Wireless 3D Point Cloud Delivery Using Deep Graph Neural Networks

Takuya Fujihashi¹, Toshiaki Koike-Akino², Siheng Chen², Takashi Watanabe¹,
¹Osaka University
²Mitsubishi Electric Research Laboratories (MERL)

IEEE International Conference on Communications 2021
June 17, 2021
Background

- Volumetric (3D) media
  - Reconstruct 3D scenes with full parallax and depth info.
  - Applications: entertainment, medical imaging, AR/MR

- Key applications in 5G and 5G beyond networks [1]
  - Google and Facebook investigate volumetric video streaming
  - Volumetric video market will grow from $578 million in 2018 to $2.78 billion by 2023 [2]

[2] Volumetric Video Market by Volumetric Capture & Content Creation (Hardware (Camera & Processing Unit), Software, and Services), Application (Sports & Entertainment, Medical, Signage, Education & Training), and Geography - Global Forecast to 2023.
Point Cloud

• Typical data structure for 3D scene
  – Consist of numerous and irregular structure of 3D points
  – Each point has 3D coordinate (x, y, z) information

• Existing schemes use digital-based compression
  – Encoder sets coding parameters based on the current wireless channel quality

Diagram:
- Octree
- Quantization (Pruned tree)
- Entropy coding
- Modulation & Channel coding
Issues of Digital-based Wireless Point Cloud Streaming

Encoder) channel quality is here! Set the best source/channel coding

1) Cliff due to errors in bit-stream

2) Constant due to quantization

Wireless Channel Quality (SNR)

High

Low

Reconstruction quality

BPSK

QPSK

16QAM

Channel quality is varying

BPSK

QPSK

16QAM

200ms

Time [ms]
Issues of Digital-based Wireless Point Cloud Streaming

Ideal:
1) No cliff,
2) Graceful improvement

Wireless Channel Quality (SNR)

Reconstruction quality

Our scheme
Purpose

Our study tackles the following challenging issues
1. Cliff effect
2. Leveling effect: Constant reconstruction quality

Propose a novel scheme for wireless point cloud delivery

- Regard 3D points as graph signals with the attributes of 3D coordinates to deal with irregular structure of holographic data formats
- Introduce graph neural network (GNN)-based and multi-layer perceptron (MLP)-based autoencoder for point coding
- Skip digital-based compression, instead, introduce near-analog modulation to realize graceful quality improvement
GNN

- Extend neural networks to process graph signals
  - Deal with signals in non-Euclidean space

- Applications
  - Physics, Molecule, Text, Social networks, Images

- Typical GNN model
  - Find graph structure
  - Design loss functions and computational modules
Proposed: Graph Construction

Regard 3D points as graph \( g = (V, \epsilon, W) \)

- **\( V \):** vertex (each 3D point), \( \epsilon \): edge
- **\( W \):** adjacency matrix of positive edge weights
  - \( W_{i,j} \): the weight of an edge connecting vertices \( i \) and \( j \)
  - 1: vertices are connected, 0: vertices are not connected
    - Use K-nearest-neighbor graph to make the connection between the vertices

Diagram:
- Original 3D Points
- Graph Construction
- Proposed GNN-based Encoder
- Near-analog modulation
Proposed: GNN-based Encoder

- Transforms 3D points into several latent variables
  - Consists of a series of graph convolution followed by leaky rectified linear unit (ReLU) activation function, Top-K pooling, and a normalization layer
Proposed: GNN-based Encoder

- Graph Convolution: Extract the graph signal features
- Leaky ReLU: learn a mapping from the source to coded signals
- Top-K Pooling: chooses the largest values from each channel to remain important features
Proposed: Near-analog Modulation

Near-analog modulation realizes graceful quality improvement according to wireless channel quality.
Proposed: MLP-based Decoding

- Latent variables $z_i$ are impaired according to a channel transfer function with pre/post equalizations
  
  \begin{align*}
  \eta_{\text{preeq}} &= |h_i|z_i + n_i \\
  \eta_{\text{posteq}} &= z_i + n_i/h_i
  \end{align*}

  - $h_i$: multiplicative fading coefficient
  - $n_i$: effective noise with a variance of $\sigma^2$

- Decoder
  
  - Consists of a series of fully-connected layers and leaky ReLU
Proposed: Loss Function

- GNN-based encoding and MLP-based decoding functions are trained to minimize a loss function
  \((\theta, \phi) = \arg\min_{\theta, \phi} E[ d(p, \hat{p}_{\theta, \phi})] \)

- \(\hat{p}_{\theta, \phi}\): reconstructed 3D coordinates via the proposed encoder and decoder with parameter sets of \(\theta\) and \(\phi\)

- Consider augmented Chamfer distance for loss function
  \[ \max \left\{ \frac{1}{|S|} \sum_{p \in S} \min_{\hat{p} \in \hat{S}} \|p - \hat{p}\|_2, \frac{1}{|\hat{S}|} \sum_{\hat{p} \in \hat{S}} \min_{p \in S} \|p - \hat{p}\|_2 \right\} \]

- If Chamfer distance is even small, the original and reconstructed 3D coordinates are close each other
Evaluation

- Comparative schemes
  - SoftCast [3]: DCT-based
  - HoloCast [4]: GFT-based
  - Proposed: GNN

- Reference point cloud
  - ShapeNet: dataset of 3D points

---

Evaluation

- **Reference point cloud**
  - **ShapeNet**: dataset of 3D points
    - Category: Airplane
    - 2115 point clouds for training, 234 point clouds for testing
Reconstruction Quality vs. Wireless Channel Quality

Graceful improvement

HoloCast (Overhead: $7.2 \times 10^6$ symbols)
HoloCast (Overhead: $2.2 \times 10^6$ symbols)
SoftCast (Overhead: $7.5 \times 10^4$ symbols)
Proposed (Overhead: $9.0 \times 10^4$ symbols)
Reconstruction Quality vs Traffic

The graph illustrates the relationship between Chamfer Distance and Communication Overhead (Symbols). The proposed method (GNN) shows a significant reduction in Chamfer Distance compared to other methods like HoloCast (GFT) and SoftCast (DCT), with a 98.0% reduction in sudden degradation.
Visual Quality of Point Cloud Reconstruction

Original

Chamfer distance: 0.020

SoftCast

HoloCast

Chamfer distance: 0.003

Proposed

Chamfer distance: 0.012
Conclusion

- Proposed a novel scheme for wireless point cloud delivery
  - Regard 3D points as **graph signals** with the attributes of 3D coordinates and color components to deal with irregular structure of holographic data formats
  - Introduce graph neural network (GNN)-based and multi-layer perceptron (MLP)-based autoencoder for point coding
  - Skip digital-based compression, instead, introduce **near-analog modulation** to realize graceful quality improvement
- Validated the advantage
  - Gracefully improve reconstruction quality with the improvement of wireless channel quality
  - Better reconstruction quality with a limited amount of traffic
Q&A

- Please send questions and comments
  fujihashi.takuya@ist.osaka-u.ac.jp