Joint Learning of Point Clouds and Motion Vectors

Presenter: Cheng-Tse Lee, National Tsing Hua University
Advisor: Cheng-Hsin Hsu, National Tsing Hua University, Taiwan

Outline

- Introduction
- Challenge and Solution
- Solution
- Algorithm: Gaussian as a Proxy (GaaP)
- Algorithm: Direct Point Cloud (DPC)
- Experiments
- Conclusion

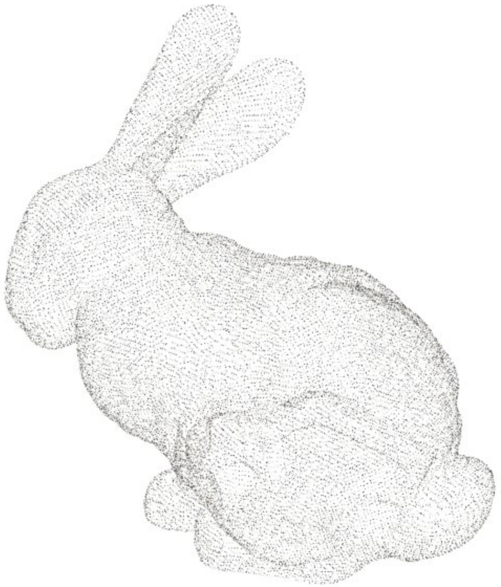
Volumetric Videos Capture 3D Worlds

Users get to navigate in 6 Degrees-of-Freedom (6DoF)

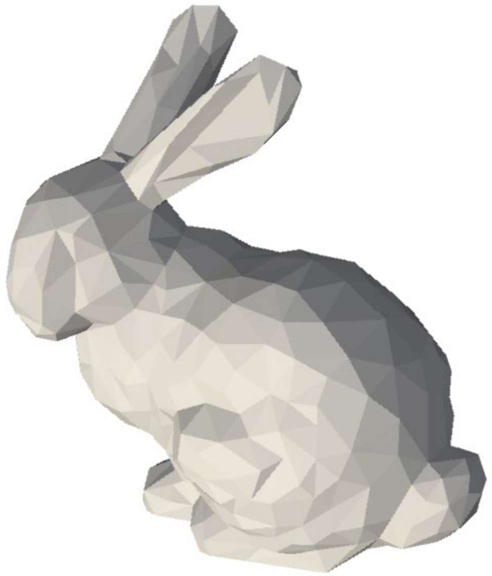
Representations of Volumetric Video

Traditional Representations

Neural Network Representations



Point Cloud



Mesh

3DGS [1]

NeRF [2]

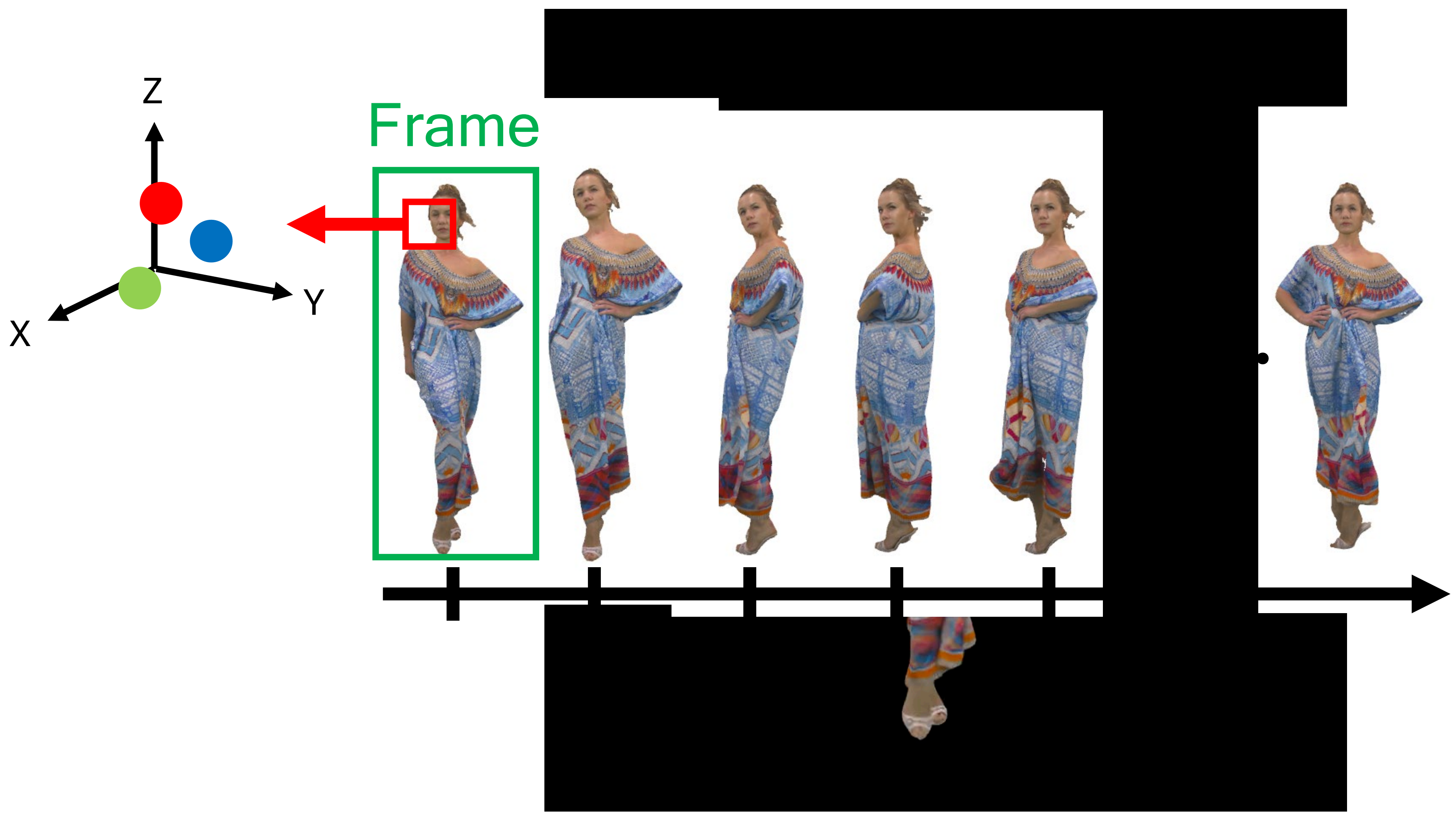
Less

More

Constructing / Rendering Overhead

[1] Yang et al., "Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction," Proc. of the IEEE/CVF conference on computer vision and pattern recognition, 2024
[2] Mildenhall et al., "Nerf: Representing scenes as neural radiance fields for view synthesis," Communications of the ACM, 2021

Dynamic Point Cloud



Group of Frame

Unstructured

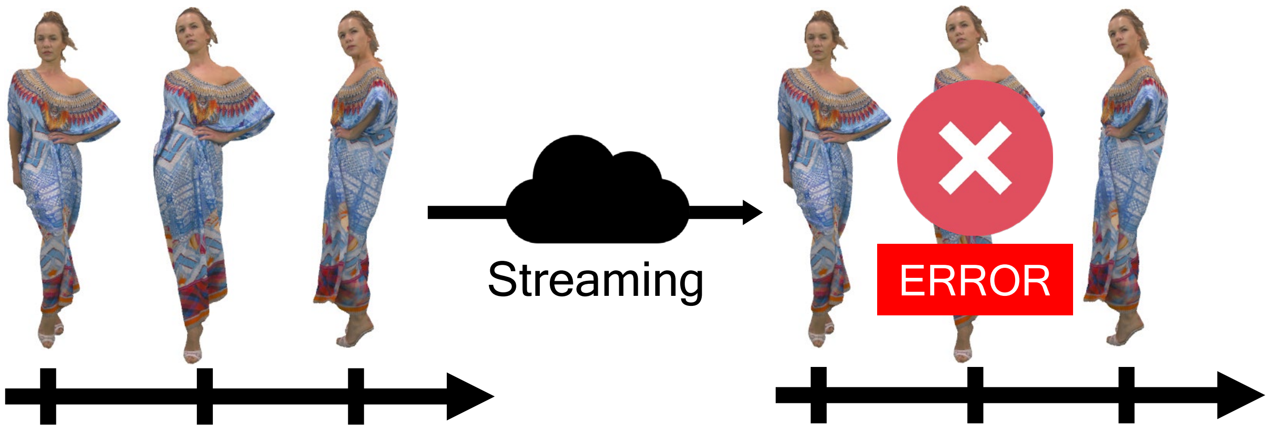
Lightweight

Outline

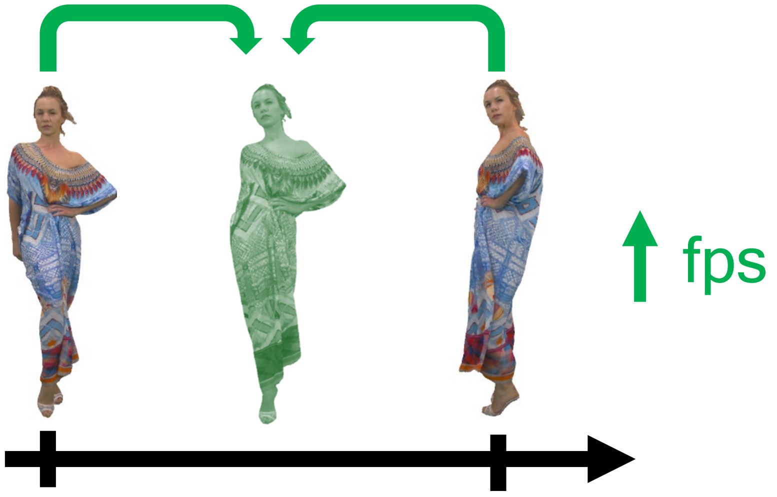
- Introduction
- Challenge and Solution
- Algorithm: Gaussian as a Proxy (GaaP)
- Algorithm: Direct Point Cloud (DPC)
- Experiments
- Conclusion

Limitations of Unstructured Dynamic Point Cloud

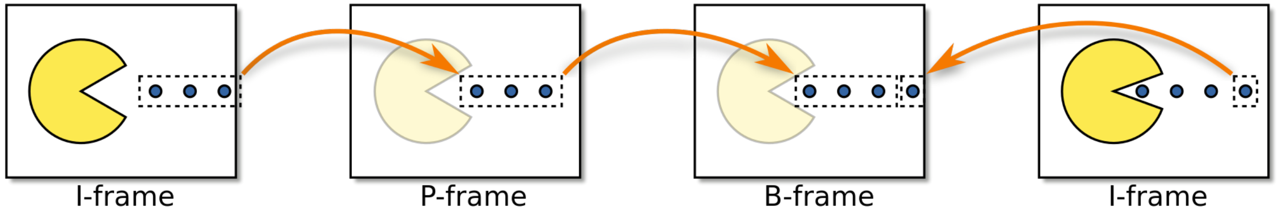
- Unstructured nature → No mapping of points across frames
- Difficult to optimize some applications without **point matching**



1. Error Concealment



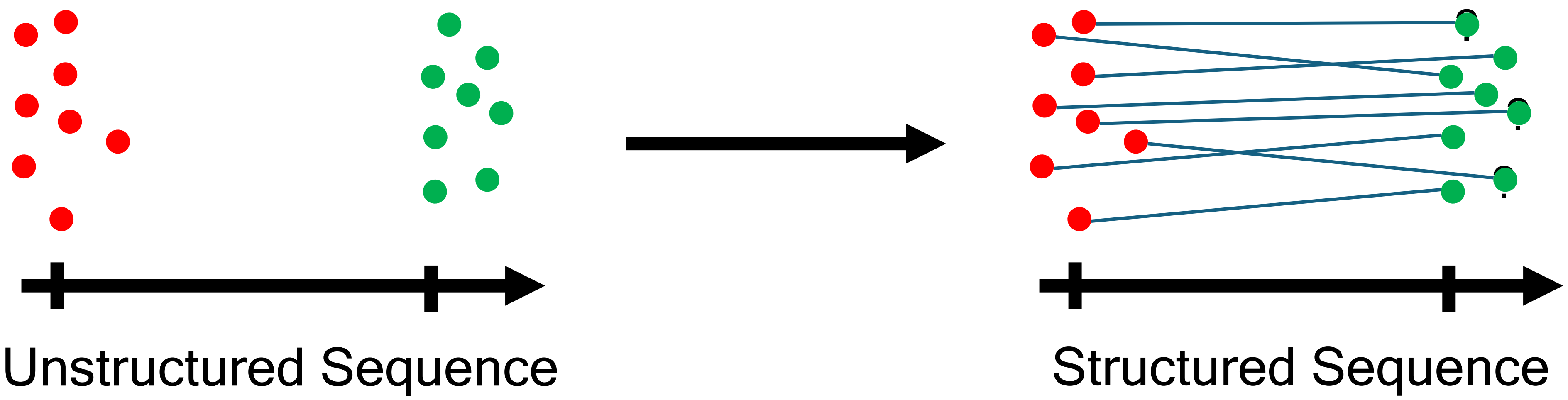
2. Temporal Super-resolution



3. Interframe Coding

The Key Idea of Our Work

Q: How to construct a structured dynamic point cloud from unstructured dynamic point cloud



