection [16] to mitigate the metadata overhead for analog decoding.
- We discuss the effects of decorrelation and graph Laplacian matrix on 3D reconstruction quality. We empirically show that the random-walk GFT realizes better energy compaction compared with the other decorrelation techniques.
- We evaluate both 3D and 2D reconstruction quality of comparative schemes. HoloCast+ can project a clean image from the reconstructed 3D point cloud irrespective of the 2D projected angles.

II. Related Works

A. Digital-based Point Cloud Delivery

The point cloud representing 3D scenes or objects require numerous data traffic in general [17]. To reduce the amount of data traffic in point cloud delivery, two transform techniques were proposed for energy compaction of the non-ordered and non-uniformly distributed signals: Fourier-based transform, e.g., GFT, and wavelet-based transform, e.g., Region-Adaptive Haar Transform (RAHT) [18]. For example, recent studies took GFT for the color components [19] and 3D coordinates [20] of the graph signals for signal decorrelation. They used quantization and entropy coding for the compression of

the decorrelated signals. [21] realized graph-domain prediction before the decorrelation for further energy compaction after the compression. The wavelet-based RAHT scheme is a hierarchical transform for the color attribute of the graph signals without the need of eigenvalue decomposition [18], [22]. Based on the RAHT feature, the study in [23] realized region of interest (ROI)-based point cloud coding for the wavelet coefficients. The recent study proposed Region Adaptive GFT (RA-GFT) [24], which is a multiresolution transform by combining spatially localized block transforms, to deal with ROI-based point cloud coding even in the GFT coefficients.

B. Graceful Image/Video Delivery

Graceful image/video delivery schemes have been proposed, e.g., in [25]–[28] to grecefully improve the reconstruction quality with the improvement of instantaneous wireless channel quality. For example, SoftCast [25] was designed to realize graceful video delivery. They skip quantization and entropy coding, and use 3D discrete-cosine transform (DCT) and analog modulation, which maps DCT coefficients directly to transmission signals. FoveaCast [26] and another study [29] considered human-perception-based graceful video delivery to enhance the reconstruction quality of the ROI parts. FoveaCast considers the foveation characteristic of human vision systems into the power allocation for better quality in visual perception. The study in [29] uses you-only-look-once (YOLO) structure to find the ROI and non-ROI parts from each image to assign unequal transmission power for human-perceptual quality enhancement.

FreeCast [28] extended the graceful video delivery towards multi-view video plus depth (MVD) signals. They use 5D-DCT for decorrelation and directly send the coefficients to realize graceful quality improvement. Another study in [30] also designed SoftCast for the MVD signals. They consider the view synthesis distortion into power allocation problems to realize optimal reconstruction quality at the reference and virtual views. OmniCast [31] and 360Cast [32] were designed for graceful 360-degree video delivery. Both schemes defined the distortion model of 2D projection to realize the optimal viewport quality on each user's head mounted display. The recent study in [33] accommodated diverse users with both heterogeneous resolutions and channel conditions in graceful video delivery by using spatial decomposition for the source signals.

In view of loss resilience of graceful delivery, some studies in [34] and [35] proposed graceful video delivery systems using block-based compressed sensing. The sender randomly/adaptively samples the source signals and the receiver reconstructs the source signals from the limited number of the received samples by using the iterative thresholding algorithm. An experimental study [36] implemented the graceful video delivery onto the software radio platform and empirically demonstrated the benefits of the graceful video delivery over real-world wireless channels. We note that all of the above-mentioned studies considered that the source image/video signals are ordered and uniformly distributed signals.

Our HoloCast [12] was the first study on wireless 3D point cloud delivery for mobile holographic displays. HoloCast integrated graceful transmission with GSP to deal with irregular structure of the 3D points. HoloCast takes GFT [13], [14] for each attribute to compact the signal power, whose output is then scaled and directly mapped to transmission signals without relying on digital modulation schemes.

C. Overhead Reduction in Graceful Image/Video Delivery

In graceful image/video delivery schemes, a sender needs to let the receiver know the power information of all the transformed signals to demodulate the signals. However, it requires a relatively large amount of overhead. To reduce the amount of overhead, the existing schemes [25] used chunk division, while it can cause a degradation due to improper power allocation. To achieve better image/video quality under a low overhead requirement, a method proposed in [11] exploited a Lorentzian model to obtain the power information at the receiver with only a few parameters, while achieving an excellent streaming quality in band-limited environments. In addition, a recent study in [37] used an end-to-end deep neural network to reconstruct the power information from the limited number of the latent variables for further overhead reduction. Although the fitting-based overhead reduction realized better reconstruction quality, it still needs a large computation cost for fitting. To reduce the communication overhead without any additional computational cost at transmitter, blind data detection (BDD) [16] was proposed to decode the signal without the power information at the receiver. Specifically, they use zero-forcing estimator and sign of the received signal to approximate the original signal.

Overhead reduction of point could delivery was discussed for HoloCast in [38], where the Givens rotation was applied to efficiently compress the GFT matrix. The end-to-end deep learning approach was also investigated for point cloud delivery without overhead requirement [39], where graph neural network was introduced to compress the 3D data.

D. Hybrid Digital-Analog Coding for Image/Video Delivery

There are several HDA coding for image/video transmission schemes [40]–[45] to exploit both the benefits of conventional digital-based and graceful analog-based delivery schemes. Most typical HDA coding schemes [42], [43], [45] use digital coding for video frames and DCT-based analog coding for the residuals. They firstly assign transmission power for the digital symbols for reliable delivery and then assign the rest of the transmission power for the analog symbols to enable graceful reconstruction quality depending on the wireless channel quality. A recent study in [46] defined an adaptive recursive distortion estimate to adaptively assign the transmission power for the digital and analog symbols based on the estimated distortion. Another study [47] utilized channel state information (CSI) to decide transmission power assignment for digital and analog components across orthogonal frequency division multiplexing (OFDM) subcarriers.

Another study [48] extended the HDA delivery for the MVD signals by solving view synthesis distortion. Although the

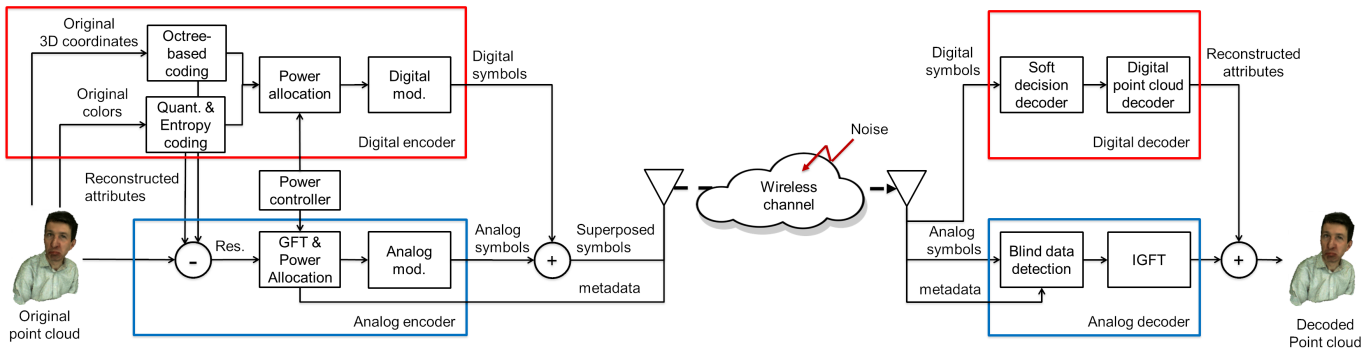


Fig. 3. End-to-end hybrid digital-analog point cloud delivery systems: HoloCast+.

above studies designed frame-level HDA delivery schemes, some studies [40], [41] designed bit-level HDA delivery schemes. Specifically, they use digital coding for lower-order bits and analog coding for higher-order bits of the pixel values to provide baseline quality through the digital symbols and quality enhancement through the analog symbols.

