On Error Concealment of Dynamic 3D Point Cloud Streaming

Tzu-Kuan Hung (hungtzukuan@gmail.com)

Advisor: Cheng-Hsin Hsu

Networking and Multimedia Systems Lab,

CS, National Tsing Hua University

NMSL@NTHU

Networking and Multimedia Systems Lab

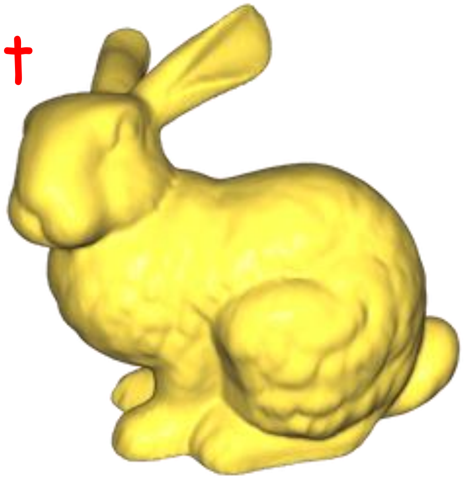
Outline

- Introduction
- Motivations
- Related Work
- Problem
- Solutions
- Experimental Setup
- Objective Results
- Subjective Results
- Conclusion
- Future Work

INTRODUCTION

3D Representations

Hard to edit
in real time

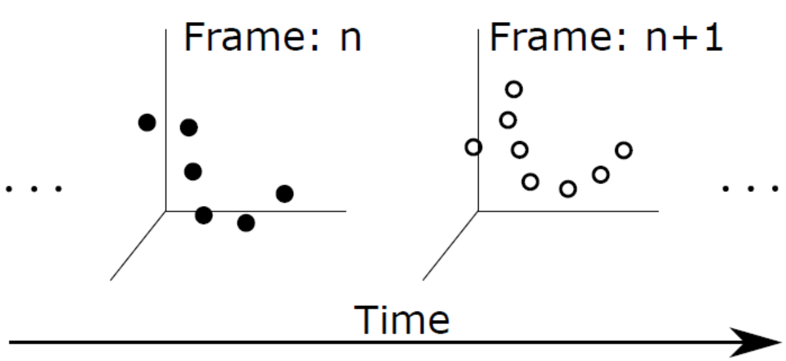


- Meshes
  - Points, edges, and faces
  - Not native output data types of any capturing sensors
- Point Clouds
  - Mandatory: 3D coordinates
  - Optional: attributes, such as colors
  - Native data format from some sensors
  - Light-weight data format
  - Applications:
    - Extended Reality (XR)
    - Entertainments
    - Teleconference
    - …….