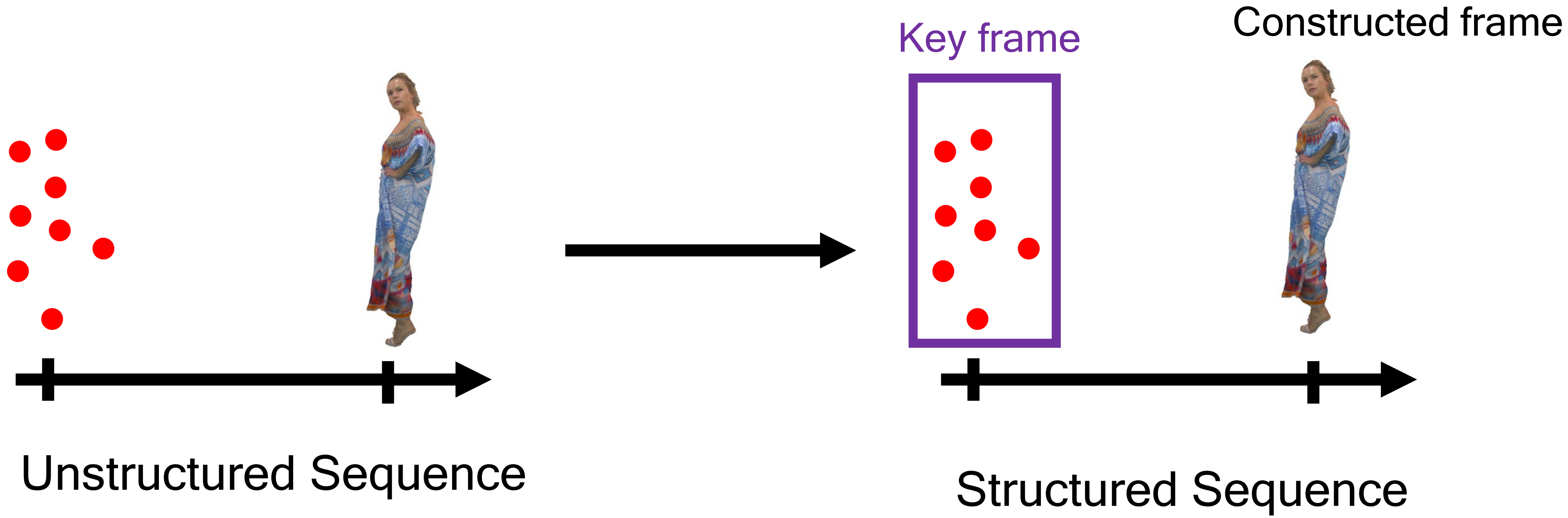
The Key Idea of Our Work

Q: How to construct a structured dynamic point cloud
from unstructured dynamic point cloud

Different point cloud can have similar rendered visual appearance

The Key Idea of Our Work

Q: How to construct a structured dynamic point cloud from unstructured dynamic point cloud



Learn and reconstruct the point cloud to obtain more accurate motion vectors

Problem Formulation

$$\{\hat{\mathbf{X}}_n^* \mid n = 2, 3, \ldots, N\} = \underset{\{\hat{\mathbf{X}}_n \mid n=2,3,\ldots,N\}}{\operatorname{argmin}} \sum_{n=2}^{N} \sum_{\theta \in \Theta} Q(R(\underbrace{\mathbf{X}_n}_{\text{Input non-key frame}}, \overbrace{\boxed{\theta}}^{\text{Camera parameter}}), R(\underbrace{\hat{\mathbf{X}}_n}_{\text{Learned non-key frame}}, \theta))$$

Given a GoF with N point cloud frames:

- $R(\mathbf{X}, \theta)$: rendered view of point cloud $\mathbf{X}$ under camera θ
- $Q(\cdot, \cdot)$: quality degradation between two rendered view

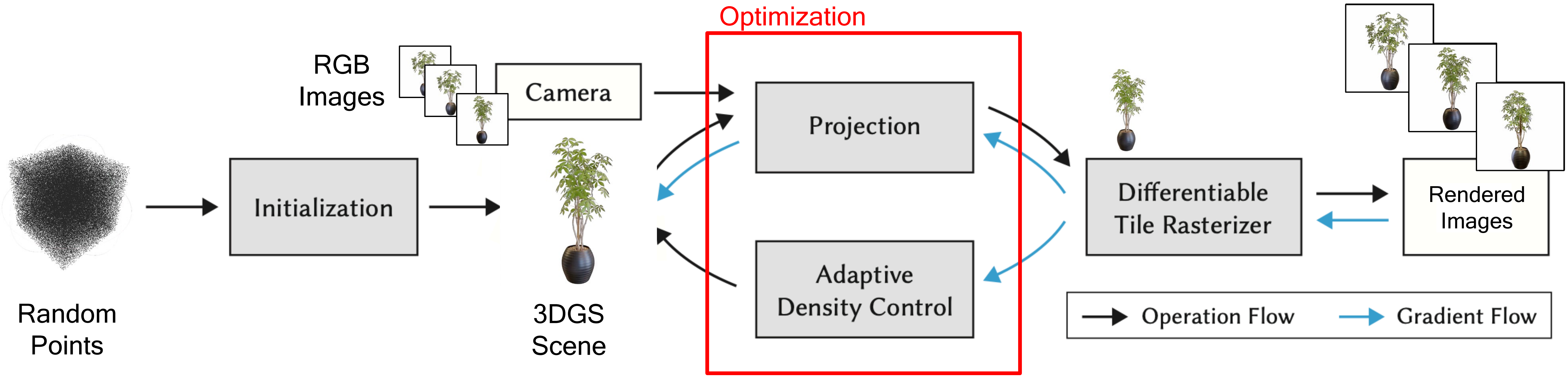
Goal: learn non-key frames whose rendered views are close to the input frames.

Outline

- Introduction
- Challenge and Solution
- Algorithm: Gaussian as a Proxy
- Algorithm: Direct Point Cloud
- Experiments
- Conclusion

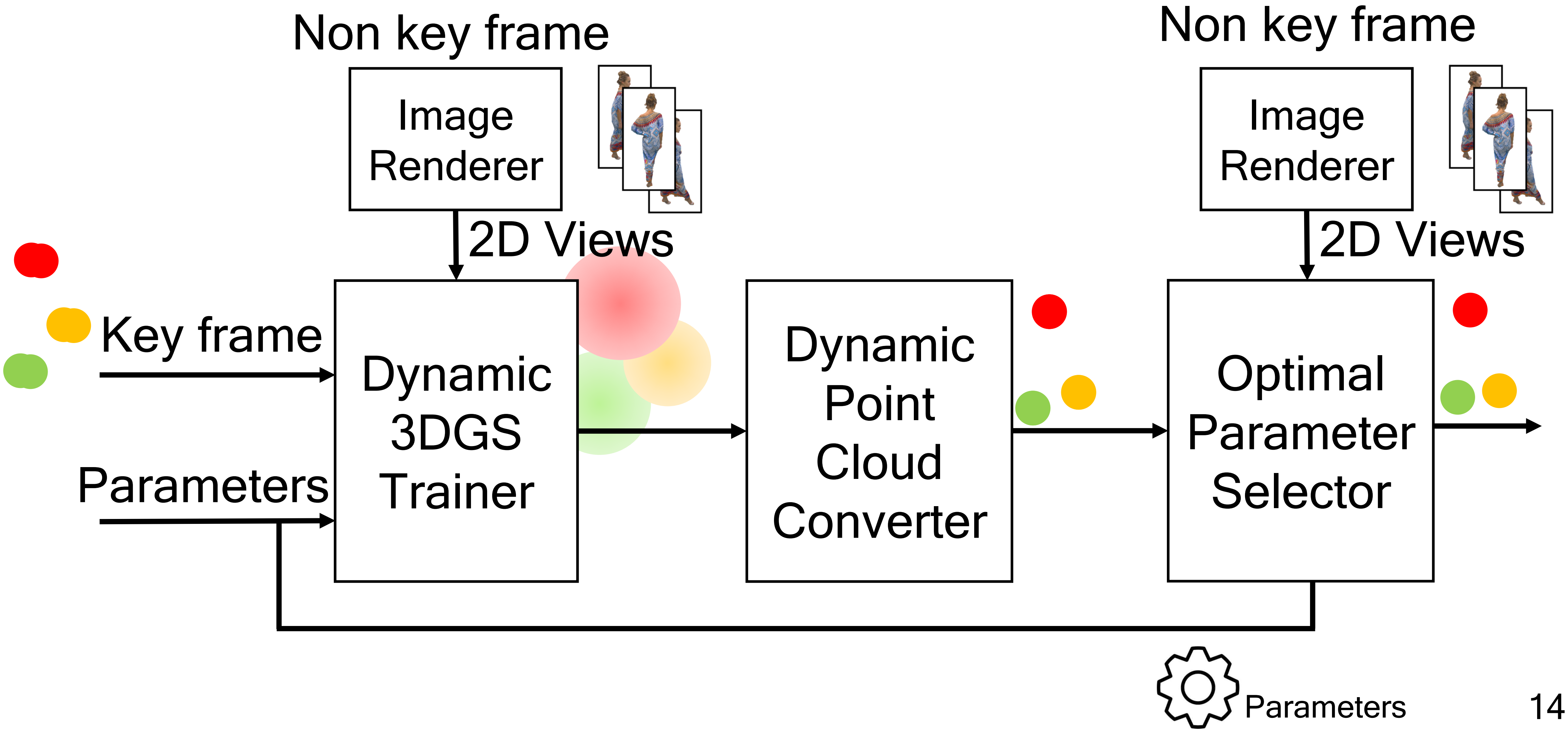
3D Gaussian Splatting (3DGS) Training [1]

- Neural-based and point-based representation
- Simple generation pipeline

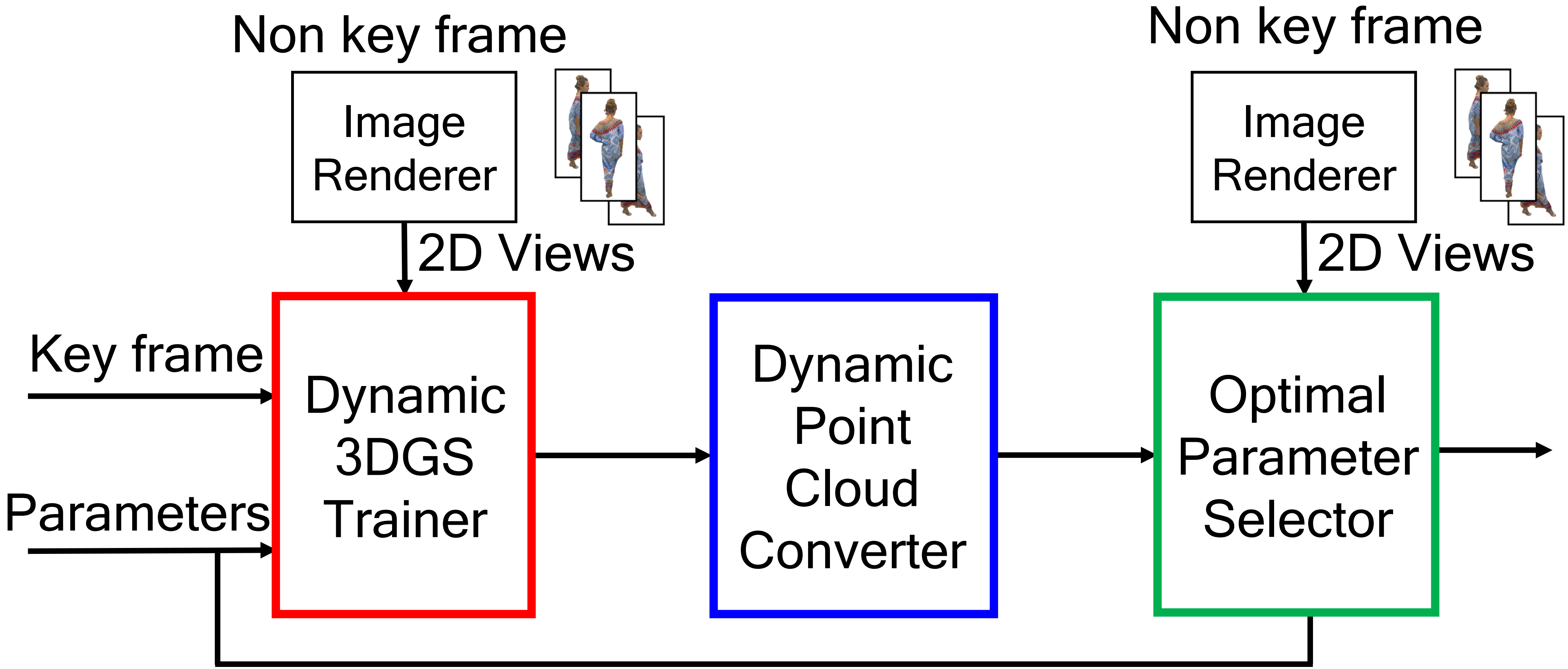


[1] Kerbl et al., "3D Gaussian Splatting for Real-time Radiance Field Rendering," ACM Transactions on Graphics, 2023