E. Distinguished Feature of Proposed Method

Our HoloCast+ is the extension of our HoloCast to yield better 3D reconstruction quality without cliff and leveling effects. HoloCast+ is the first HDA scheme for the graph-based point cloud delivery to deal with the irregular structure of the 3D points. The digital coding can provide the baseline quality and realize the energy compaction of the point cloud signal for analog delivery. The graph-based analog coding decorrelates the residuals in graph-domain to boost the quality enhancement according to the improvement of wireless channel quality. To reduce the communication overhead for the analog decoding, HoloCast+ uses BDD for the analog decoding. Although BDD causes quality degradation due to noise enhancement issues, we show that HoloCast+ achieves metadata-free delivery with better 3D reconstruction quality.

III. HoloCast+: Graceful Point Cloud Delivery

The objectives of our study are four-fold; 1) to prevent cliff/leveling effects in 3D scene reconstruction due to channel quality fluctuation, 2) to gracefully improve 3D reconstruction quality with the improvement of wireless channel quality, 3) to achieve highly efficient energy compaction for soft point cloud delivery, and 4) to realize lower-overhead communications.

Fig. 3 shows the end-to-end architecture of our proposed HoloCast+, where both the encoder and decoder consist of digital and analog parts. The digital encoder separately encodes 3D coordinates and the corresponding colors to generate a bit stream. The digital bits of the 3D coordinates and color components are channel-coded by a low-rate convolutional encoder and modulated by binary phase-shift-keying (BPSK) scheme for reliable transmission. The analog encoder obtains the reconstructed 3D coordinates and colors from the digital encoder and calculates the residuals between the original and reconstructed attributes. The residuals are then transformed into the graph spectrum domain by using GFT. The GFT coefficients are mapped to the quadrature (Q) component to avoid the interference to the digital-modulated symbols. The power controller assigns unequal transmission power for the digital-modulated and analog-modulated symbols, and then the superposed symbols are transmitted to the receiver. The receiver extracts the digital-modulated and analog-modulated symbols from the received symbols. The digital part for the 3D coordinates and color components is reconstructed by digital decoding, while the residual part is obtained by the analog decoding. The receiver finally adds the reconstructed residuals to the output of the digital decoder for the 3D display.

A. Digital Encoder

The 3D coordinates and color components in point cloud residing on N points are represented with $N \times 6$ vectors. We let x, y, and z denote the first three column vectors specifying the 3D coordinates, and $\boldsymbol{r}$, $\boldsymbol{g}$, and $\boldsymbol{b}$ be the rest three vectors for the color components.

In HoloCast+, the original 3D coordinates are encoded by using octree decomposition [7], [49]. The octree is a tree structure where every branch node represents a certain volume in the 3D space. When a volume containing at least one signal from the 3D point cloud, it is said to be an occupied volume. In the octree decomposition, the 3D space is hierarchically partitioned into volumes. The partition starts from the root volume which contains all the 3D points and each volume can generate eight children volumes.

We assume that a 3D point cloud is contained in a volume of $D \times D \times D$ voxels. The volume is vertically and horizontally decomposed into eight sub-volumes with the dimension of $D/2 \times D/2 \times D/2$ voxels. This process is recursively repeated for each occupied sub-volume until D will be 1. In each decomposition step, the digital encoder is verified which volumes are occupied. The occupied volumes are marked as 1 and the unoccupied volumes are marked as 0. The octets generated during this process represent an octree node occupancy state in 1-byte word. The octets in the octree are traversed in breadth-first order and the ordered octets are compressed by an arithmetic coder considering the correlation with neighboring octets.

Here, the compression level of the 3D coordinates depends on the number of the decomposition steps. If the recursive decomposition provides all the end sub-volumes containing

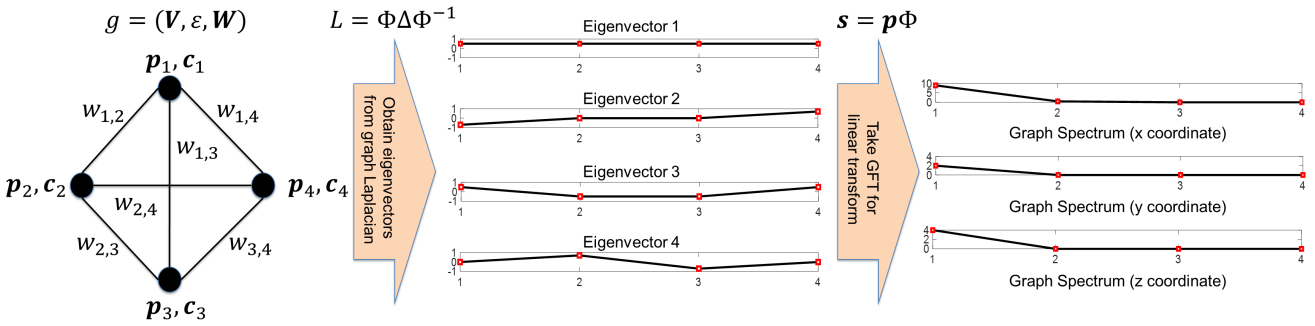


Fig. 4. Graph Fourier transform (GFT) to convert point cloud attributes into the graph spectrum density via eigen-space of the graph Laplacian matrix.

only one 3D point, the compression is lossless. Unless otherwise (containing multiple 3D points in some sub-volumes), the compression is lossy. In this case, we regard the 3D coordinates of the points in the sub-volume as the center of the sub-volume. In this paper, we adjust the available number of 3D points in each volume to control the compression level of the 3D coordinates.

On the other hand, HoloCast+ encodes the color components by using the digital-based operations of quantization and entropy coding. We let f_{color} denote the column vector containing the each color component. Each vector of the color components is uniformly quantized as $f'_{\text{color}} = \text{round}(f_{\text{color}}/Q_{\text{color}})$, where Q_{color} is the quantization step size for the color components. The quantized vector f'_{color} is then coded by an arithmetic coding to generate digital bits.

B. Analog Encoder

The digital encoder is typically lossy due to finite quntization and incomplete octree decomposition. In addition, channel coding and digital modulation are difficult to be adaptive in a real-time fashion depending on the rapid change of wireless fading channels. Therefore, the digital part alone suffers from the cliff/leveling effects. To realize the graceful performance, HoloCast+ uses the analog encoder for the residual signals of the analog part, i.e., the residual 3D coordinates $\boldsymbol{p} = [\mathsf{x}'', \mathsf{y}'', \mathsf{z}'']^{\mathrm{T}} = [\mathsf{x}-\mathsf{x}', \mathsf{y}-\mathsf{y}', \mathsf{z}-\mathsf{z}']^{\mathrm{T}} \in \mathbb{R}^{3\times N}$ and the residual color components $\boldsymbol{c} = [\boldsymbol{r}'', \boldsymbol{g}'', \boldsymbol{b}'']^{\mathrm{T}} = [\boldsymbol{r} - \boldsymbol{r}', \boldsymbol{g} - \boldsymbol{g}', \boldsymbol{b} - \boldsymbol{b}']^{\mathrm{T}} \in \mathbb{R}^{3\times N}$.

1) Graph Construction: Given the residual of the digital compression, HoloCast+ uses a weighted and undirected graph $\mathcal{G} = (\boldsymbol{V}, \mathcal{E}, \boldsymbol{W})$ where $\boldsymbol{V}$ and $\mathcal{E}$ are the vertex and edge sets of $\mathcal{G}$, respectively. $\boldsymbol{W}$ is an adjacency matrix having positive edge weights and the (i,j)th entry $\boldsymbol{W}_{i,j}$ represents the weight of an edge connecting vertices i and j. We consider the attributes of the point cloud, i.e., the residuals of the 3D coordinates $\boldsymbol{p}$ and the residuals of the color components $\boldsymbol{c}$, as signals that reside on the vertices in the graph.