Point Cloud Characteristics

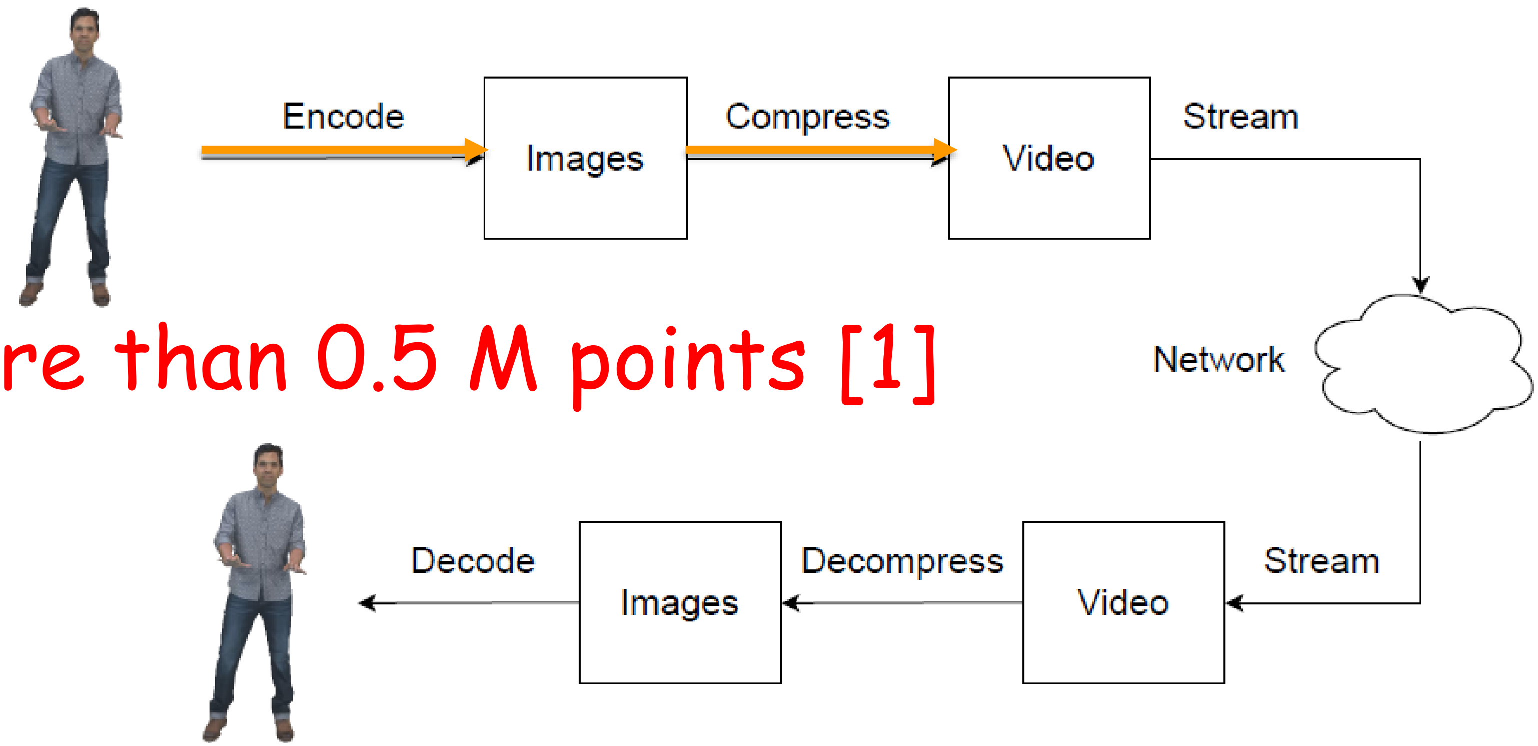
- No connectivity among points
  - No edge or face information
  - More points are needed compare to meshes
- Unordered
  - No specific order among points
  - No 1-1 matches among points across frames
- Heterogeneity
  - Sparseness levels
  - Optional attributes

Dense point clouds
with colors



MOTIVATIONS

Issue of Dynamic Point Cloud Streaming (1/3)