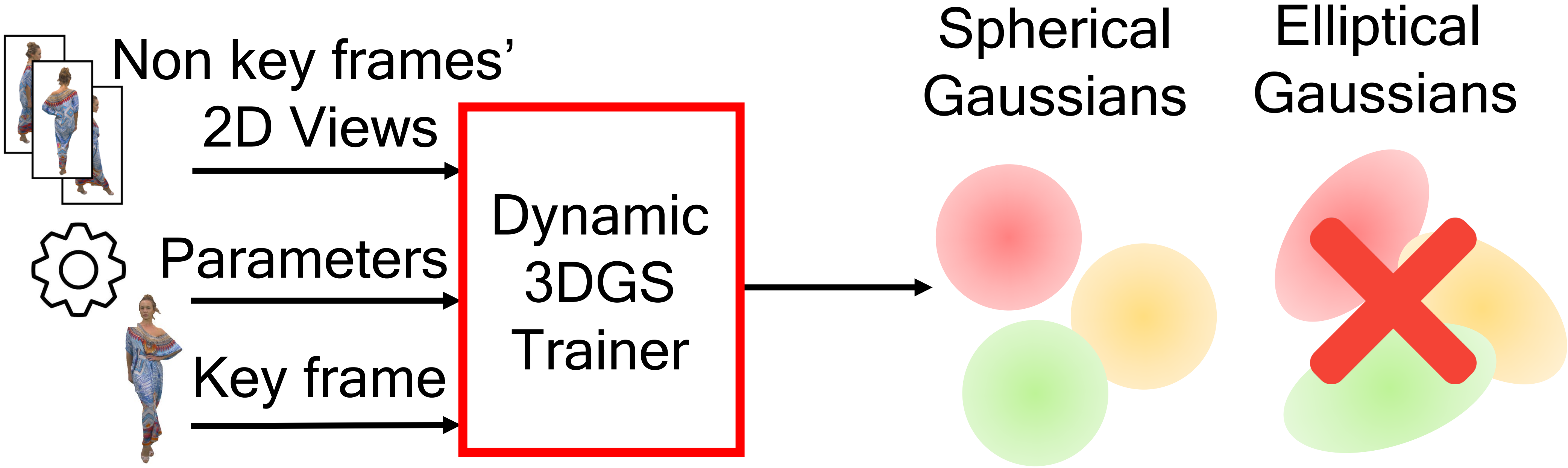
Pipeline of Our GaaP Algorithm



Our Proposed GaaP Algorithm

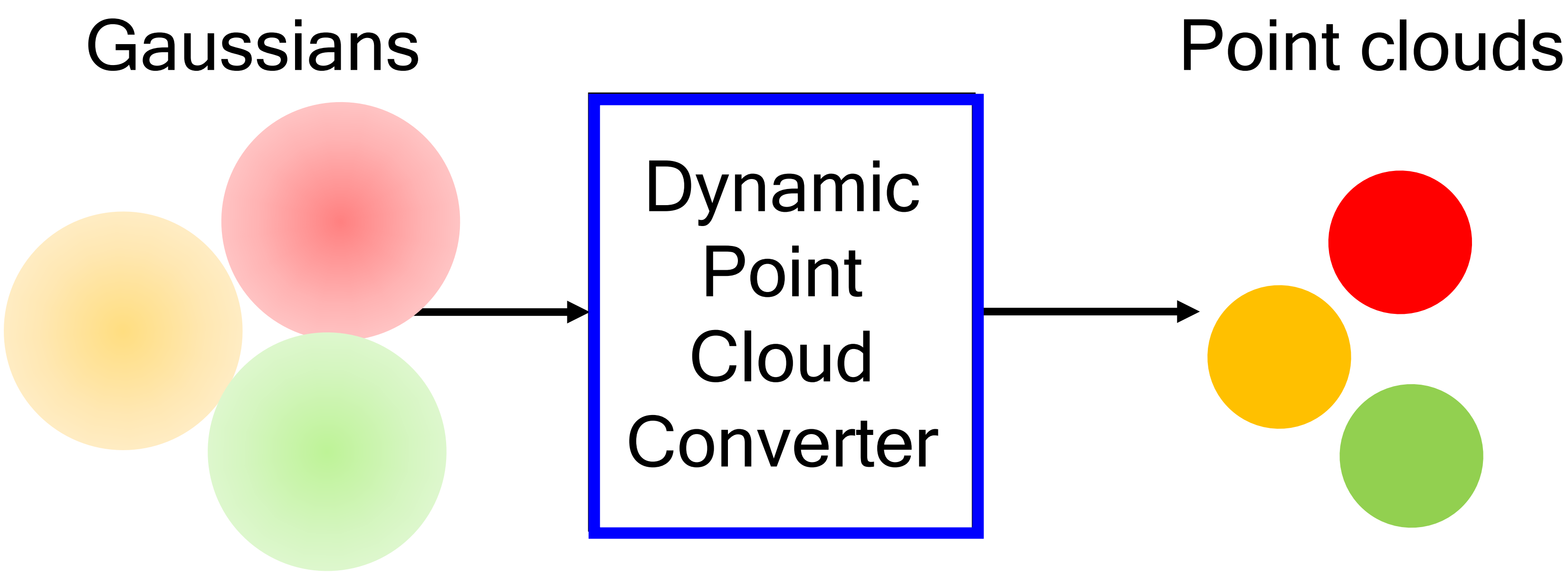


Dynamic 3DGS Trainer



Scale of Gaussians significant affect constructing results

Dynamic Point Cloud Converter

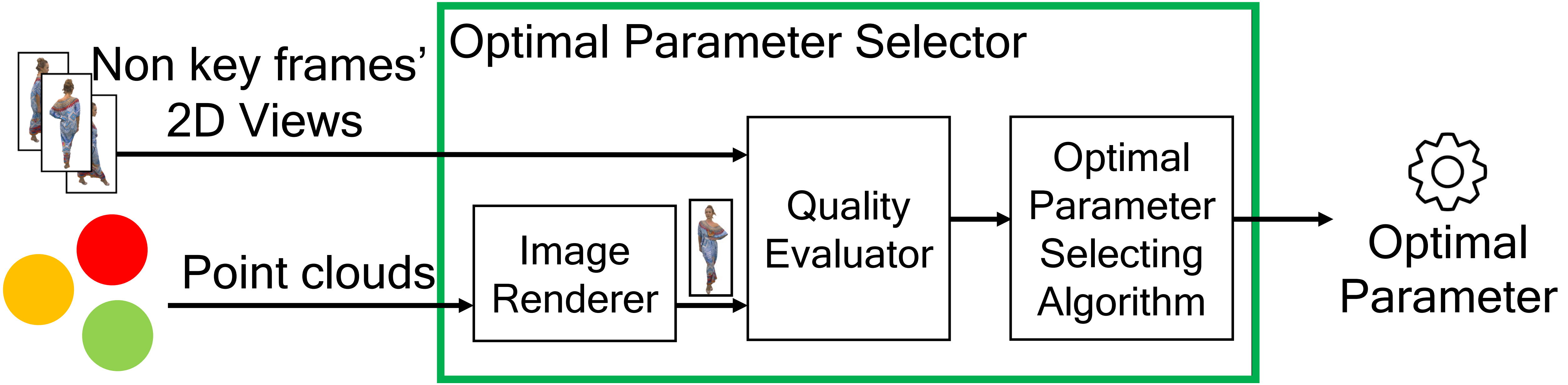


Optimal point size

Large point size

Size of point cloud significant affect the visual quality

Optimal Parameter Selector



Parameters to select: **scale of Gaussian** **and** **size of point cloud**

Steps:

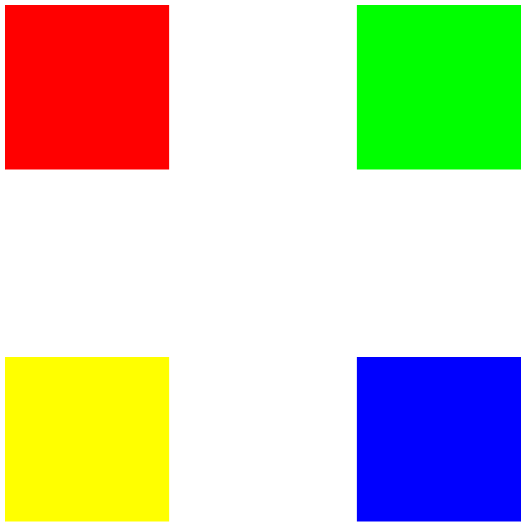
1. Fixed the size and use binary search to find the optimal scale
2. Linear search to get the optimal size

Outline

- Introduction
- Challenge and Solution
- Algorithm: Gaussian as a Proxy (GaaP)
- Algorithm: Direct Point Cloud (DPC)
- Experiments
- Conclusion

The Importance of Differentiable Renderer

- We need a differentiable renderer that allow for gradient backpropagation



Open3D: Hard blending, non-differentiable [1]

Pulsar: Soft blending, differentiable [2]

[1] Zhou et al. "Open3D: A modern library for 3D data processing." *arXiv preprint arXiv:1801.09847* (2018).

[2] Lassner et al. "Pulsar: Efficient sphere-based neural rendering." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

Two Point Cloud Differentiable Renderer

PyTorch3D [1]