2) Graph Fourier Transform for Residuals: Fig. 4 illustrates how to transform a graph signal into the graph spectrum domain by using GFT. From the attributes of the graph signal, each weight $\boldsymbol{W}_{i,j}$ can be calculated, e.g., by the bilateral Gaussian kernel [50] as follows:

$$\boldsymbol{W}_{i,j} = \exp\left(-\left(\frac{\|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2^2}{\kappa_p} + \frac{\|\boldsymbol{c}_i - \boldsymbol{c}_j\|_2^2}{\kappa_c}\right)\right), \quad (1)$$

where κ_p and κ_c are hyperparameters specifying the kernel width for 3D coordinates and color components, respectively. In HoloCast+, we use the standard deviation across the corresponding attributes for the hyperparameters κ_p and κ_c. A sender then transforms the residuals into spectral representation using GFT. The GFT is defined through the graph Laplacian operator $\boldsymbol{L}$ using edge weight matrix $\boldsymbol{W}$ and degree matrix $\boldsymbol{D}$, where $\boldsymbol{D}$ is a diagonal matrix whose ith diagonal element is equal to the sum of the weights of all the edges incident to the ith vertex. Specifically, the diagonal matrix is represented as follows:

$$\boldsymbol{D}_{i,j} = \begin{cases} \sum_{n=1}^{N} \boldsymbol{W}_{i,n}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Based on the degree matrix, we can calculate some variants of the graph Laplacian matrix [51]:

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}, \quad (3)$$

$$\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{W}, \quad (4)$$

where $\boldsymbol{I}$ denotes an identity matrix of proper dimension. We refer to the above graph Laplacian matrices as regular and random-walk graph Laplacian, respectively.

In general, the graph Laplacian is a real symmetric matrix that has a complete set of orthonormal eigenvectors with corresponding nonnegative eigenvalues. To obtain the eigenvectors and eigenvalues, the eigenvalue decomposition of the graph Laplacian matrix is performed as:

$$\boldsymbol{L} = \boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{\Phi}^{-1}, \quad (5)$$

where $\boldsymbol{\Phi}$ is the eigenvectors matrix and $\boldsymbol{\Delta}$ is a diagonal matrix containing the eigenvalues.[1] The multiplicity of the smaller eigenvalue indicates the number of connected components of the graph. The GFT coefficients of each attribute are obtained

[1]For non-diagonalizable graph Laplacian matrix, the singular value decomposition (SVD) is instead used to express as $\boldsymbol{L} = \boldsymbol{\Psi}\boldsymbol{\Delta}\boldsymbol{\Phi}^{-1}$ where $\boldsymbol{\Psi}$, $\boldsymbol{\Delta}$ and $\boldsymbol{\Phi}$ denote left singular vectors matrix, diagonal matrix containing singular values, and right singular vectors matrix, respectively. In this case, we use the right singular vectors of $\boldsymbol{\Phi}$ as the graph-based transform basis matrix $\boldsymbol{\Phi}$.

by multiplying the graph-based transform basis matrix by the corresponding residual vector $\boldsymbol{e} \in \mathbb{R}^N$ as follows:

$$\boldsymbol{s} = \boldsymbol{e}\,\boldsymbol{\Phi}, \tag{6}$$

where $\boldsymbol{s}$ is a vector of GFT coefficients corresponding to the residuals of $\boldsymbol{e}$.

C. Power Allocation

In HoloCast+, the power controller decides transmission powers for digital and analog encoders based on the wireless channel quality. The transmitter first decides power allocation for digital encoder to ensure enough power to decode the entropy-coded bit stream correctly. When the channel quality is low, the receiver will face difficulty in decoding the bit stream correctly. For that case, our scheme switches to pure analog transmission to prevent the cliff effect. To decide the transmission power for the digital encoder, the power controller calculates the power threshold to decode the bit stream correctly:

$$P_{\mathsf{th}} = N_0 \cdot \gamma_0, \tag{7}$$

where P_{th} is the power threshold and N_0 is the average noise power of the wireless channel. Here, γ_0 is the required SNR to guarantee that the decoding bit-error rate (BER) is not larger than a target BER. After the threshold calculation, the transmitter decides the transmission powers for digital encoder P_{d} and analog encoder P_{a}, respectively, as follows:

$$P_{\mathsf{d}} = \begin{cases} P_{\mathsf{th}}, & P_{\mathsf{th}} \le P_{\mathsf{t}}, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

$$P_{\mathsf{a}} = P_{\mathsf{t}} - P_{\mathsf{d}}, \tag{9}$$

where P_{t} is the total power budget.

Let x_i denote the ith transmission symbol. The symbol x_i is formed by superposing a BPSK-modulated symbol $x_i^{\langle \mathsf{d} \rangle}$ and analog-modulated symbol $x_i^{\langle \mathsf{a} \rangle}$ as follows:

$$x_i = x_i^{\langle \mathsf{d} \rangle} + \jmath\, x_i^{\langle \mathsf{a} \rangle}, \tag{10}$$

where $\jmath = \sqrt{-1}$ denotes the imaginary unit. The BPSK-modulated symbol and the analog-modulated symbol are scaled by P_{d} and P_{a}, respectively, as follows:

$$x_i^{\langle \mathsf{d} \rangle} = \sqrt{P_{\mathsf{d}}} \cdot b_i, \qquad x_i^{\langle \mathsf{a} \rangle} = g_i \cdot s_i, \tag{11}$$

where $b_i \in \mathbb{X} = \{\pm 1\}$ is the BPSK-modulated symbol and $s_i \in \mathbf{s}_i$ is the ith GFT coefficient. Here, g_i is a scale factor for ith GFT coefficient. The optimal scale factor g_i is obtained by minimizing the mean square error (MSE) under the power constraint for analog encoder of P_{a} as follows:

$$\min_{\{g_i\}} \quad \mathsf{MSE} = \mathbb{E}\left[(s_i - \hat{s}_i)^2\right] = \frac{1}{N}\sum_i^N \frac{\sigma^2 \lambda_i}{g_i^2 \lambda_i + \sigma^2}, \tag{12}$$

$$\text{s.t.} \quad \frac{1}{N}\sum_i^N g_i^2 \lambda_i = P_{\mathsf{a}}, \tag{13}$$

where $\mathbb{E}[\cdot]$ denotes expectation, $\hat{s}_i$ is a receiver estimate of the transmitted GFT coefficient, $\lambda_i = |s_i|^2$ is the power of the ith GFT coefficient, N is the total number of coefficients, and σ^2 is a receiver noise variance. As shown in [25], the near-optimal solution is expressed as

$$g_i = m\lambda_i^{-1/4}, \qquad m = \sqrt{\frac{NP_{\mathsf{a}}}{\sum_j \lambda_j^{1/2}}}. \tag{14}$$

Over the wireless links, the receiver obtains the received BPSK-modulated and analog-modulated symbols, which are modeled as follows:

$$y_i = x_i + n_i, \tag{15}$$

where y_i is the ith received symbol and n_i is an effective AWGN with a variance of σ^2. We assume an effective fading attenuation is considered in the noise variance.

D. Decoder

1) Digital Part: The receiver first extracts BPSK-modulated symbol from the in-phase (I) component of each received symbol, i.e., $\Re(y_i)$. To decode the modulated symbol, the digital decoder calculates log-likelihood ratio (LLR) values from the received symbols:

$$L_i = \ln \frac{\Pr(y_i|1)}{\Pr(y_i|0)}, \tag{16}$$

where L_i is the LLR value of the received symbol. Here, $\Pr(y_i|\omega)$ denotes the probability that the received signal is y_i conditioned on the transmitted bits of ω, i.e., $\Pr(y_i|\omega) = \frac{1}{\pi\sigma_i^2}\exp\left(-\frac{1}{\sigma_i^2}(\Re(y_i) - M(\omega))^2\right)$ where $M(\omega) \in (-1)^\omega\sqrt{P_{\mathsf{d}}}$ is the BPSK modulated symbol for ω. After computing the LLR values for all received symbols, the receiver deinterleaves the LLR values and feeds them into the Viterbi decoder. The output of the Viterbi decoder is the entropy-coded bit stream, which will be further decoded by an arithmetic decoder to reconstruct the 3D coordinates and color components.