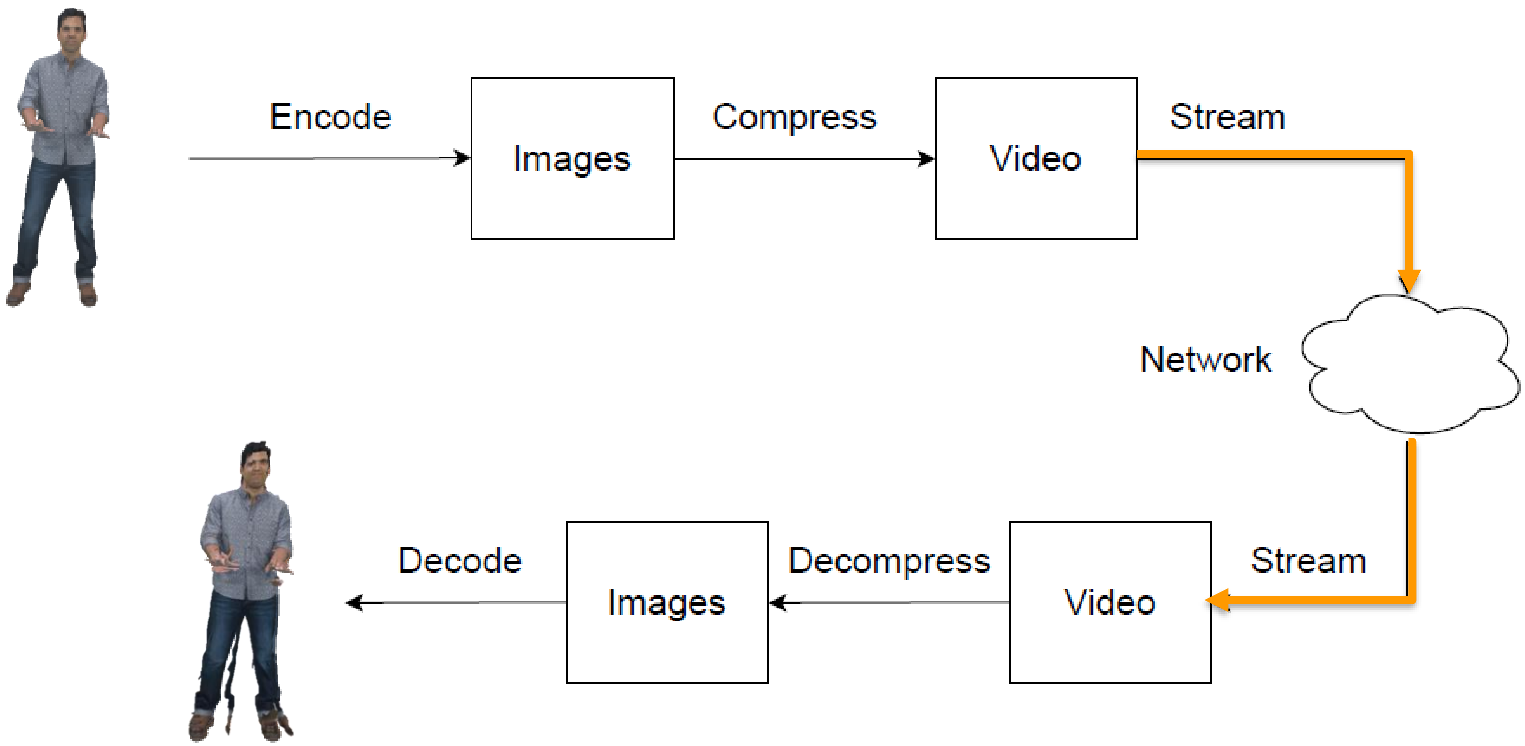
More than 0.5 M points [1]

- Streaming uncompressed dynamic point cloud dictates more than 4 Gbps

Compression before streaming is essential

[1] C. Cao, M. Preda, and T. Zaharia, "3D point cloud compression: A survey," ACM International Conference on 3D Web Technology (Web3D'19), pages 1–9, July 2019.

Issue of Dynamic Point Cloud Streaming (2/3)



- Lost or late packets of encoded bitstreams degrade visual quality

Issue of Dynamic Point Cloud Streaming (3/3)

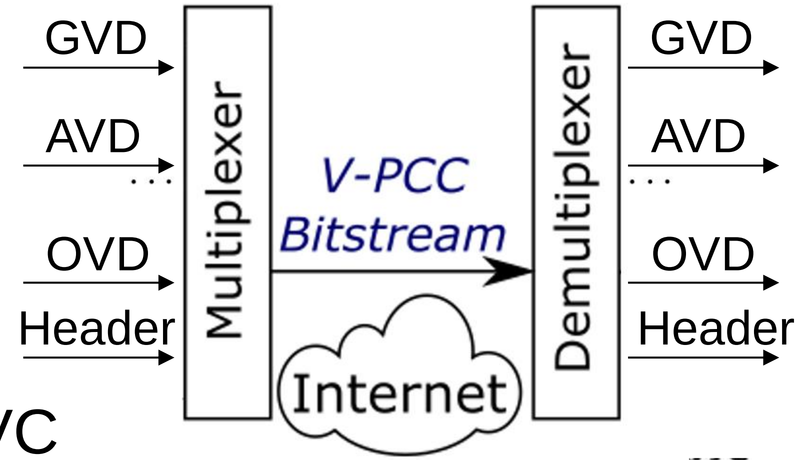
- Lost or late packets of encoded bitstreams degrade visual quality

That's why we need error concealment

RELATED WORK

MPEG Video-based Point Cloud Compression (V-PCC)

- V-PCC[2] Reference codec used in our work
  - Project each point cloud into:
    - Geometry (Near and Far map)
    - Attribute (Near and Far map)
    - Occupancy
    - Metadata and parameters
  - Encode sub-bitsream by HEVC



[2] MPEG 3DGC. V-PCC codec description v12. International Organization for Standardization Meeting Document ISO/IEC JTC1/SC29/WG7 MPEG/N0012, 2020. Meeting held online.