- Soft point splatting with alpha compositing

\+ Easy to implement
− Lower visual quality and slower learning

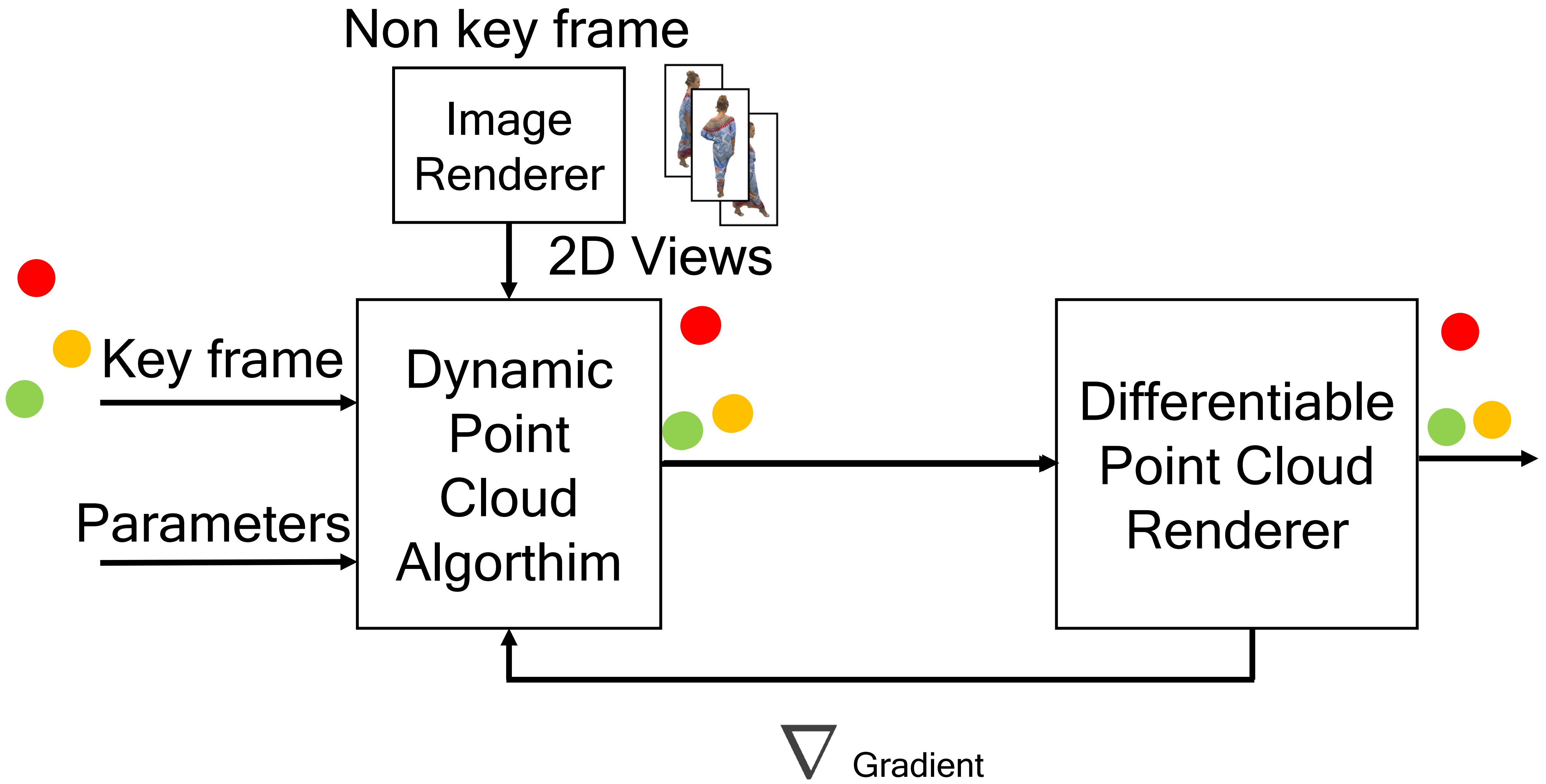
Pulsar

- Sphere-based ray rendering with soft compositing

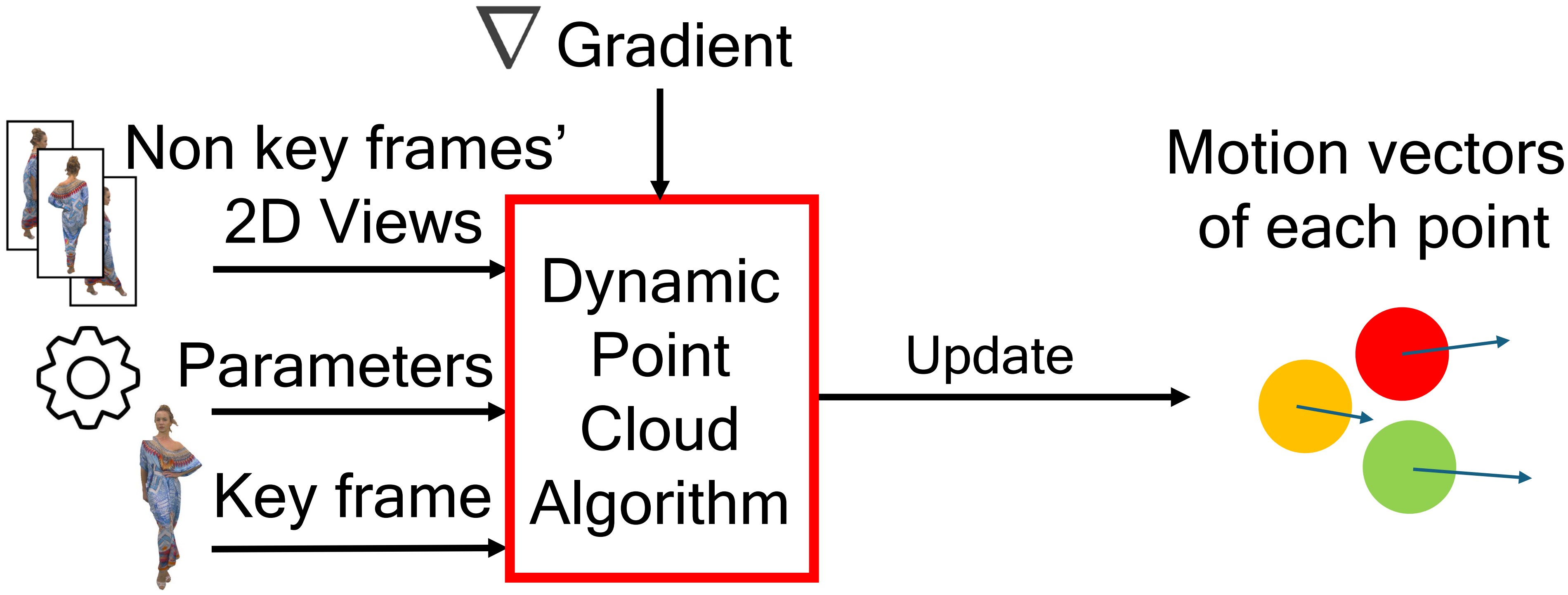
\+ Higher visual quality and faster learning
− Sensitive to depth sharpness parameter γ

Ravi et al. "Accelerating 3d deep learning with pytorch3d." *arXiv preprint arXiv:2007.08501* (2020).

Our Proposed DPC Algorithm



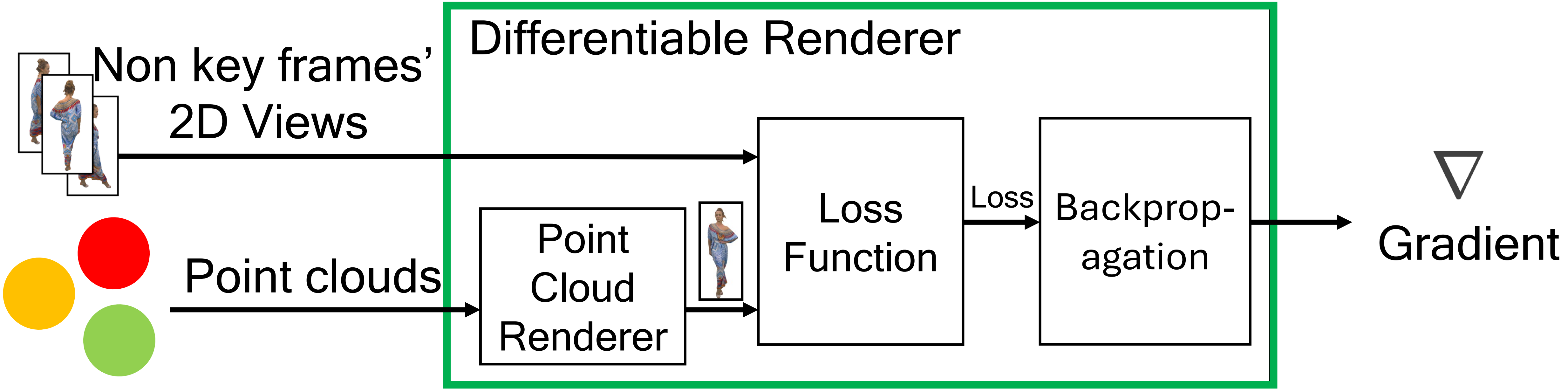
Dynamic Point Cloud Algorithm



Parameters

1. Rendering point size - controls point influence on image plane

Differentiable Renderer

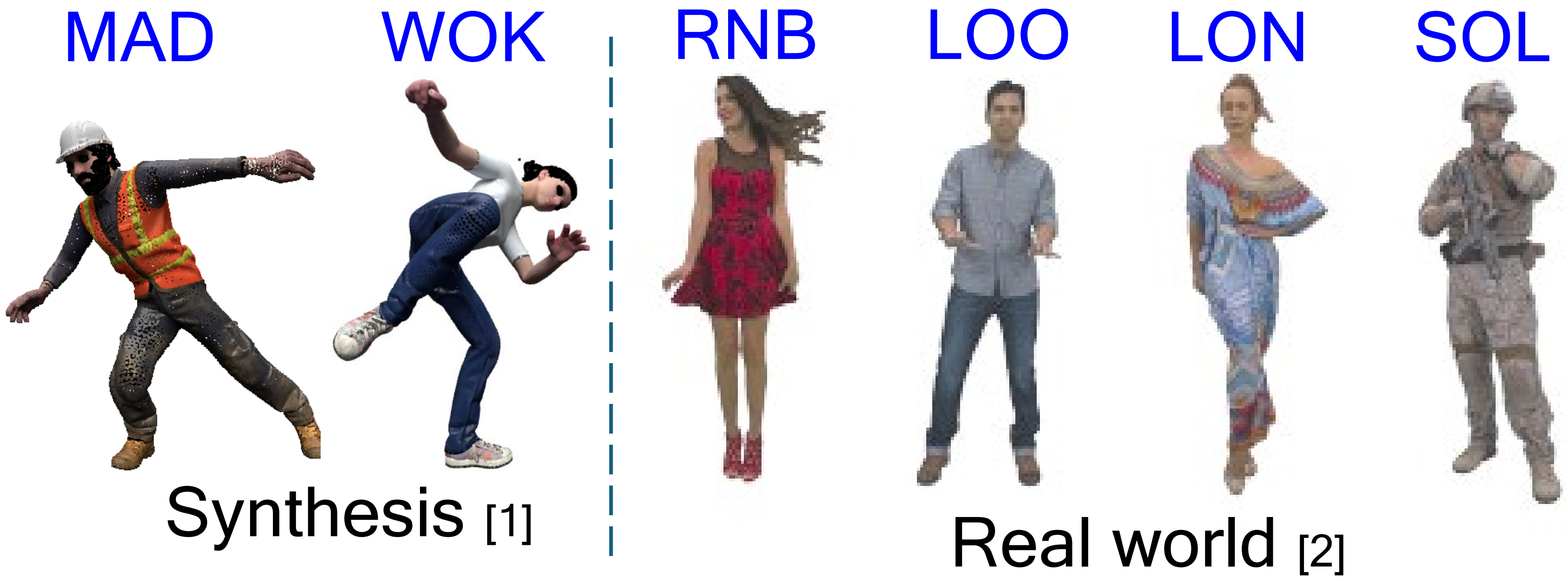


Outline

- Introduction
- Challenge and Solution
- Algorithm: Gaussian as a Proxy (GaaP)
- Algorithm: Direct Point Cloud (DPC)
- Experiments
- Conclusion

Experiment Setup

- Datasets [1][2]:



- Metrics:
  - Visual quality: foreground PSNR and SSIM
  - 3D quality: PSNR-D1 and YPSNR
  - Motion Vector quality: Temporal/Spatial Smoothness

[1] Sun et al. "A dynamic 3d point cloud dataset for immersive applications." Proceedings of the 14th Conference on ACM Multimedia Systems. 2023.
[2] Eon et al., “8i Voxelized Full Bodies, version 2 – A Voxelized Point Cloud Dataset,” 2017

Experiment Setup cont.

- Each sequence has 72 sampled frames
- Downstream applications:
  - Error Concealment: drop 1, 2, or 4 frames, then evaluate visual quality
  - Temporal Super-Resolution: insert 1, 2, or 4 frames, then evaluate warping error
  - Source Coding: encode motion data and report bitstream size

Impact of Optimal Parameter Selection in GaaP

Default Parameters
(PSNR=19.5 dB)

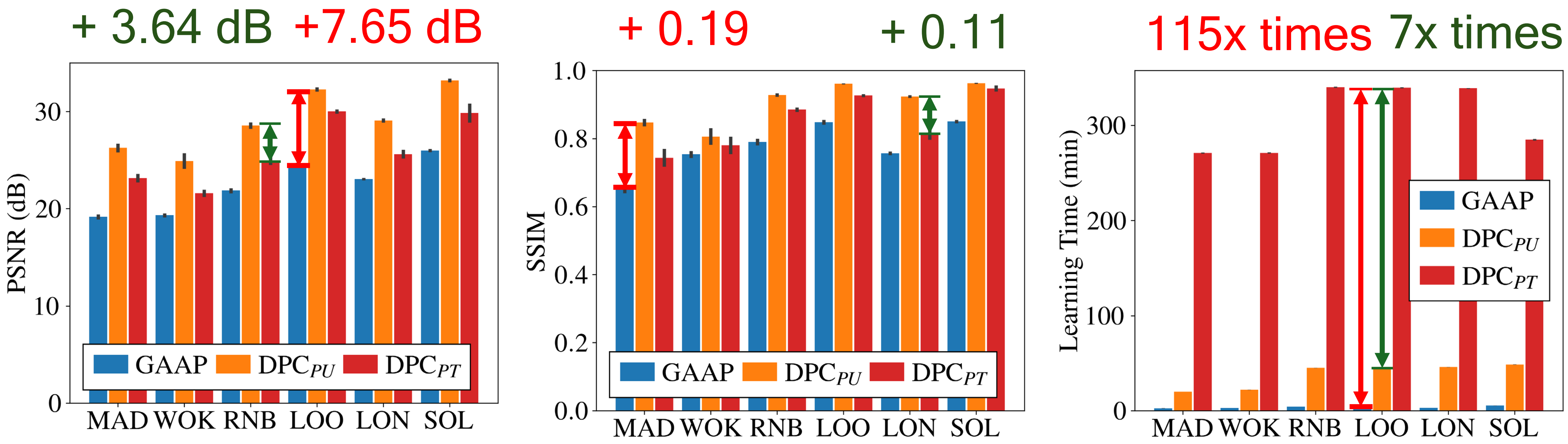
With
Optimal Selection Alg.
(PSNR=28.2 dB)