2) Analog Part: The receiver extracts residual values from the Q component of each received symbol, i.e., $\Im(y_i)$. The receiver first uses the MMSE filter [25] for the extracted value:

$$\hat{s}_i = \frac{g_i \lambda_i}{g_i^2 \lambda_i + \sigma^2} \cdot \Im(y_i). \tag{17}$$

The analog decoder then reconstructs corresponding residuals $\hat{\boldsymbol{e}}$ by taking the inverse GFT (IGFT) for the filtered GFT coefficients in each attribute $\hat{\boldsymbol{s}}$ as follows:

$$\hat{\boldsymbol{e}} = \hat{\boldsymbol{s}}\,\boldsymbol{\Phi}^{-1}. \tag{18}$$

By adding the decoded residuals $\hat{\boldsymbol{e}}$ into the reconstructed 3D point clouds based on the digital part, we can achieve graceful quality.

E. Overhead Reduction for Analog Decoding

To carry out the MMSE filtering in (17), the sender needs to correctly notify the receiver the value of λ_i of all the GFT coefficients as the metadata. For example, to transmit a point cloud with $N = 800{,}000$ points, there will be $6N = 4{,}800{,}000$ coefficients. This overhead will cause performance degradation and consume extra transmission power. To reduce a large overhead, the conventional graceful delivery scheme [25] divides the coefficients into multiple chunks and carry out

chunk-wise scaling and MMSE filter. However, the overhead is still high in general and the chunk division can cause another factor of performance degradation due to a loss of optimality for scaling with respect to (17).

Although the existing metadata reduction schemes [11], [37] based on a signal model can realize the approximation of the metadata power, they need a large computational cost and prior knowledge. HoloCast+ instead uses the BDD [16] to decode the residual without large computational cost and the prior knowledge. Specifically, HoloCast+ first scales each GFT coefficient with an optimal scaling factor at the analog encoder. With $\lambda_i = |s_i|^2$, (14) can be rewritten as:

$$g_i = m|s_i|^{-1/2}. \tag{19}$$

The received signal of the analog part can be then modeled:

$$\Im(y_i) = g_i \cdot s_i + n_i = m|s_i|^{-1/2}s_i + n_i. \tag{20}$$

Here, we can estimate the amplitude of s_i via a zero-forcing estimator:

$$|\hat{s}_i| = (\Im(y_i)/m)^2, \tag{21}$$

and use the sign of the received symbol to predict the sign of s_i, i.e., $\mathrm{sign}(\hat{s}_i) = \mathrm{sign}(\Im(y_i))$. Accordingly, we obtain an estimate of s_i as follows:

$$\hat{s}_i = |\hat{s}_i| \cdot \mathrm{sign}(\hat{s}_i) = (\Im(y_i)/c)^2 \cdot \mathrm{sign}(\Im(y_i)). \tag{22}$$

The above equation shows that the amplitude of the GFT coefficients is proportional to the squared amplitude of the received analog-modulated symbol. For decoding all the GFT coefficients, the analog decoder only needs to know the value of the constant m. As a result, BDD in the proposed HoloCast+ realizes almost metadata-free, i.e., single metadata transmission of m. Nevertheless, the zero-forcing estimation used in the BDD can generally cause a noise enhancement issue, which may degrade the reconstruction quality.

IV. Performance Evaluation

A. Simulation Settings

Performance Metric in 3D Point Cloud: We evaluate the 3D reconstruction quality of point cloud delivery in terms of the symmetric MSE based on [52] in each attribute of 3D coordinates $\boldsymbol{p}$ and color components $\boldsymbol{c}$. The symmetric MSE of the 3D coordinates, $\mathsf{sMSE}_{\mathsf{xyz}}$, can be obtained as follows:

$$\mathsf{sMSE}_{\mathsf{xyz}} = \frac{1}{2}\big(d(\boldsymbol{p}_{\mathrm{org}} \to \boldsymbol{p}_{\mathrm{dec}}) + d(\boldsymbol{p}_{\mathrm{dec}} \to \boldsymbol{p}_{\mathrm{org}})\big), \tag{23}$$

where $\boldsymbol{p}_{\mathrm{org}}$ is the original 3D coordinates and $\boldsymbol{p}_{\mathrm{dec}}$ is the decoded 3D coordinates. Here, each way of the asymmetric MSE in the 3D coordinates are defined as follows:

$$d(\boldsymbol{p}_{\mathrm{org}} \to \boldsymbol{p}_{\mathrm{dec}}) = \frac{1}{N}\sum_{\boldsymbol{p}\in\boldsymbol{p}_{\mathrm{org}}}\left(\min_{\boldsymbol{p}'\in\boldsymbol{p}_{\mathrm{dec}}}\left\|\boldsymbol{p}-\boldsymbol{p}'\right\|_2^2\right),$$

$$d(\boldsymbol{p}_{\mathrm{dec}} \to \boldsymbol{p}_{\mathrm{org}}) = \frac{1}{N}\sum_{\boldsymbol{p}\in\boldsymbol{p}_{\mathrm{dec}}}\left(\min_{\boldsymbol{p}'\in\boldsymbol{p}_{\mathrm{org}}}\left\|\boldsymbol{p}-\boldsymbol{p}'\right\|_2^2\right).$$

Note that the sMSE is closely related to the augmented Chamfer distance.

The symmetric peak SNR (sPSNR) of the color components is derived analogously as follows:

$$\mathsf{sPSNR} = \frac{255^2}{\frac{1}{2}\big(d(\boldsymbol{c}_{\mathrm{org}} \to \boldsymbol{c}_{\mathrm{dec}}) + d(\boldsymbol{c}_{\mathrm{dec}} \to \boldsymbol{c}_{\mathrm{org}})\big)}, \tag{24}$$

where $\boldsymbol{c}_{\mathrm{org}}$ and $\boldsymbol{c}_{\mathrm{dec}}$ are the original and decoded color components, respectively. The distance of the color component is defined as follows:

$$d(\boldsymbol{c}_{\mathrm{org}} \to \boldsymbol{c}_{\mathrm{dec}}) = \frac{1}{N}\sum_{\boldsymbol{c}\in\boldsymbol{c}_{\mathrm{org}}}\left(\left\|\boldsymbol{c}-\boldsymbol{c}_{\mathrm{dec}}(\boldsymbol{p}'_{\min})\right\|_2^2\right),$$
$$\boldsymbol{p}'_{\min} = \arg\min_{\boldsymbol{p}'\in\boldsymbol{p}_{\mathrm{dec}}}\left\|\boldsymbol{p}_{\mathrm{org}}-\boldsymbol{p}'\right\|_2^2,$$

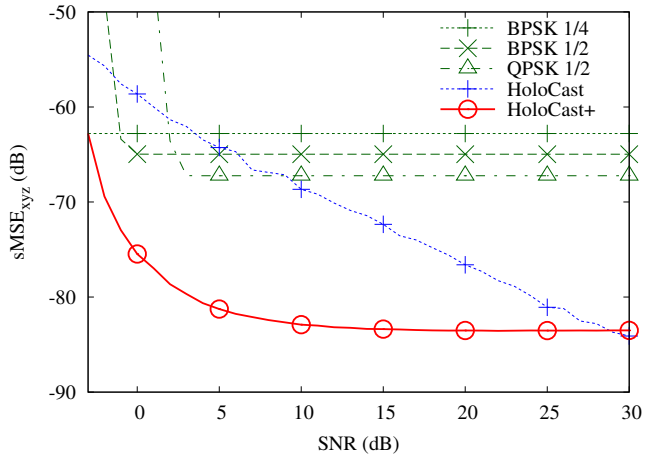
$$d(\boldsymbol{c}_{\mathrm{dec}} \to \boldsymbol{c}_{\mathrm{org}}) = \frac{1}{N}\sum_{\boldsymbol{c}\in\boldsymbol{c}_{\mathrm{dec}}}\left(\left\|\boldsymbol{c}-\boldsymbol{c}_{\mathrm{org}}(\boldsymbol{p}''_{\min})\right\|_2^2\right),$$
$$\boldsymbol{p}''_{\min} = \arg\min_{\boldsymbol{p}''\in\boldsymbol{p}_{\mathrm{org}}}\left\|\boldsymbol{p}_{\mathrm{dec}}-\boldsymbol{p}''\right\|_2^2,$$

where $\boldsymbol{c}_{\mathrm{org}}(\boldsymbol{p})$ and $\boldsymbol{c}_{\mathrm{dec}}(\boldsymbol{p})$ represent the original and decoded color components of the corresponding 3D coordinates $\boldsymbol{p}$, respectively.

