Poster: Mobile Volumetric Video Streaming Enhanced by Super Resolution

Anlan Zhang ChenDong Wang Xing Liu Bo Han* Feng Qian
University of Minnesota, Twin Cities *AT&T Labs - Research

Introduction: Volumetric Videos

- 6DoF (degree of freedom) during playback:
  - positions (X, Y, Z) + orientations (yaw, pitch, roll)
  - highly immersive and interactive
- Capture:
  - RGB-D cameras with depth sensors (Figure 1)
- Representation:
  - **Point Cloud (PtCl)**: a collection of points
  - 3D Mesh
- Enable novel applications
  - Entertainment, health care, education, and etc.

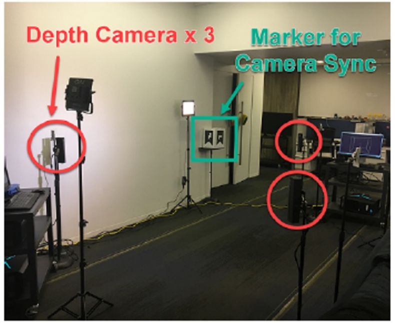
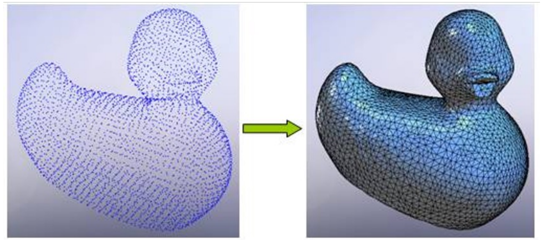


Figure 1: The volumetric video capturing system in our lab.



PtCl to 3D Mesh

Introduction: Motivation

- Streaming PtCl Volumetric Videos
  - Extremely bandwidth demanding, even after compression
  - For a high-resolution PtCl footage, its data rate can be as high as 6Gbps[1]

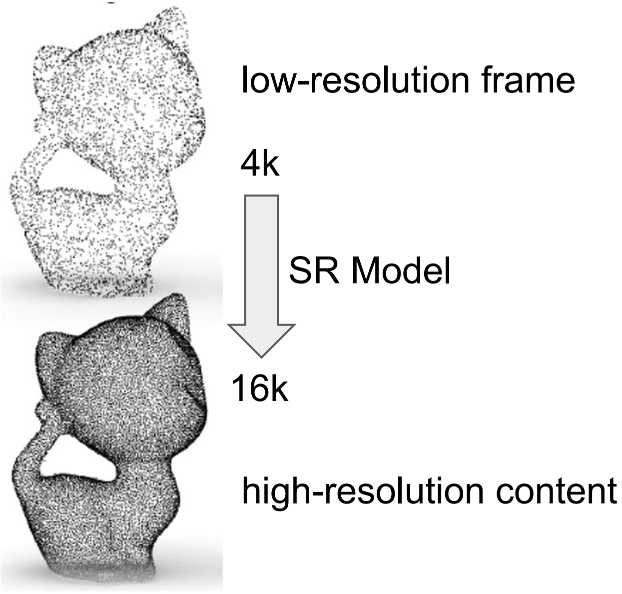
Challenge

How to stream high-quality volumetric contents wirelessly to commodity mobile devices in real-time (30FPS) while maintaining a good user's Quality of Experience (QoE)?

VoluSR: Kea Idea

- **Apply 3D super-resolution (SR) to enhance volumetric video quality**, striking a tradeoff between the network resource and the client-side computation capability.

- SR (or Upsampling) for 3D PtCls
  - A DNN SR model learning:
    low-resolution content → high-resolution details
  - Online inference stage



Enhancing PtCl Video Quality by 3D SR

- A Benchmark
  - Setup:
    - SR model: PU-GAN[1]
    - PtCl video: captured in our lab, ~100k points/frame
    - Hardware: A desktop PC with an Nvidia 2080Ti GPU
    - **Upsampling Ratio: x4, i.e. ~25k to ~100k**
    - Evaluation across 100 frames
  - Pilot Results:
    - Average Chamfer Distance (CD): $0.33 * 10^{-3}$ m^2
    - Video FPS: < 2

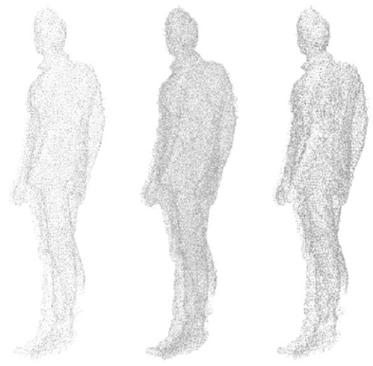


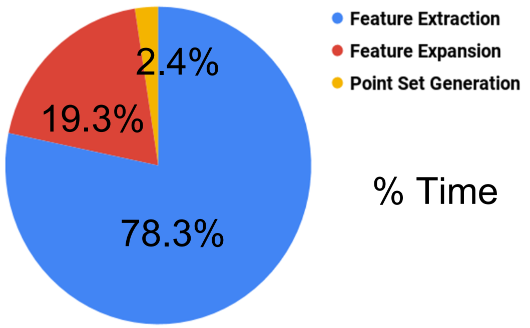
Figure 2: Left: a low-resolution frame (input to PU-GAN), Middle: the high-resolution frame inferred by PU-GAN, Right: the ground truth high-resolution frame.

$$\mathrm{CD}(S_1, S_2) = \frac{1}{|S_1|}\sum_{x\in S_1}\min_{y\in S_2}\|x-y\|_2^2 + \frac{1}{|S_2|}\sum_{y\in S_2}\min_{x\in S_1}\|x-y\|_2^2$$

SR model achieves a good accuracy by leveraging overfitting while suffering poor runtime performance

VoluSR: Proposed Optimizations

- Speeding Up Model Inference
  - Observations:
    - Feature Extraction is the bottleneck of Vanilla PUGAN[1]
  - Strategies:
    - Reducing model complexity
      - Replace feature extraction with a more efficient spherical kernel[2] for 3D PtCls graph convolution
    - General DNN Model Acceleration Methods
      - Pruning & Quantization



Profile the running time of the PU-GAN model.

VoluSR: Proposed Optimizations

- Caching & Reusing Inference Results
  - Observations:
    - Volumetric videos exhibit significant similarities across frames
  - Strategies:
    - Partition each frame into 3D tiles
    - Cache the inference results of the tiles and reuse them aggressively
    - Approximate a tile using a geometrically similar cached tile and a lightweight patch that delta-encodes the difference

VoluSR: Proposed Optimizations

- Adapting to User's Perception
  - Observations:
    - Use human users' perception to reduce the computational workload
  - Strategies:
    - Conduct SR for tiles:
      - Fall into the predicted viewport
      - Are not blocked by other tiles
      - Bear a close physical distance to the viewpoint
      - Have sufficiently high brightness

VoluSR: Proposed Optimizations

- Adapting to Devices' Computation Capabilities
  - Observations:
    - Heterogeneity of mobile devices
    - Different tiles have different visual importance
      - E.g., a closer tile may take a higher priority than a tile that is far from the viewpoint.
  - Strategies:
    - Dynamically adjusts the upsampling ratio for each tile based on their visual importance

System-level Optimization and Integration

- We are working on developing the holistic VoluSR system:
  - Pipelining
  - Offloading some tasks to the edge/server
  - Will thoroughly evaluate our prototype using real PtCl videos, real users' viewport traces, and off-the-shelf mobile devices.

Thanks for listening!