Scale of Gaussians significant affect constructing results

+8.7 dB in PSNR

+0.2 in SSIM

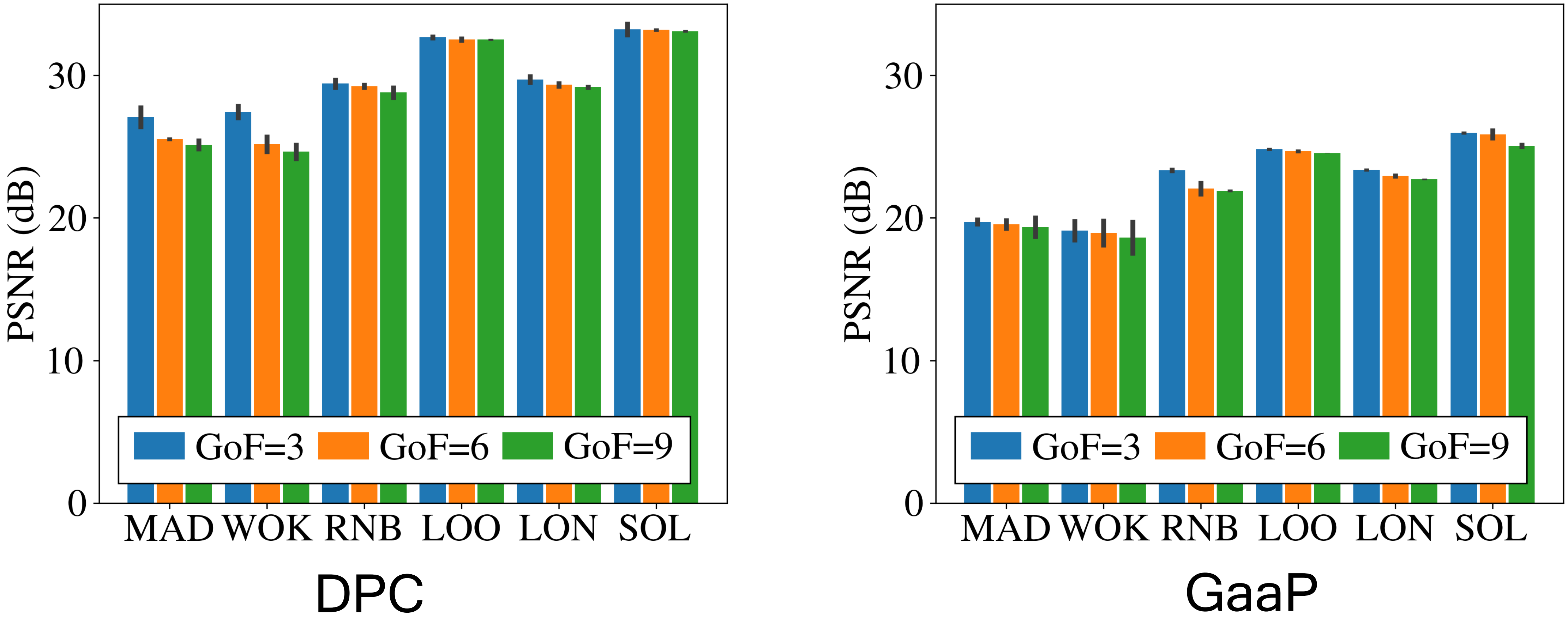
Comparison between Different Algorithms



Visual Quality: DPC$_{PU}$ > DPC$_{PT}$ > GaaP

Computation Overhead: GaaP > DPC$_{PU}$ > DPC$_{PT}$

Impact of Group of Frame (GoF) Size

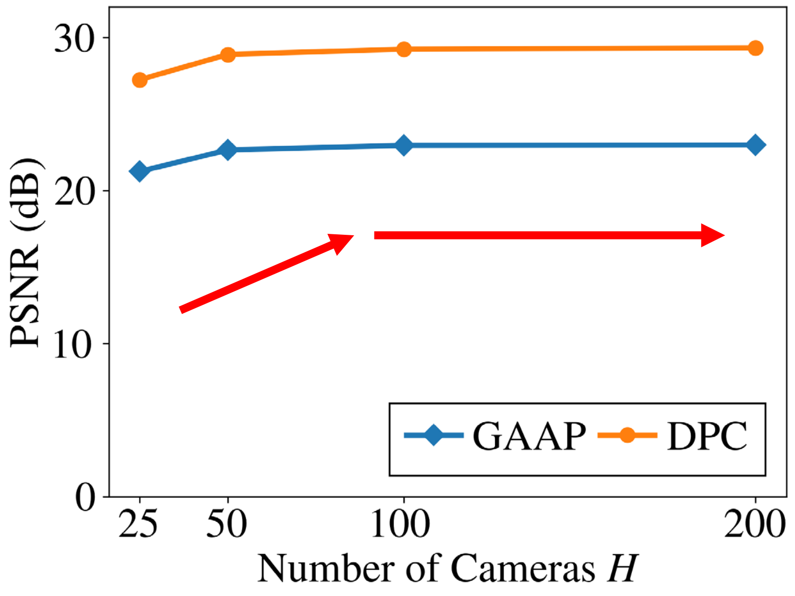


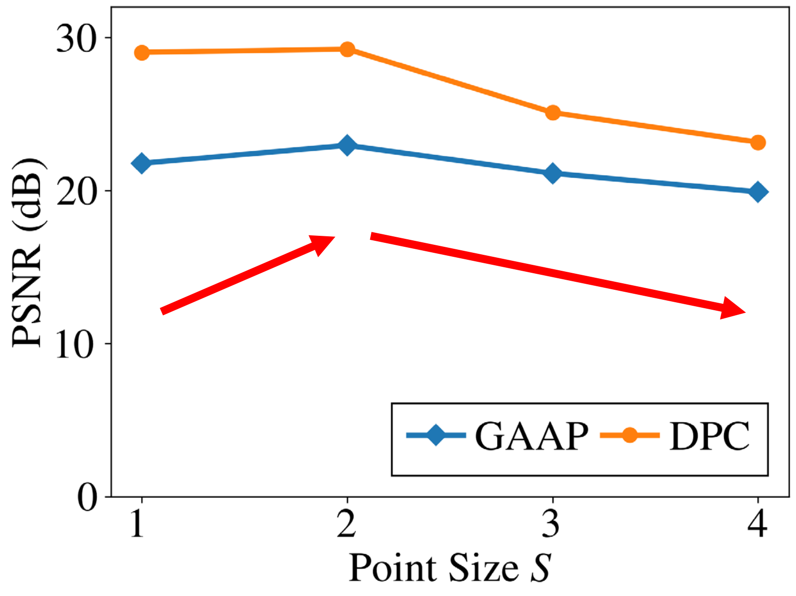
GoF Size = 6 is chosen as default setting

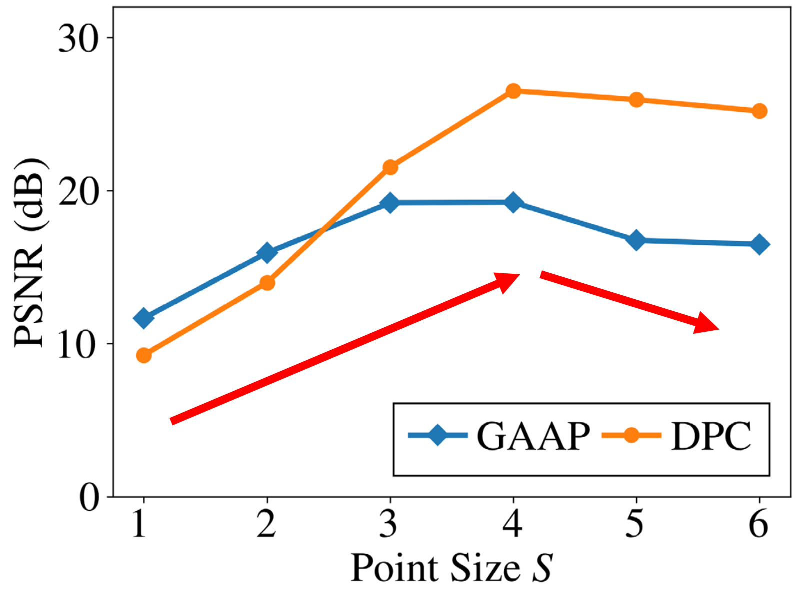
Impact of Number of Cameras and Point Size

Camera

Point Size



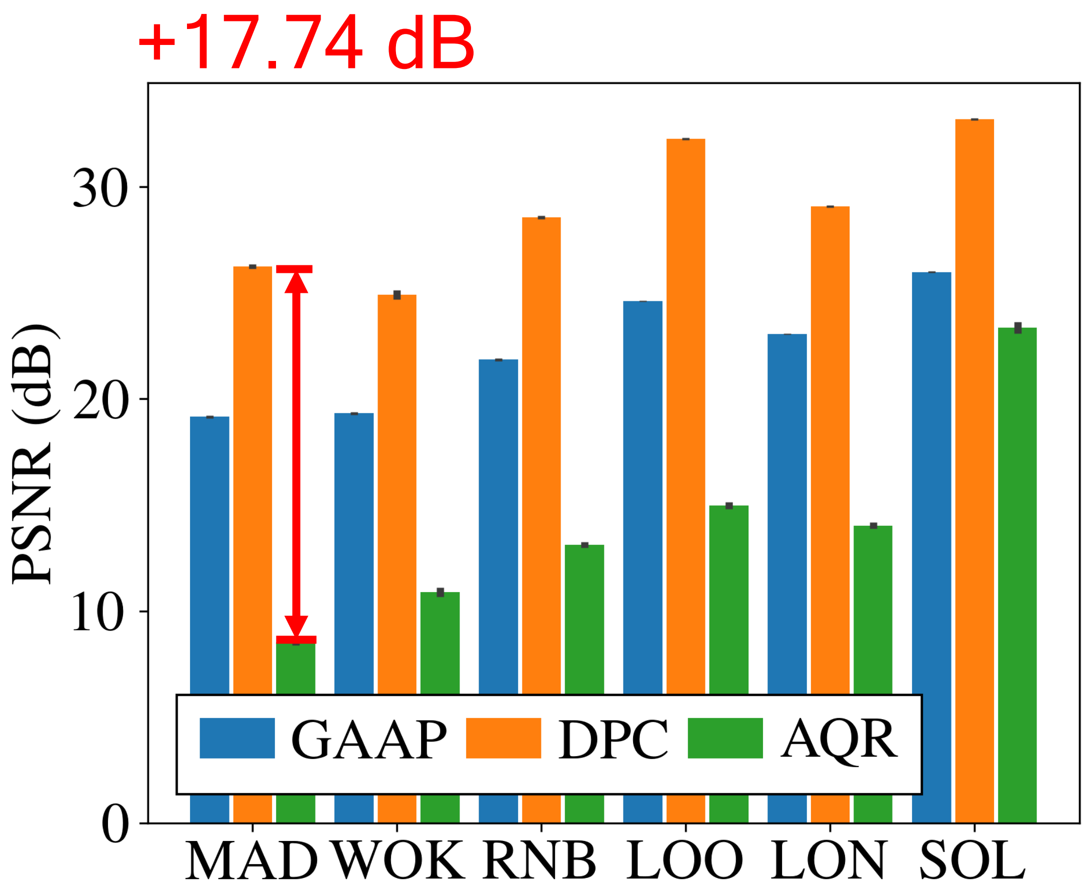


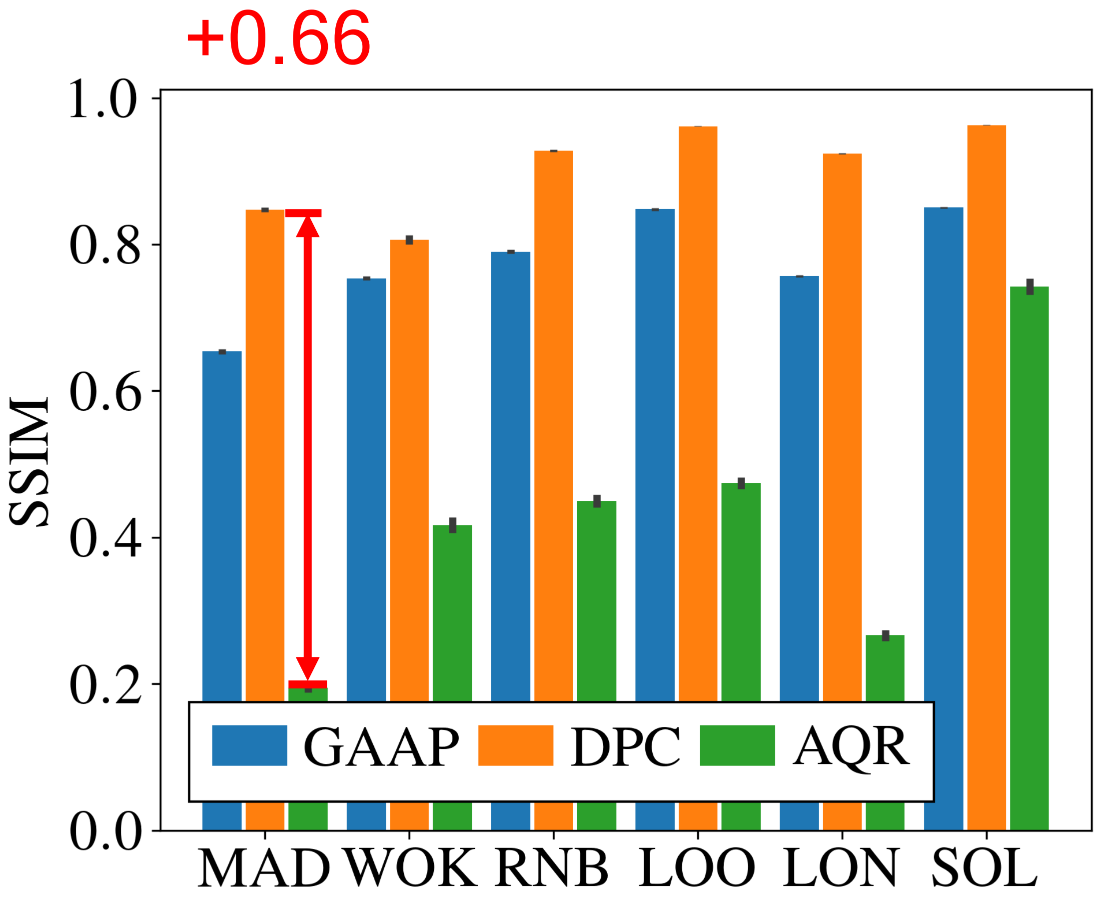


Default Setting:

- 100 Cameras + Point size set to 2 for real world, 4 for synthesis

Visual Quality of Our Algorithms

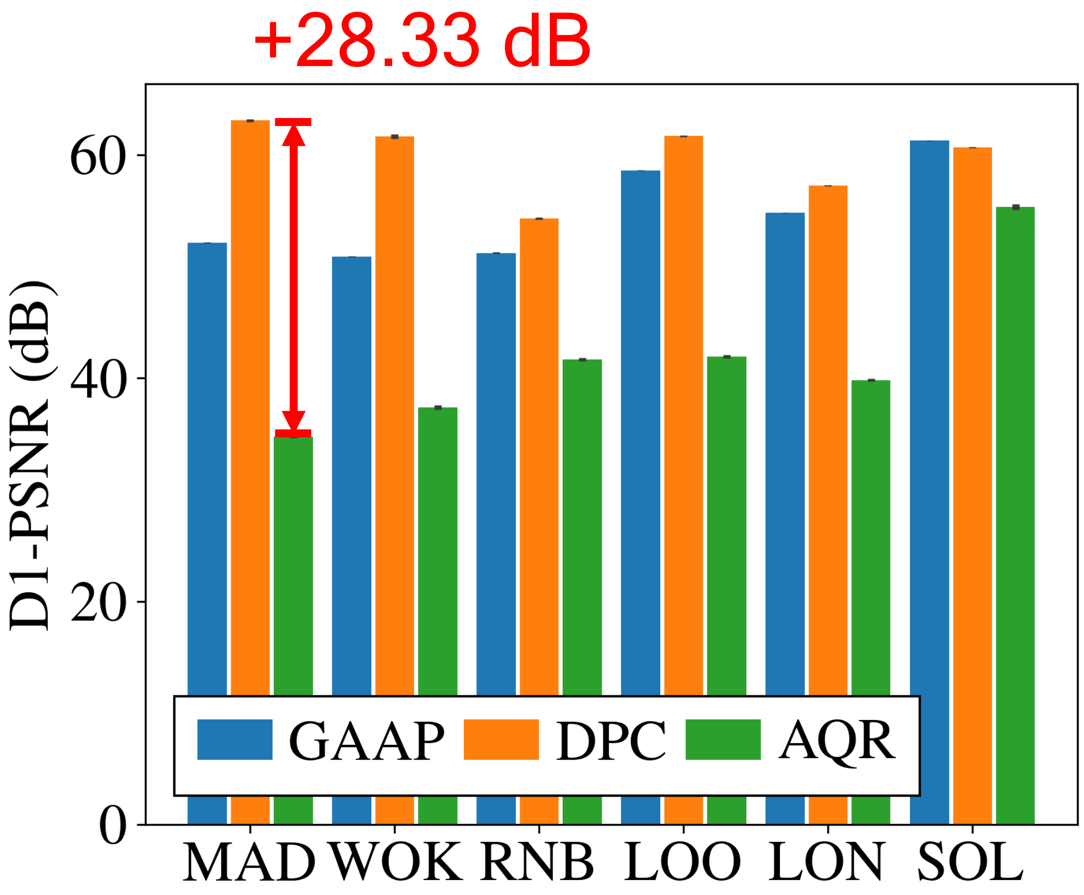


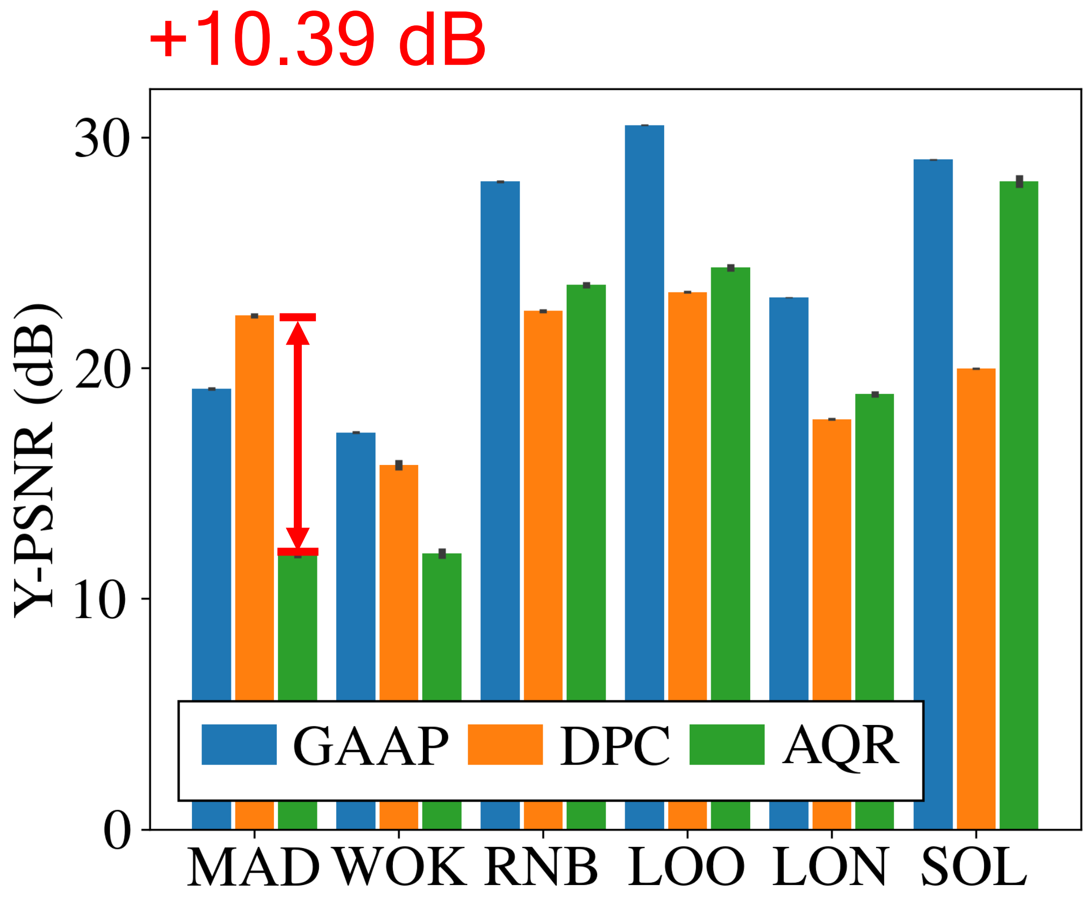


AQR: Adaptive Query Radius [1], the state-of-the-art heuristic algorithm

[1] Huang, I-Chun, et al. "Composing Error Concealment Pipelines for Dynamic 3D Point Cloud Streaming." ACM Transactions on Multimedia Computing, Communications and Applications 21.6 (2025): 1-28.