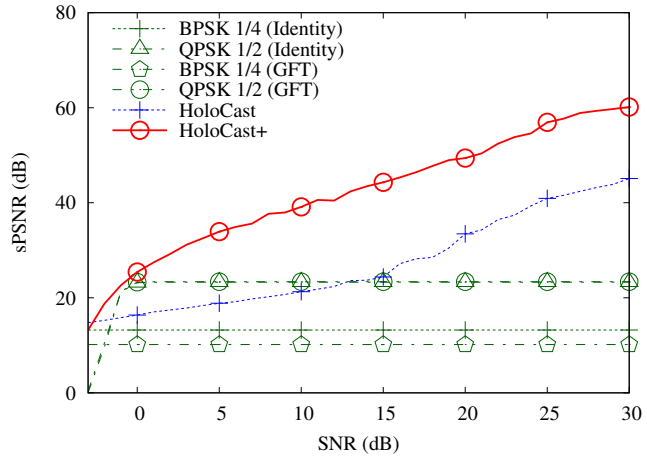
Point Cloud Dataset: We use publicly available point cloud data, namely, *pencil__10_0*, *pencil__9_0*, *pencil__4_0*, *pen__4_0*, and *milk_color* whose number of points N is 2,731, 6,712, 5,712, 23,649, and 13,704, respectively. To deal with a large number of 3D points in both digital and analog coding, we discretize the 3D points into multiple voxels using the octree decomposition and take a decorrelation method for each voxel. We consider each voxel contains up to 6,000 3D points.

Wireless Settings: The received symbols are impaired by an AWGN channel. For digital-based delivery schemes, we use a rate-$1/2$ or $1/4$ convolutional code with a constraint length of 10 for the compressed bit stream. We use the digital modulation formats of BPSK and Quadrature Phase-Shift Keying (QPSK) to send the channel-coded symbols. For the proposed HoloCast+ scheme, we use a rate-$1/4$ convolutional code with a constraint length of 10 and BPSK modulation format for the digitally-coded bit stream. The BPSK-modulated symbols are superposed with the analog-modulated symbols for HDA delivery. Here, we set γ_0 to -3 dB to prevent bit errors in the digital part from preliminary evaluation results. We consider an instantaneous N_0 can be precisely estimated by the sender unless otherwise stated. The effect of the estimation error between the estimated and instantaneous N_0 will be discussed later.

Comparative Schemes: We compare the proposed HoloCast+ with the conventional digital or analog point cloud delivery schemes. For the digital-based delivery schemes, the octree-based compression is used for 3D coordinates compression [49]. For the color components, the sender uses the quantization and entropy coding for the compression. To discuss the impact of the signal decorrelation in the digital compression, we consider either no decorrelation or GFT-based decorrelation [19] for the color components before the quantization. Here, the random-walk graph Laplacian matrix is used for the GFT decorrelation, given the original 3D coordinates. As a baseline of the conventional analog methods,



(a) 3D coordinates $\boldsymbol{p}$



(b) Color components $\boldsymbol{c}$

Fig. 5. Average reconstruction quality of 3D coordinates and color attributes in digital-based delivery, HoloCast, and HoloCast+ schemes for *pencil__10_0*, *pencil__9_0*, *pencil__4_0*, *pen__4_0*, and *milk_color*.

we consider HoloCast [12], which takes GFT for the 3D coordinates and color components based on the random-walk graph Laplacian matrix. The GFT coefficients are scaled and analog-modulated before transmission.

B. HoloCast+ vs. Conventional Schemes

We first compare the 3D reconstruction quality of HoloCast+ with the conventional digital-based delivery and analog-based HoloCast schemes. Here, the proposed HoloCast+ uses the random-walk graph Laplacian $\boldsymbol{L}$. Fig. 5 (a) shows the average symmetric MSE of the 3D coordinates for the digital-based delivery, HoloCast, and HoloCast+ schemes as a function of wireless channel SNRs. Here, we set the available number of transmission symbols for the 3D coordinates of *pencil__10_0*, *pencil__9_0*, and *pencil__4_0* to 4.1 ksymbols, whereas the available number of transmission symbols for the 3D coordinates of the other cases is 23.0 ksymbols. From Fig. 5 (a), we can find the following observations:

- HoloCast+ and HoloCast gracefully improve the reconstruction quality of 3D coordinate attributes with the improvement of wireless channel quality.
- HoloCast+ achieves the best MSE performance irrespective of wireless channel quality.
- The digital-based delivery schemes suffer from cliff effect at low channel SNR regimes because bit errors cause synthesis failure of entropy decoding.

For example, HoloCast+ achieves 20.1 dB, 17.9 dB, and 14.2 dB improvement compared with the rate-$1/4$ BPSK, rate-$1/2$ BPSK, and HoloCast schemes, respectively, at a wireless channel SNR of 10 dB.

Fig. 5 (b) shows the average symmetric PSNR performance of the color components as a function of wireless channel SNRs. Here, we set the available number of transmission symbols for the color components of *pencil__10_0* to 43.0 ksymbols, *pencil__9_0* and *pencil__4_0* to 70.0 ksymbols, and the other cases to 230.0 ksymbols. In this case, the total number of transmission symbols for the 3D coordinates is identical to that in Fig. 5 (a). It is confirmed that HoloCast+ realizes graceful quality improvement even for the color components. In addition, HoloCast+ scheme also offers better reconstruction quality compared with the digital-based delivery and analog-based HoloCast schemes. This is because the digital part of HoloCast+ compacts the power of the 3D coordinates and color components to boost the quality enhancement of the analog coding.

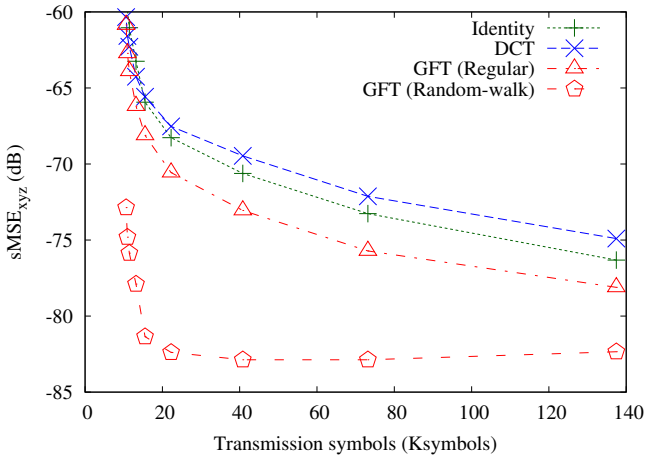
C. Impact of Decorrelation and Laplacian Matrix

In the previous section, we demonstrated that HoloCast+ yields better 3D reconstruction quality compared with the existing digital-based and analog-based schemes. In Figs. 6 (a) and (b), we discuss the performance of the proposed HoloCast+ in more details. Specifically, we consider HoloCast+ schemes with different decorrelation and graph Laplacian matrices for the residuals to clarify an impact of the GFT on quality improvement.