Error Concealment for 2D Videos

- Reduce the distortion by:
  - Frame copy
  - Temporal concealment
  - Spatial concealment

Can we apply them to V-PCC?

No! Patches are at different places[3]

[3] L. Li, Z. Li, V. Zakharchenko, J. Chen and H. Li, "Advanced 3D Motion Prediction for Video-Based Dynamic Point Cloud Compression," in *IEEE Transactions on Image Processing*, vol. 29, pp. 289-302, 2020, doi: 10.1109/TIP.2019.2931621.

Error Concealment for 3D Point Clouds

- Point cloud completion Not for streaming
  - Estimate the complete geometry of objects and scenes
  - Mostly by deep learning

Too slow!

- Inpainting[4]
  - Reduce cracks due to imperfect data acquisition
    - Self-similarity blocks
    - Inter-frame consistency
  - Computationally expensive
  - Not applicable to catastrophic distortion

[4] Wei Hu, Zeqing Fu, and Zongming Guo. 2019. Local frequency interpretation and non-local self-similarity on graph for point cloud inpainting. IEEE Transactions on Image Processing 28, 8 (2019), 4087–4100.

PROBLEM

Create V-PCC Loss Patterns

- Bitstreams consist of Network Abstraction Layer Units (NALUs)
  - Geometry Video Data (GVD)
  - Attribute Video Data (AVD)
  - Occupancy Video Data (OVD)
  - V-PCC header
  - V3C Parameter Set

  Headers
- Simulate packet loss
  - Encode 5 frames as a Group of Frame (GoF)
  - Mark NALUs of 3rd frame to drop
  - Overwrite NALUs with zeros
  - Decode corrupted bitstreams with V-PCC

Results from Loss Pattern

Pattern	I	P	S	I+P	I+S	P+S	I+P+S
O	C_G	-	-	-	-	-	-
G	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End
A	C_A	X	N	X	C_A	X	X
O+G	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End
O+A	C_G	X	N	X	C_G	X	X
G+A	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End
O+G+A	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End

I: Near map, P: Far map
S: Supplemental Enhancement Information (SEI)

- Outcomes
  - N: No clear visual impairment
  - C_A: Point cloud frame 3 is distorted in attributes only
  - C_G: Distorted in both geometry and attributes
  - C_G-End: 3-5 frames are distorted
  - X: Not decoded due to assertion errors of V-PCC

Concealment Strategies

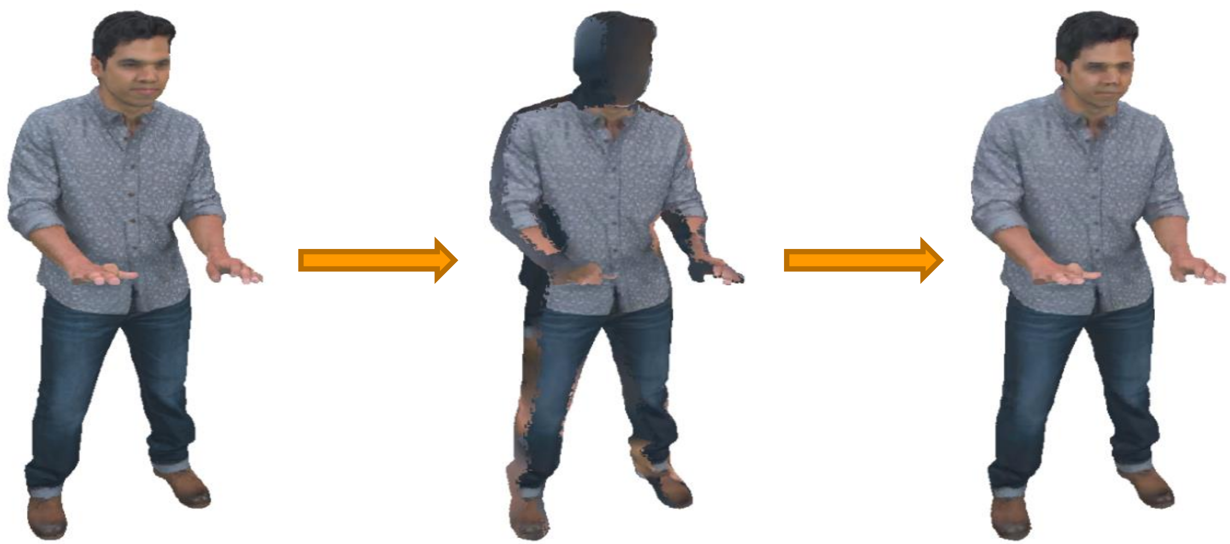
Pattern	I	P	S	I+P	I+S	P+S	I+P+S
O	C_G	-	-	-	-	-	-
G	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End
A	C_A	X	N	X	C_A	X	X
O+G	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End
O+A	C_G	X	N	X	C_G	X	X
G+A	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End
O+G+A	C_G	C_G-End	N	C_G-End	C_G	C_G-End	C_G-End