3D Quality of Our Algorithms



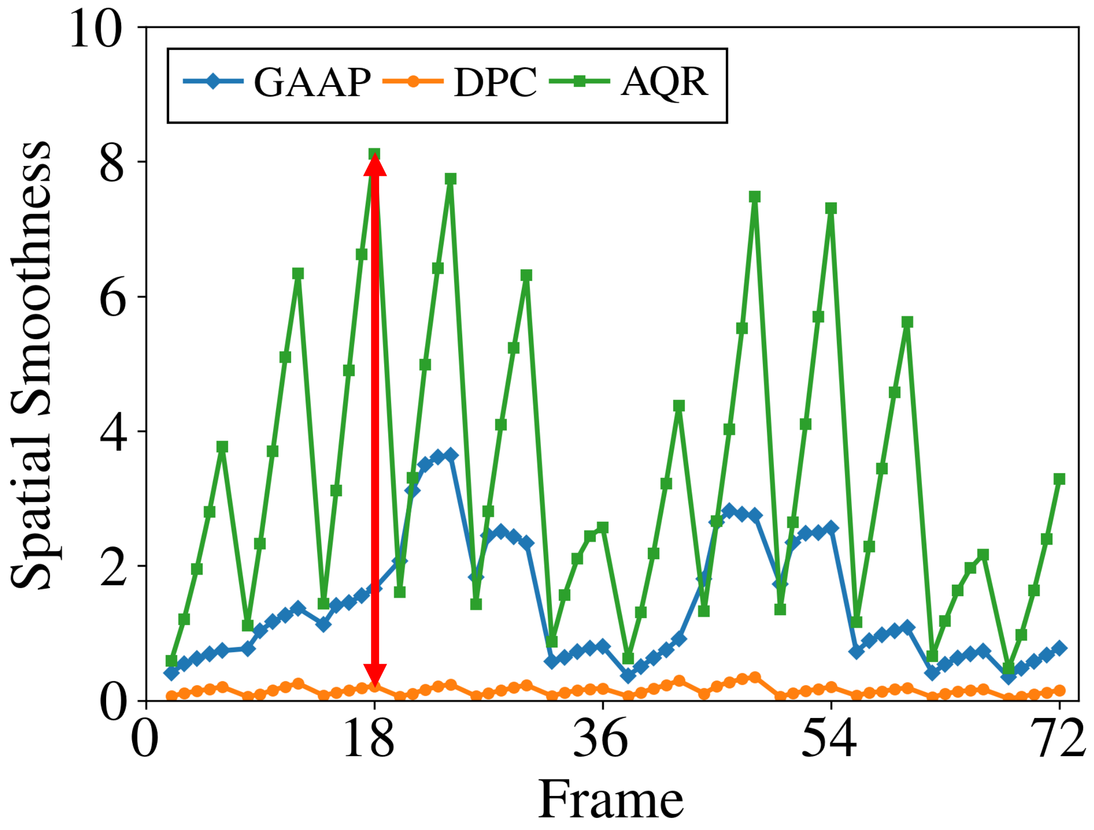


D1-PSNR: geometry quality

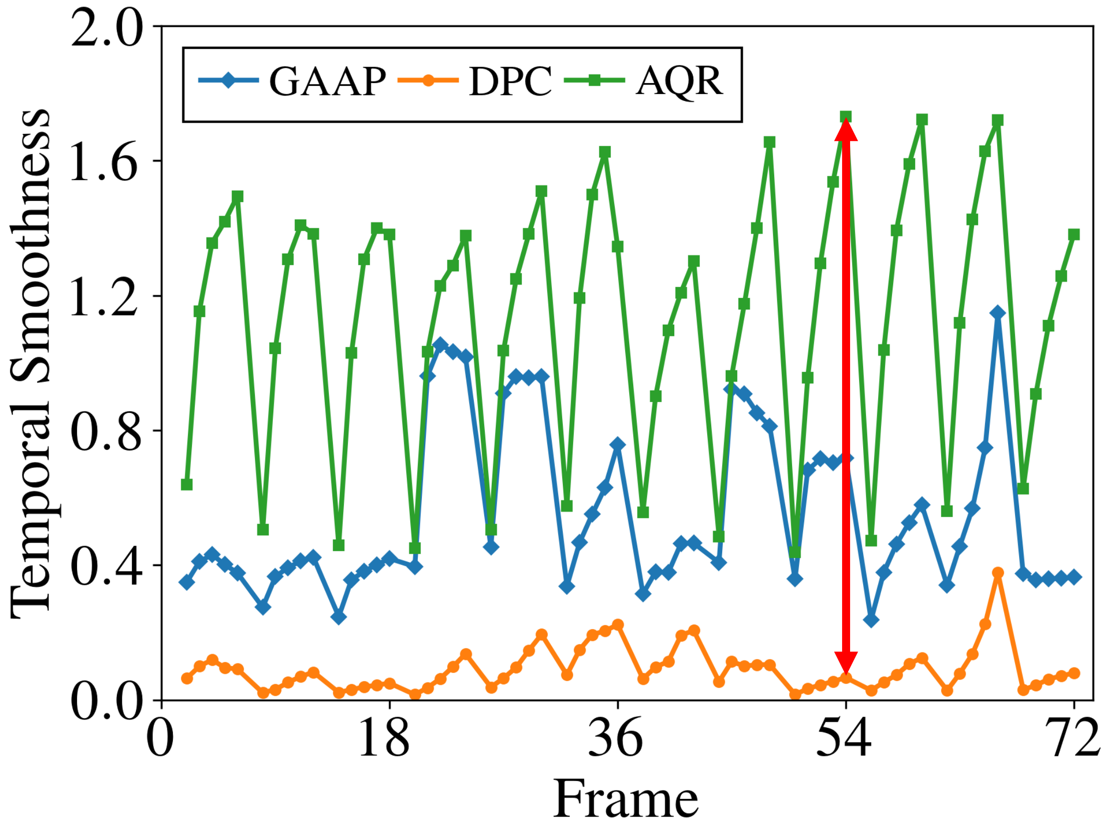
Y-PSNR: point-cloud color quality

Motion Vector Quality of Our Algorithms

-97.35%

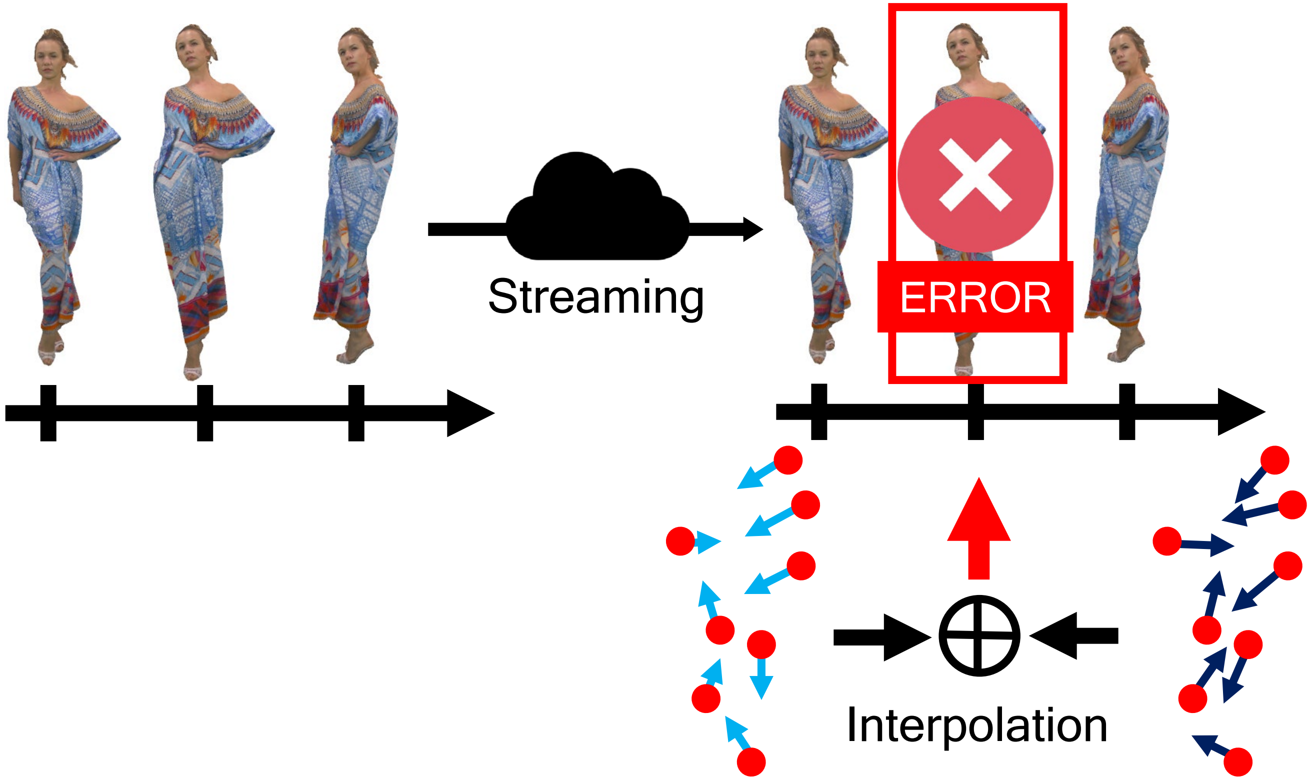


-93.83%



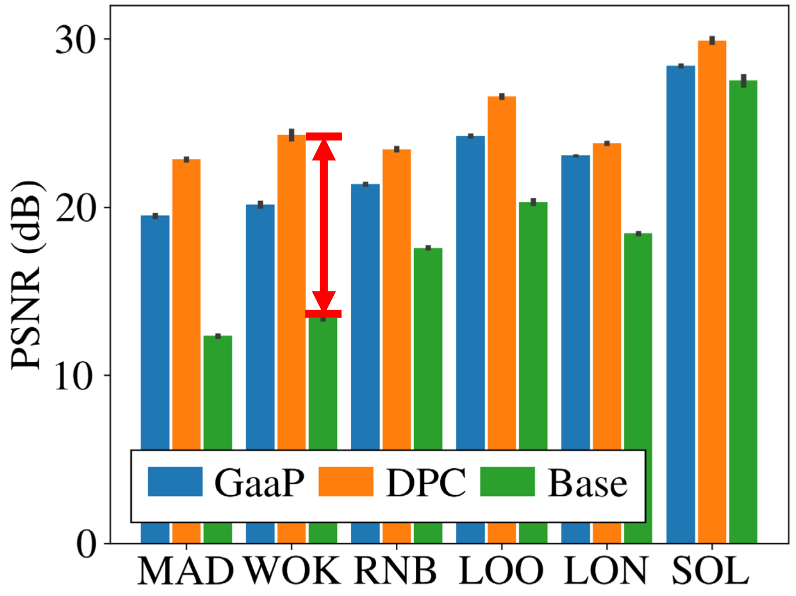
Smaller smoothness values indicate smoother motion vectors.