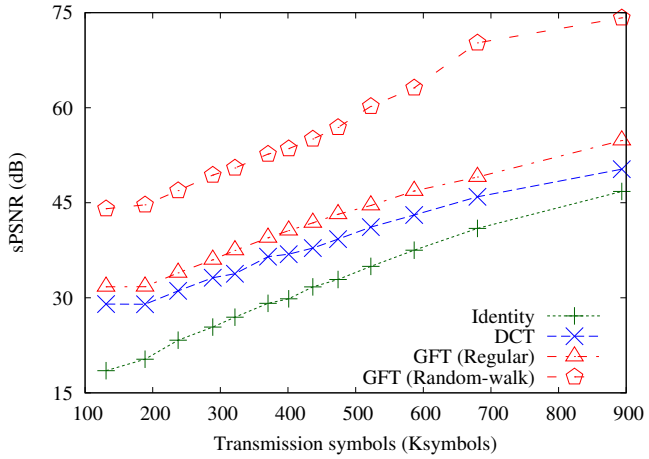
From the results, one can see that GFT-based HoloCast+ achieves better reconstruction quality compared with the DCT-based HoloCast+ and HoloCast+ without decorrelation. For example, the random-walk GFT-based HoloCast+ improves the synmetric MSE performance of the 3D coordinates by 12.8 dB and 13.0 dB compared with the DCT-based HoloCast+ and HoloCast+ w/o decorrelation, respectively, on average across the number of the transmission symbols between 10.6 ksymbols and 137.4 ksymbols. In addition, we can see that the random-walk Laplacian matrix was best in both 3D coordinate and color component attributes. When we use the regular graph Laplacian matrix for the analog coding, the reconstruction quality of the 3D coordinates and color components causes up to 10.4 dB and 13.9 dB degradation, respectively, compared with the random-walk graph Laplacian matrix.

D. Effect of Noise Power Accuracy

In our evaluation, we assumed that the sender has a perfect knowledge of an instantaneous N_0, i.e., SNR, to calculate

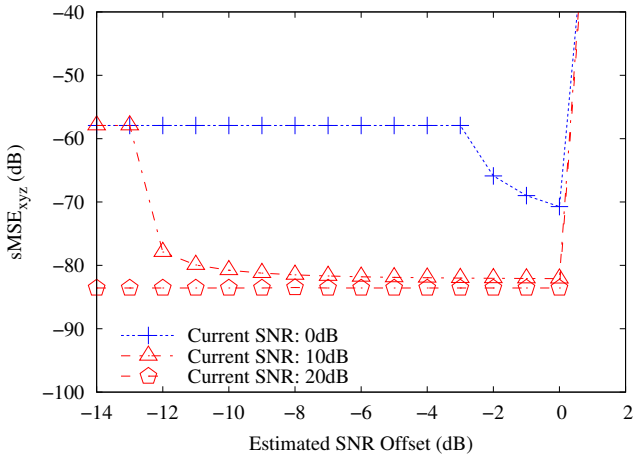


(a) 3D coordinates $\boldsymbol{p}$

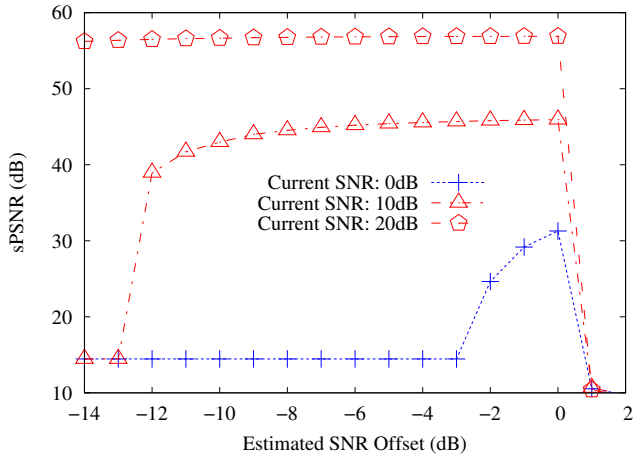


(b) Color components $\boldsymbol{c}$

Fig. 6. Reconstruction quality of 3D coordinates and color attributes in HoloCast+ scheme as a function of the number of the transmission symbols for the point cloud of *milk_color*.



(a) 3D coordinate



(b) Color components

Fig. 7. Effect of the offset of the estimated SNR against the true SNR on the 3D reconstruction quality for the point cloud of *milk_color*.

the power threshold P_{th} for transmission power determination in both digital and analog parts. If the gap between the estimated and instantaneous SNR is large, it will affect the 3D reconstruction quality of the proposed HoloCast+ scheme.

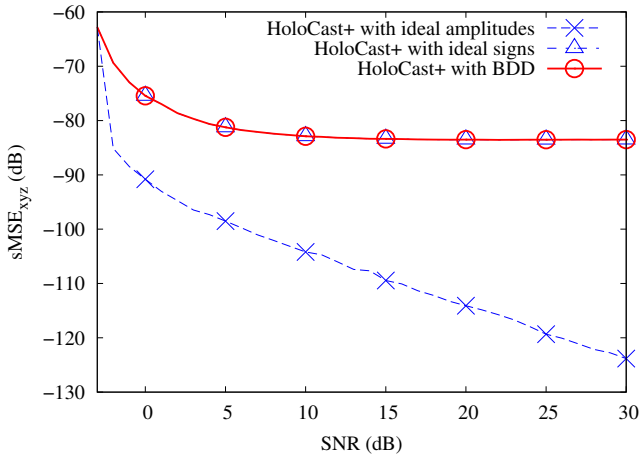
Figs. 7 (a) and (b) show the symmetric MSE and PSNR performance for the point cloud of *milk_color* as a function of the offset of the estimated SNR against the true SNR. Here, we consider three cases of the true SNR, i.e., 0, 10, and 20 dB, to discuss the effect of the gap in low, middle, and high average wireless channel SNRs. We can find the following results:

- If the estimated SNR is lower than the true SNR, the proposed HoloCast+ assigns unnecessarily larger transmission power to the digital part while decreasing the analog power. Hence, the 3D reconstruction quality will converge to that of the digital-only scheme if the estimation error is large.
- The estimation error margin is wider for higher SNRs.
- When the transmitter over-estimated the channel SNR, the proposed HoloCast+ allocates much transmission power to the analog part. In this case, cliff effects occur even in the proposed HoloCast+ due to an insufficient power allocation for the digital part.

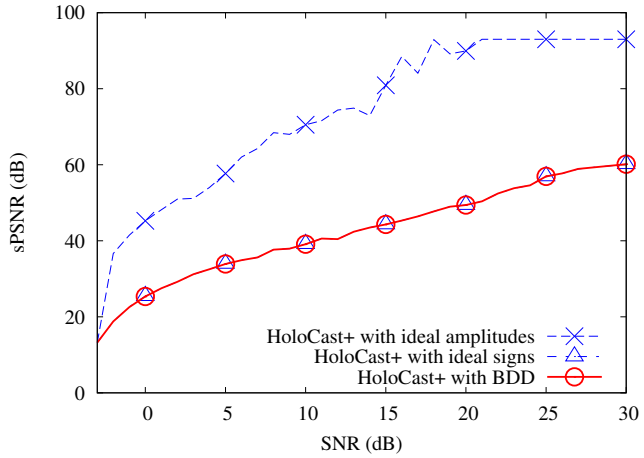
It confirms that an appropriate power allocation is necessary for HDA framework. Nevertheless, having a sufficient margin for γ_0, the cliff effect can be prevented in practice.

E. Effect of Estimation Accuracy in Blind Data Detection

The BDD used in HoloCast+ realizes the reconstruction without large computational cost and the prior knowledge. However, the estimated amplitudes and signs may cause quality degradation due to the estimation error. To discuss an impact of the estimation error on the 3D reconstruction quality, we evaluate three cases: HoloCast+ with BDD, HoloCast+ with ideal amplitudes, and HoloCast+ with ideal signs. For HoloCast+ with ideal amplitudes, the receiver is assumed to have the prior knowledge of $|s_i|$. In this case, the receiver estimates the sign of s_i from the received symbol for decoding the residual signals. We note that this scheme needs to send all the amplitude information of the residuals without errors, and thus it causes a significant communication overhead. For HoloCast+ with ideal signs, the receiver has the prior



(a) 3D coordinate



(b) Color components

Fig. 8. Performance of BDD with/without amplitude and sign information.

knowledge of the sign of s_i, and thus estimates the amplitude of s_i via the zero-forcing estimator. Although sign information need to be sent as metadata, the overhead may not be large as it is just a binary data.

Figs. 8 (a) and (b) show the average symmetric MSE and PSNR of the 3D coordinate and color component attributes as a function of the wireless channel SNRs. We can see the following results:

- HoloCast+ with BDD and HoloCast+ with ideal signs achieve almost the same 3D reconstruction quality.
- The knowledge of amplitude information improves the performance of HoloCast+ significantly at the cost of traffic overhead.