- Strategy
  - N: No concealment required
  - C_A: Attribute Concealment
  - C_G: Geometry Concealment
  - C_G-End: Geometry Concealment
  - X: Geometry Concealment

SOLUTIONS

Nearest Point (NP)

- Conceal point cloud frames without attribute (color) data only
- For each point in the current frame
  - Search for the closest point in the previous point cloud frame
  - Copy the attributes over



Error Concealment Schemes

- We propose a suite of error concealment algorithms for geometry distortion

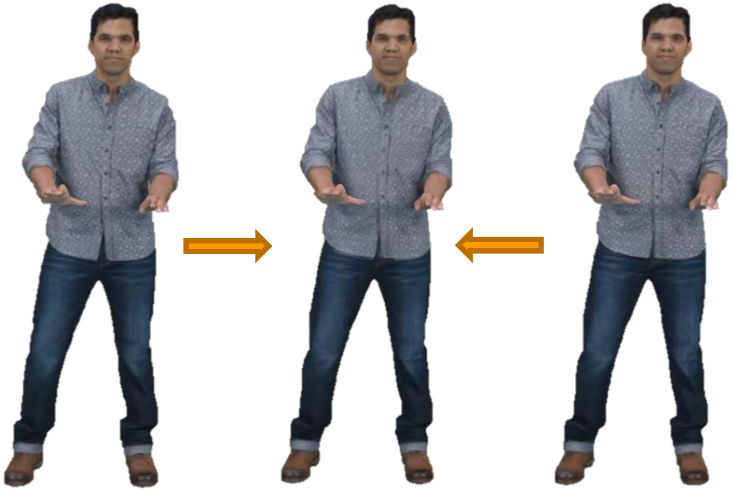
Name	f_1	f_2	Motion Estimation	f_2'	Matching (m)	Prediction $P(\cdot,\cdot)$
PI	Prev. frame	Next frame	-	$f_2' = f_2$	most similar point in f_2	interpolates between p_1 and $m(p_1)$
TI	Prev. frame	Next frame	-	$f_2' = f_2$	most similar triangle in f_2	interpolates between p_1 and $m(p_1)$
CMI	Prev. frame	Next frame	Cube-based motion	$f_2' = f_1 + M$	-	$f_3 = f_2'$
NCI	Prev. frame	Next frame	Cube-based motion	$f_2' = \Sigma_{i=1}^{27}(M_i/V_i) / \Sigma_{i=1}^{27}(1/V_i)$ where $M_i = (x_i, y_i, z_i)$, $V_i = \lvert x_i\rvert\times\lvert y_i\rvert\times\lvert z_i\rvert$	-	$f_3 = f_2'$

- Assume all geometry and attribute data are lost
  - Catastrophic distortion for decoded point clouds with V-PCC
  - Point-base (first 2) and cube-based (next 2) algorithms