Application: Error Concealment



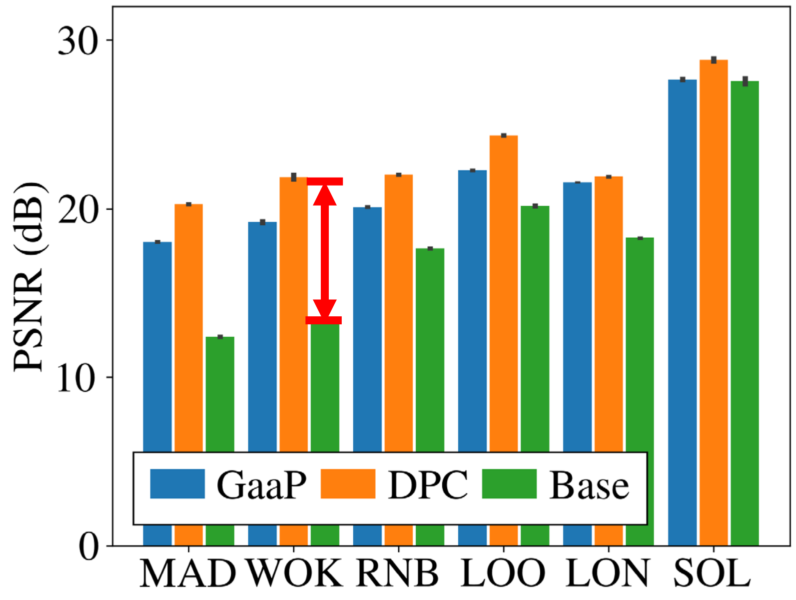
Results of Error Concealment

+10.90 dB



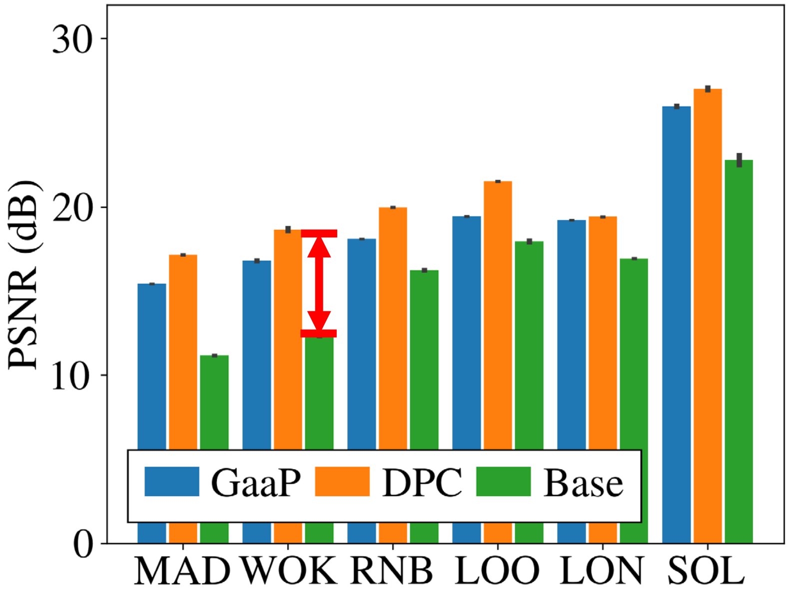
Drop 1 Frames

+8.52 dB



Drop 2 Frames

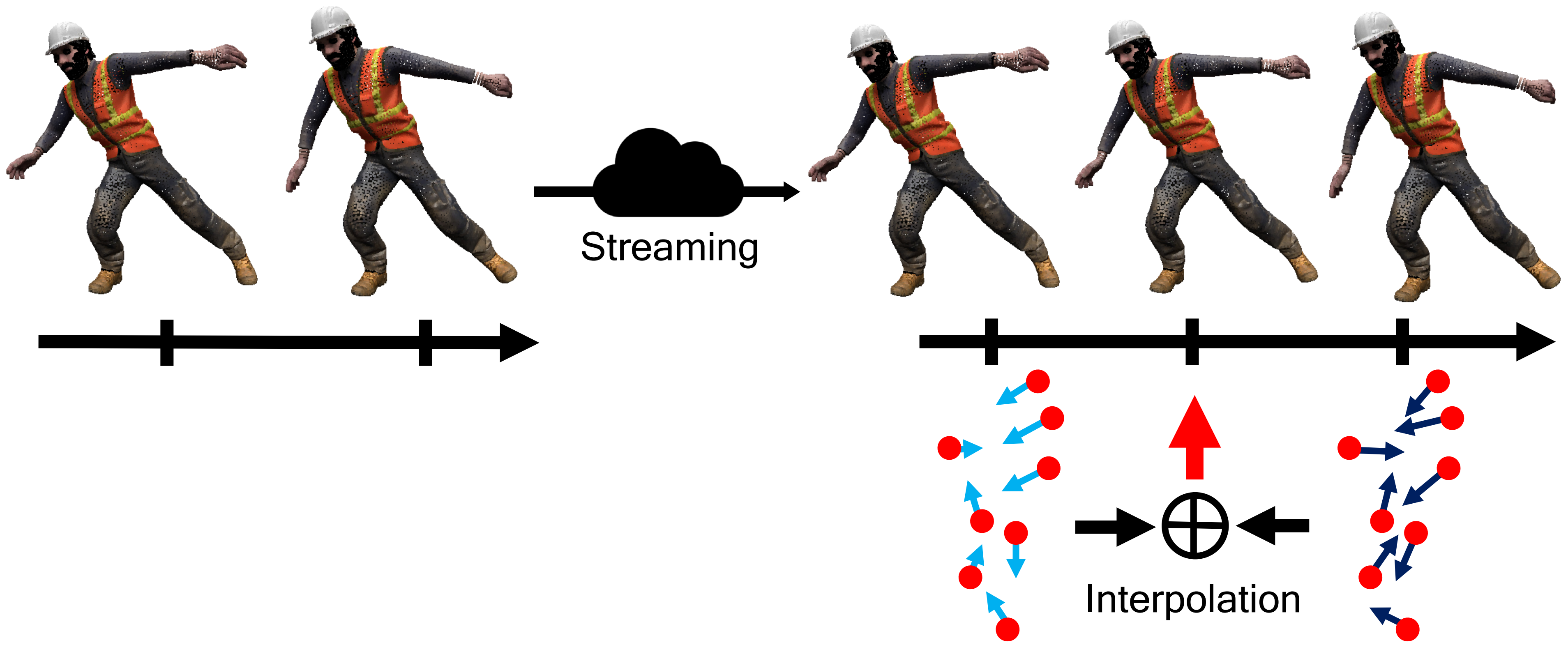
+6.33 dB



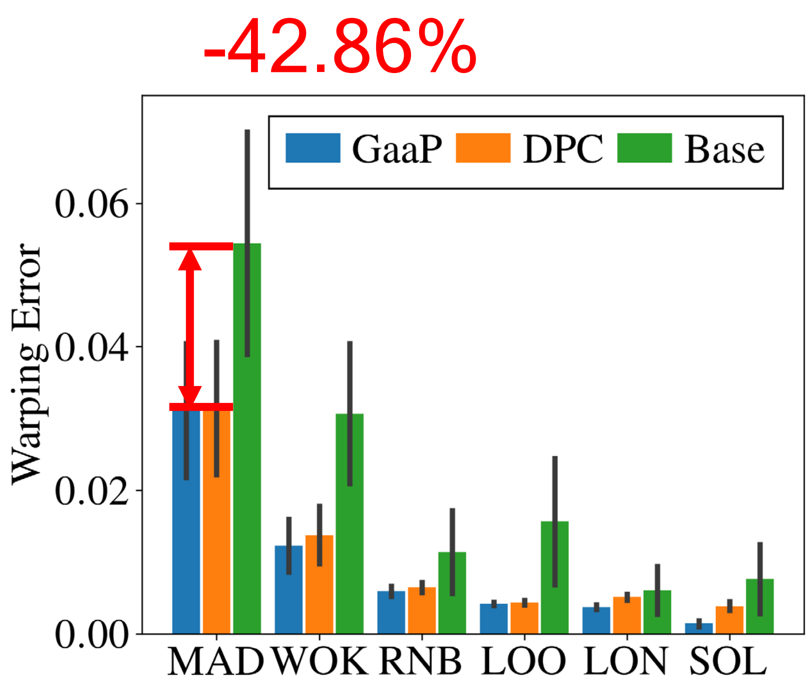
Drop 4 Frames

We outperform baseline in all situation

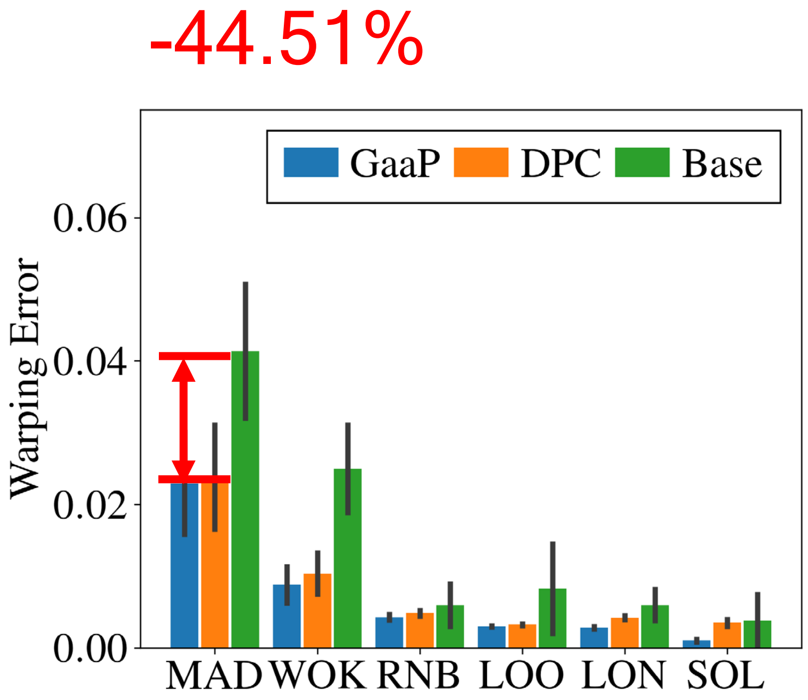
Application: Temporal Super-resolution



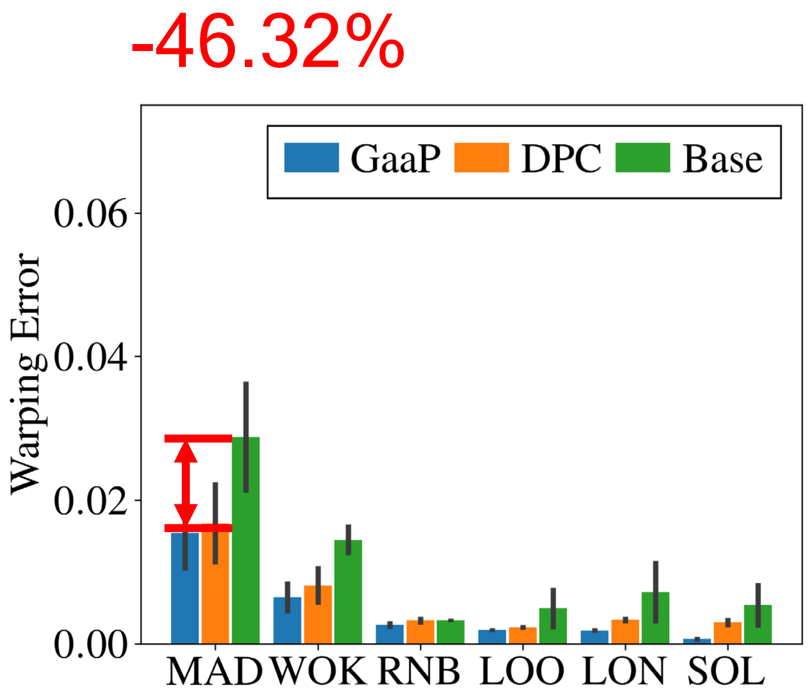
Results of Temporal Super-resolution



Insert 1 Frames



Insert 2 Frames

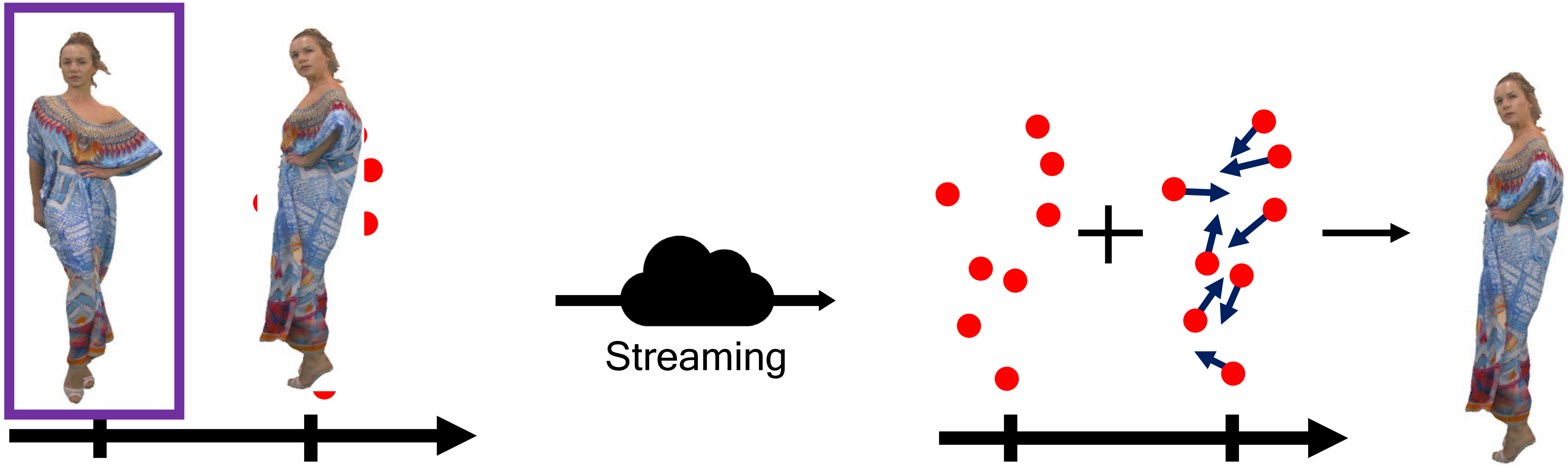


Insert 4 Frames

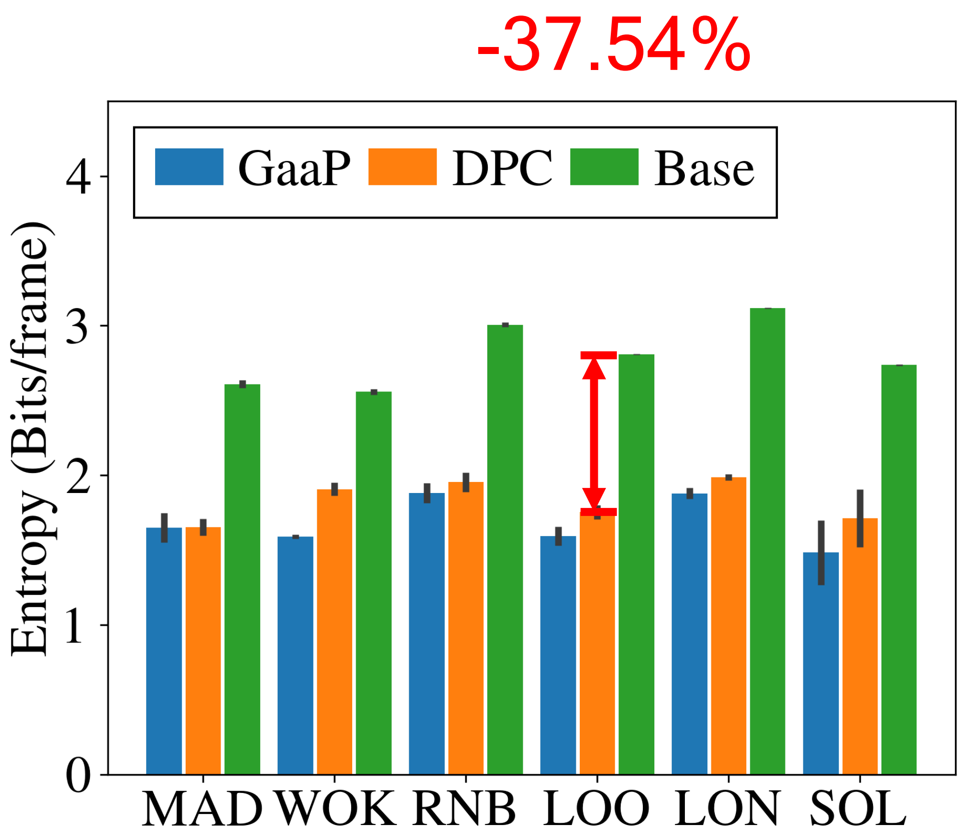
In Man Dance, our algorithms reduce warping error by about 45%

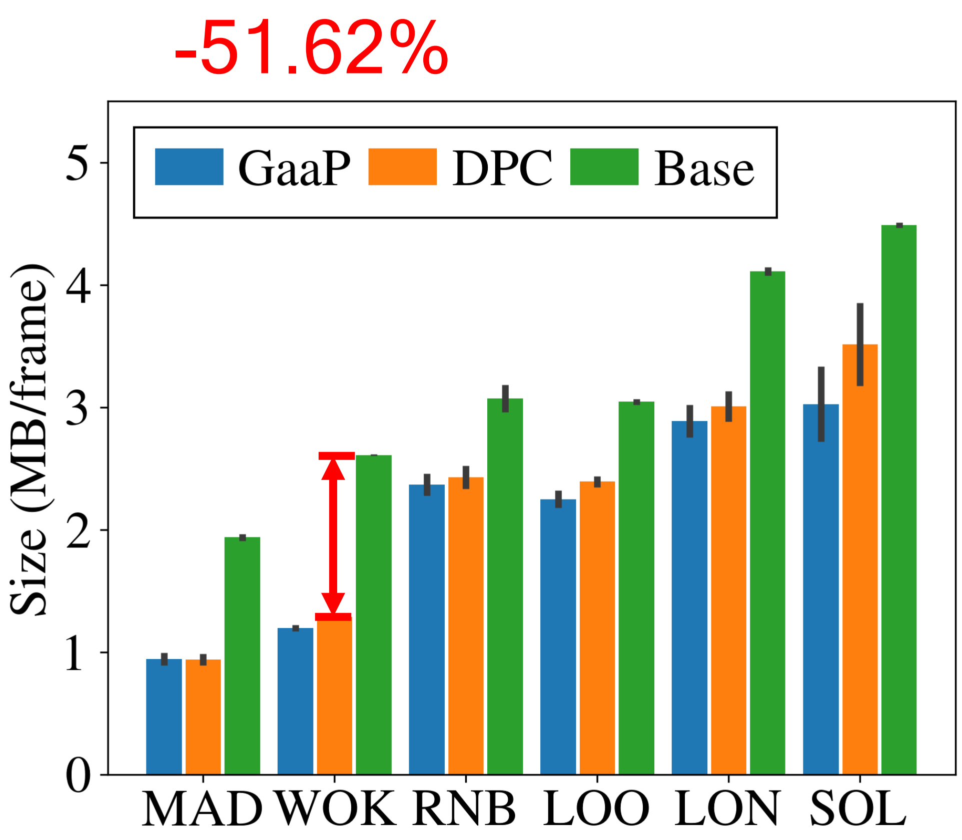
Source Coding

Key frame



Our Motion Vectors are more Compressible





We compute the entropy and use bzip for lossless compression

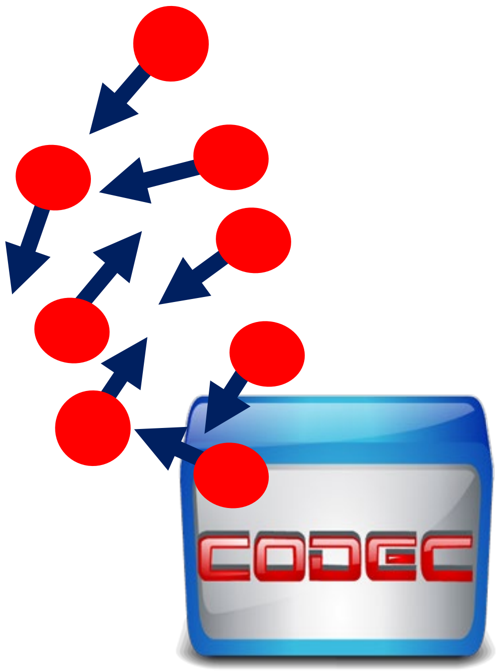
Outline

- Introduction
- Challenge
- Solution
- Algorithm: Gaussian as a Proxy (GaaP)
- Algorithm: Direct Point Cloud (DPC)
- Experiments
- Conclusion

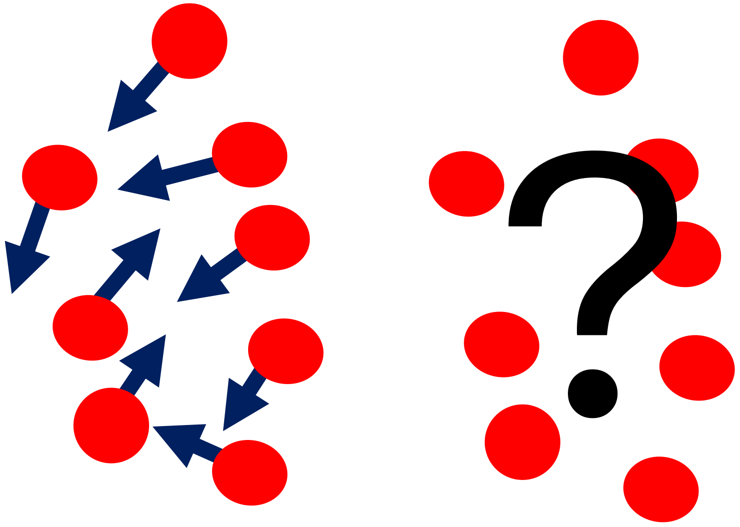
Conclusion

- We study a new paradigm that jointly learns non-key point cloud frames and point-wise motion vectors from dynamic point clouds
- We propose two algorithms: GaaP and DPC, both outperform the state-of-the-art heuristic algorithm
- Our learned motion vectors support error concealment, temporal super-resolution, and source coding

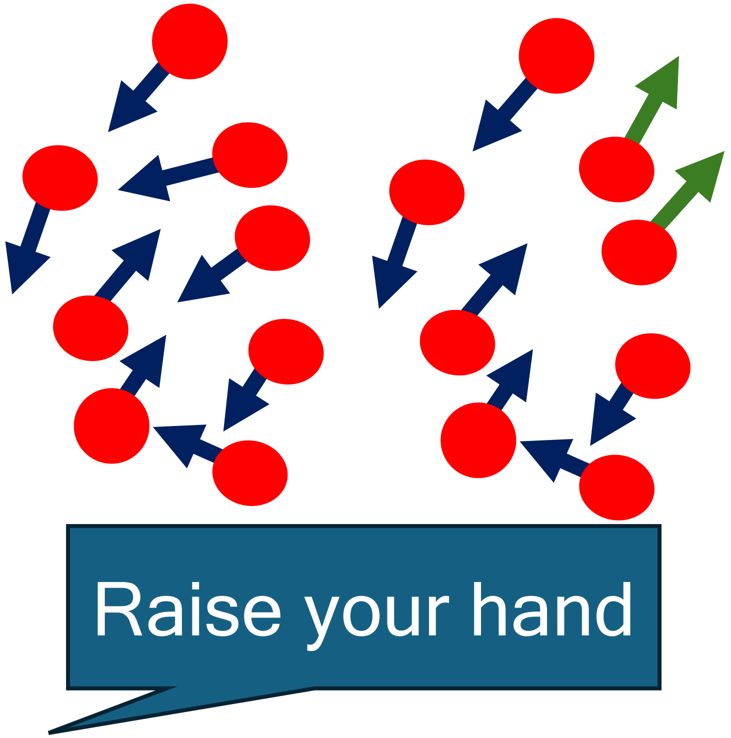
Future Work



Motion-Vector
Codec



Motion-Vector
Prediction



Motion-Guided
Content Editing

Thank you for listening