For example, the performance improvement is about 15.3 dB at a channel SNR of 0 dB, while the gap increases to 40.3 dB at a channel SNR of 30 dB. We leave how to realize an accurate prediction of the amplitude information without the need of large overheads as a future work.

F. 2D Projected Point Cloud Quality

Finally, we evaluate the reconstruction performance in terms of visual quality, PSNR, and structural similarity (SSIM) [53]

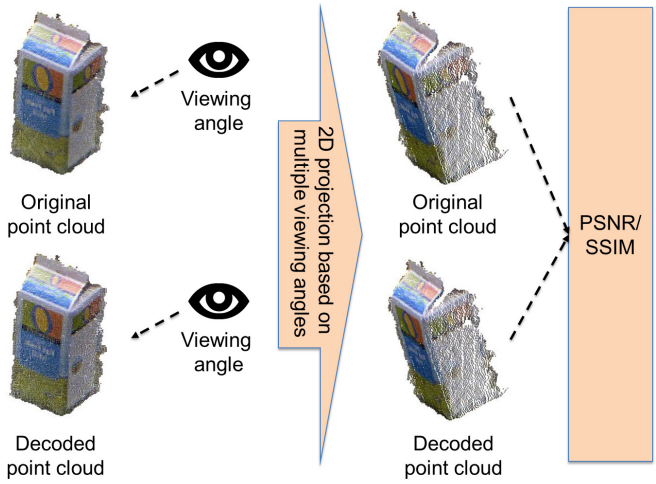


Fig. 9. Quality measurement of point cloud delivery from 2D projected images.

of point cloud data projected on 2D image from a particular angle. Fig. 9 shows how to measure the reconstruction quality of the 2D projected images in our evaluation. Depending on the perspective of the user, 3D point clouds are projected on the 2D view plane of the user. We then measure the performance of PSNR and SSIM by using the original and decoded 2D images. The 2D projected metrics may be more relevant to discuss the perceptual distortion for practical holographic display systems, compared to 3D symmetric MSE.

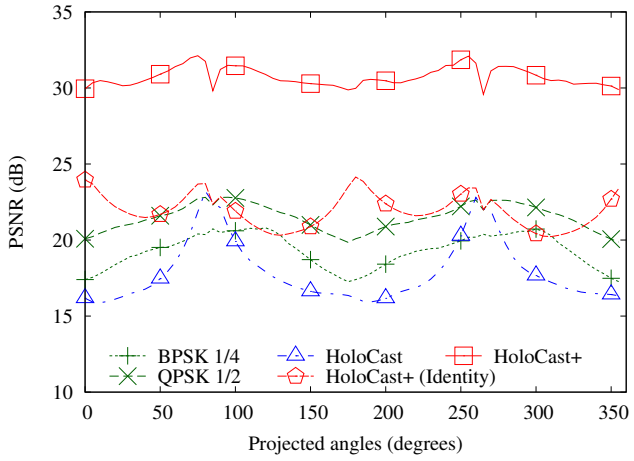
PSNR is defined as follows:

$$\mathsf{PSNR} = 10 \log_{10} \frac{(2^L - 1)^2}{\epsilon_{\mathsf{MSE}}}, \tag{25}$$

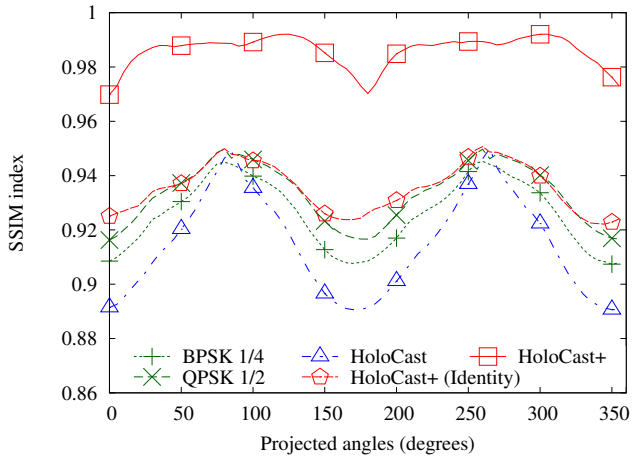
where L is the number of bits used to encode pixel luminance (typically eight bits), and $\varepsilon_{\mathrm{MSE}}$ is the MSE between all pixels of the decoded and the original 2D projected images. SSIM is known as a better metric than PSNR to predict the perceptual quality between the original and decoded point cloud images. Larger values of SSIM close to 1 indicates higher perceptual similarity between original and decoded images.

Figs. 10 (a) and (b) show the PSNR and SSIM performance as a function of the 2D projected angles for the reference point cloud of *milk_color* at a wireless channel SNR of 10 dB. Here, we also set the available number of transmission symbols for the 3D coordinates and color components to 23.0 ksymbols and 230.0 ksymbols, respectively. The original and decoded point clouds are projected onto a 2D plane at different angles horizontally rotated by 5 degree steps. In Figs. 10 (a) and (b), it observed that the HoloCast+ achieves the highest quality irrespective of the 2D projected angles. Although HoloCast+ without decorrelation is still better than the conventional digital-based delivery and analog-based HoloCast schemes, the reconstruction quality considerably varies over the 2D projection angle. The stable gain of the GFT-based HoloCast+ is especially beneficial for multi-user holographic displays which require good quality from any possible directions.

Figs. 11 (a)–(f) show snapshots of 2D projected images to discuss the visual quality of the comparative schemes for



(a) PSNR



(b) SSIM index

Fig. 10. Reconstruction quality of 2D projected images for point cloud of *milk_color*.

the reference point cloud of *milk_color* at a wireless channel SNR of 10 dB. The available number of transmission symbols is the same as Fig. 10. We can see that HoloCast+ exhibits better visual quality compared with the other schemes in terms of the 3D coordinates and color components. Specifically, HoloCast+ can reproduce a clean 3D scene with fine details.

V. Conclusion and Discussion

In this paper, we proposed a novel HDA approach called HoloCast+ to realize graceful point cloud delivery over wireless links/networks. Specifically, HoloCast+ integrates digital coding and graph-based analog coding to achieve high-quality delivery of non-ordered and non-uniformly distributed 3D point clouds. We confirmed that HoloCast+ achieves better 3D and 2D reconstruction quality with the improvement of the instantaneous wireless channel quality. In addition, we demonstrated that random-walk graph Laplacian matrix can boost the quality enhancement compared with the other decorrelation methods.

The proposed HoloCast+ still has three-fold drawbacks to tackle in the future work. The first drawback is a quality degradation due to the estimation error caused by the BDD operation as discussed in Section IV-E. We found that more precise prediction of the amplitude information with limited prior knowledge is required to further improve the 3D reconstruction quality.

The second drawback is a communication overhead due to the graph-based transform basis matrix for the analog part. Even though the BDD can remove the metadata overhead of amplitude information, the graph-based delivery schemes still need to send the GFT matrix as additional metadata, which will cause rate and power losses for the analog-modulated symbols. This issue may be partly resolved by integrating some overhead reduction techniques, e.g., the Givens rotation [38] or graph neural network [39], into HDA framework.

The third drawback is a challenge to integrate with the standardized digital-based point cloud coding (PCC), i.e., geometry-based PCC and video-based PCC. As this paper focused on the proof-of-principle study to show the feasibility of the GFT-based HDA scheme for the wireless point cloud delivery, we will discuss the integration of the MPEG-Immersive standard activities as a future work.

Acknowledgment

T. Fujihashi's work was partly supported by JSPS KAKENHI Grant Number JP20K19783.

References

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

[4] 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.

[5] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuca, S. Lassere, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging MPEG standards for point cloud compression," *IEEE Journal of Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2019.

[6] J. Huang, Z. Chen, D. Ceylan, and H. Jin, "6-DOF VR videos with a single 360-camera," in *IEEE Virtual Reality*, 2017, pp. 1–8.

[7] 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.

[8] 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.

[9] R. schnabel and R. Klein, "Octree-based point-cloud compression," in *Eurographics Symposium on Point-Based Graphics*, 2006, pp. 111–121.