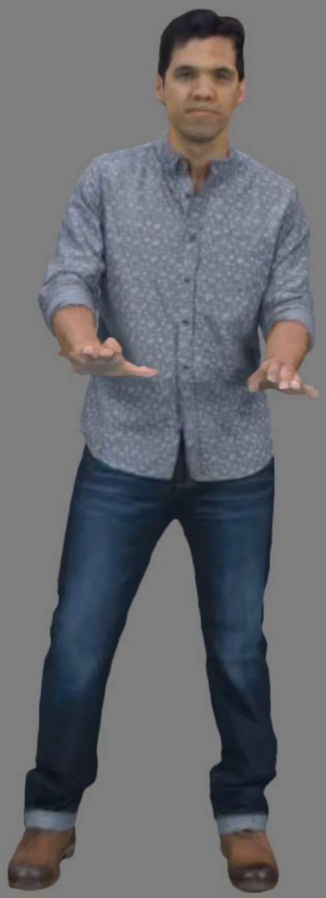
Point-to-Point Interpolation (PI)

- Conceal point cloud frames without geometry data
  - If geometry data are distorted or missing, the attribute data become useless
- For each point in the previous frame
  - Interpolate with the point in the future frame **within a specific radius**

$\Delta(p, q) = \alpha\Delta_g(p, q)+(1-\alpha)\Delta_a(p, q)$



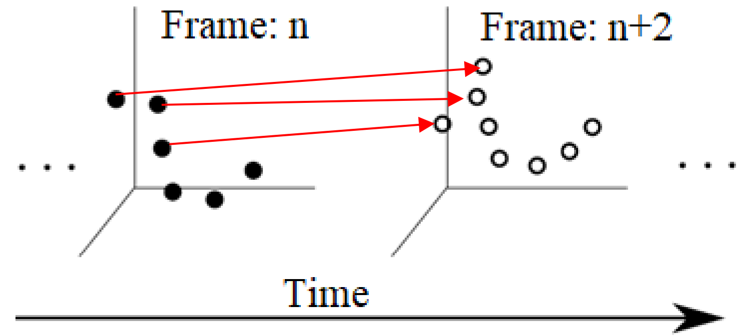
VPCC



PI

Triangular Interpolation (TI)

- Matching subroutine is done among triangles instead of points



PI (left)
TI (right)
Smaller cracks
No cracks

VPCC

TI

Cube-based Motion Interpolation (CMI)

- Divide point clouds into non-overlapped cubes with the same dimension
- Average all point-to-point outcomes within a cube for a **rigid motion vector of the whole cube**
- Enlarge cubes when gap happens
  - Let l be the length of each cube C
  - Dist. between every center to neighbor cube is exactly l
  - After interpolation, if dist. of centers between any adjacent cubes l ‘ > l, we enlarge length of the cube from l to l’

TI’s cracks

VPCC

CMI

Neighbor Cube-based Motion Interpolation (NCI)

- Use the same method to divide cubes and derive motion vectors for each cube
- Interpolate each point by inversely proportional to volume of vectors to 27-neighbors' centers
  - Get 27 vectors from each point to center of 27-neighbor cubes
  - Get volume of each vector (x_i, y_i, z_i) by ($|x_i|$ * $|y_i|$ * $|z_i|$)
  - Weighted sum by inverse volume of each vector

CMI's extrusion

VPCC

NCI

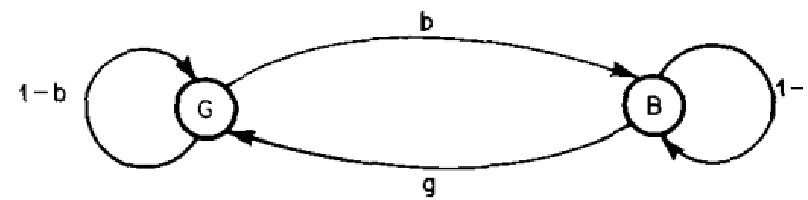
EXPERIMENTAL SETUP

Experimental Setup

- Datasets
  - MPEG dynamic 3D point cloud sequences

	Queen	Loot	Red&Blk	Soldier	LongDress	Basketball	Dancer
Cplx.	Low	Low	Low	Low	Medium	High	High
Pt.#	1.00 M	0.78	0.70	1.50	0.80	2.90	2.60

- Gilbert-Elliot Models[5] parameters
  - 5%, 10%, 15%



- Baseline
  - 2D frame copy (2DFC): naive frame copy mechanism by V-PCC codec
  - 3D frame copy (3DFC): copy the nearest undistorted frame over

[5] M. Mushkin and I. Bar-David, "Capacity and coding for the gilbert-elliott channels," in IEEE Transactions on Information Theory, vol. 35, no. 6, pp. 1277-1290, Nov. 1989, doi: 10.1109/18.45284.