[10] 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.

[11] 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.

(a) Original

(b) BPSK 1/4
$\mathsf{sMSE}_{\mathsf{xyz}}$: -57.92 dB
sPSNR: 15.52 dB
PSNR: 17.40 dB
SSIM: 0.91

(c) QPSK 1/2
$\mathsf{sMSE}_{\mathsf{xyz}}$: -62.81 dB
sPSNR: 29.10 dB
PSNR: 20.08 dB
SSIM: 0.92

(d) HoloCast
$\mathsf{sMSE}_{\mathsf{xyz}}$: -62.36 dB
sPSNR: 26.12 dB
PSNR: 16.17 dB
SSIM: 0.89

(e) HoloCast+ (Identity)
$\mathsf{sMSE}_{\mathsf{xyz}}$: -68.30 dB
sPSNR: 22.47 dB
PSNR: 23.95 dB
SSIM: 0.93

(f) HoloCast+
$\mathsf{sMSE}_{\mathsf{xyz}}$: $\mathbf{-82.03}$ dB
sPSNR: **46.05** dB
PSNR: **29.95** dB
SSIM: **0.97**

Fig. 11. Snapshot of *milk_color* in the digital-based delivery, HoloCast, and the proposed HoloCast+ schemes at a wireless channel SNR of 10 dB.

[12] 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.

[13] 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.

[14] G. Cheung, E. Magli, Y. Tanaka, and M. K. Ng, "Graph spectral image processing," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 907–930, 2018.

[15] V. Prabhakaran, R. Puri, and K. Ramchandran, "Hybrid digital-analog codes for source-channel broadcast of Gaussian sources over Gaussian channels," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4573–4588, 2011.

[16] T. Zhang and S. Mao, "Metadata reduction for soft video delivery," *IEEE Networking Letters*, vol. 1, no. 2, pp. 84–88, 2019.

[17] J. Park, P. A. Chou, and J. N. Hwang, "Rate-utility optimized streaming of volumetric media for augmented reality," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 149–162, mar 2019.

[18] D. Queiroz, R. L., and P. A. Chou, "Compression of 3D point clouds using a region-adaptive hierarchical transform," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, aug 2016.

[19] 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.

[20] P. de Oliveira Rente, C. Brites, J. Ascenso, and F. Pereira, "Graph-based static 3D point clouds geometry coding," *IEEE Transactions on Multimedia*, vol. 21, no. 2, pp. 284–299, 2019.

[21] S. Gu, J. Hou, H. Zeng, and H. Yuan, "3D point cloud attribute compression via graph prediction," *IEEE Signal Processing Letters*, vol. 27, pp. 176–180, 2020.

[22] P. A. Chou, M. Koroteev, and M. Krivokuca, "A volumetric approach to point cloud compression — Part I: Attribute compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 2203–2216, 2020.

[23] G. Sandri, V. F. Figueiredo, P. A. Chou, and R. D. Queiroz, "Point cloud compression incorporating region of interest coding," in *International Conference on Image Processing*, sep 2019, pp. 4370–4374.

[24] E. Pavez, B. Girault, A. Ortega, and P. A. Chou, "Region adaptive graph Fourier transform for 3D point clouds," 2020. [Online]. Available: https://github.com/STAC-USC/RA-GFT.

[25] 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.

[26] J. Shen, L. Yu, L. Li, and H. Li, "Foveation-based wireless soft image delivery," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2788–2800, 2018.

[27] D. He, C. Luo, F. Wu, and W. Zeng, "Swift: A hybrid digital-analog scheme for low-delay transmission of mobile stereo video," in *ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, Cancun, Mexico, nov 2015, pp. 327–336.

[28] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "FreeCast: Graceful free-viewpoint video delivery," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–11, 2019.

[29] 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*, pp. 1–14, may 2020.

[30] L. Luo, T. Yang, C. Zhu, Z. Jin, and S. Tang, "Joint texture/depth power allocation for 3-D video SoftCast," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 2973–2984, dec 2019.

[31] J. Zhao, R. Xiong, and J. Xu, "OmniCast: Wireless pseudo-analog transmission for omnidirectional video," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 58–70, mar 2019.

[32] 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*, jun 2020, pp. 1–6.

[33] Y. Gui, H. Lu, F. Wu, and C. W. Chen, "Robust video broadcast for users with heterogeneous resolution in mobile networks," *IEEE Transactions on Mobile Computing*, pp. 1–1, jun 2020.

[34] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Compressive sensing for loss-resilient hybrid wireless video transmission," in *IEEE Globecom*, San Diego, CA, dec 2015, pp. 1–5.

[35] H. Hadizadeh and I. V. Bajic, "Soft video multicasting using adaptive compressed sensing," *IEEE Transactions on Multimedia*, pp. 1–1, feb 2020.

[36] X.-W. Tang and X.-L. Huang, "A design of SDR-based pseudo-analog wireless video transmission system," may 2020. [Online]. Available: https://arxiv.org/abs/2005.04558v1

[37] T. Fujihashi, T. Koike-Akino, P. V. Orlik, and T. Watanabe, "DNN-based overhead reduction for high-quality soft delivery," in *IEEE Global Communications Conference*, Dec. 2019, pp. 1–6.

[38] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. Orlik, "Overhead reduction in graph-based point cloud delivery," in *IEEE International Conference on Communications*, 2020, pp. 1–7.

[39] T. Fujihashi, T. K. Akino, S. Chen, and T. Watanabe, "Wireless 3D point cloud delivery using deep graph neural networks," in *IEEE International Conference on Communications*, 2021, pp. 1–6.

[40] X. Fan, F. Wu, and D. Zhao, "D-Cast: DSC based soft mobile video broadcast," in *ACM International Conference on Mobile and Ubiquitous Multimedia*, 2011, pp. 226–235.

[41] Z. Song, R. Xiong, S. Ma, X. Fan, and W. Gao, "Layered image/video softcast with hybrid digital-analog transmission for robust wireless visual communication," in *IEEE International Conference on Multimedia and Expo*, 2014, pp. 1–6.

[42] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 2, pp. 331–345, 2014.

[43] ——, "Wireless cooperative video coding using a hybrid digital-analog scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 436–450, 2015.

[44] H. Cui, Z. Song, Z. Yang, C. Luo, R. Xiong, and F. Wu, "Cactus: A hybrid digital-analog wireless video communication system," in *ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems*, 2013, pp. 273–278.

[45] X. Zhao, H. Lu, C. W. Chen, and J. Wu, "Adaptive hybrid digital-analog video transmission in wireless fading channel," *IEEE Transactions on*

Circuits and Systems for Video Technology, vol. PP, no. 99, pp. 1–14, 2015.

[46] J. Zhang, A. Wang, J. Liang, H. Wang, S. Li, and X. Zhang, "Distortion estimation-based adaptive power allocation for hybrid digital-analog video transmission," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 6, pp. 1806–1818, jun 2019.

[47] P. Yahampath, "Video coding for OFDM systems with imperfect CSI: A hybrid digital–analog approach," *Signal Processing: Image Communication*, vol. 87, p. 115903, sep 2020.

[48] P. Li, F. Yang, J. Zhang, Y. Guan, A. Wang, and J. Liang, "Synthesis-distortion-aware hybrid digital analog transmission for 3D videos," *IEEE Access*, vol. 8, pp. 85 128–85 139, 2020.

[49] 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.

[50] X. Liu, G. Cheung, X. Wu, and D. Zhao, "Random walk graph Laplacian based smoothness prior for soft decoding of JPEG images," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 509–524, 2017.

[51] R. Horaud, "A short tutorial on graph Laplacians, Laplacian embedding, and spectral clustering." [Online]. Available: http://csustan.csustan.edu/ tom/Lecture-Notes/Clustering/GraphLaplacian-tutorial.pdf

[52] P. A. Chou, E. Pavez, R. L. de Queiroz, and A. Ortega, "Dynamic polygon clouds: Representation and compression for VR/AR," Microsoft Research Technical Report, Tech. Rep., 2017.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Imge quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, apr 2004.