Performance Metrics

- 3D Visual Metrics
  - **GPSNR** - The PSNR of Chamfer distance between pair-wise closest points in the target and ref. frames
  - **Hausdorff distance**: The maximal shortest distance between the points in the target and ref. frames The lower the better
  - **CPSNR:** The luminance component of color distortion between the nearest points in the target and ref. frames
- 2D Visual Metrics
  - **PSNR:** The PSNR of the foreground object (avatar) only
  - **SSIM:** The luminance SSIM of the foreground object only
  - **VMAF:** Predicts subjective video quality consider the whole video sequences
- Running time The lower the better

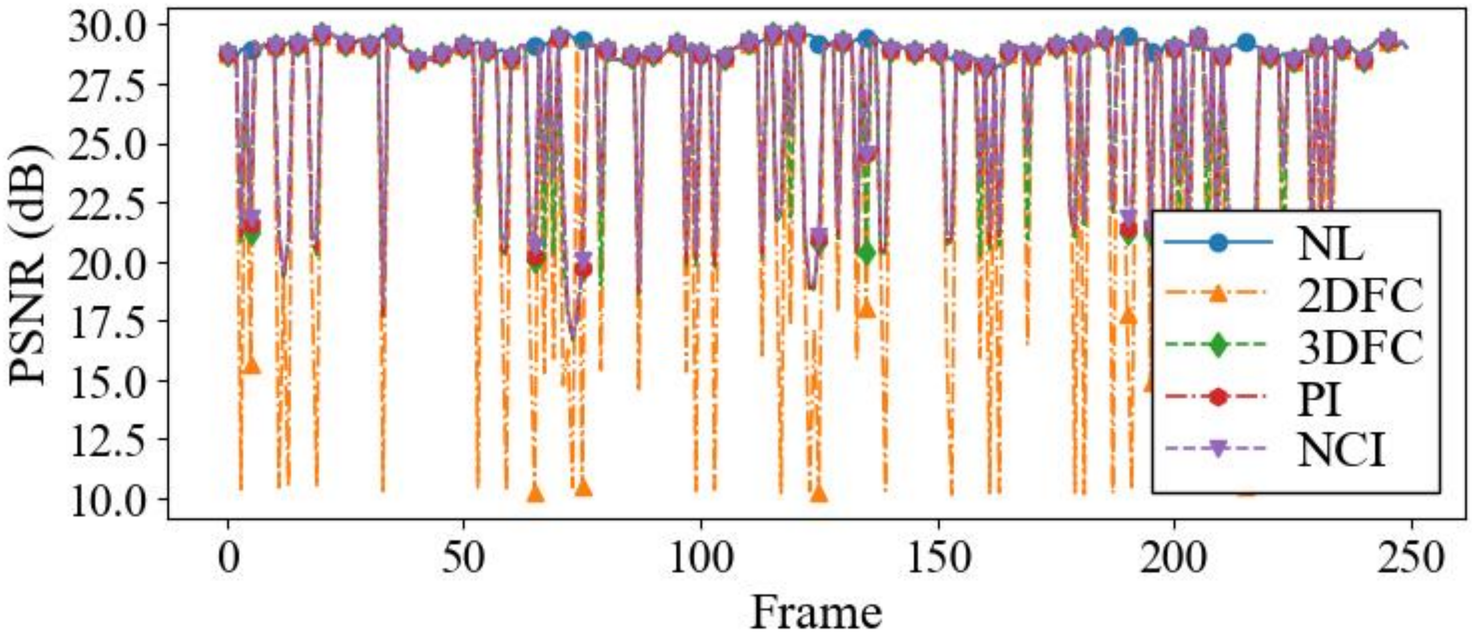
OBJECTIVE RESULTS

Per-Frame Line Figure

10% lost

- Key observations:
  - NCI > PI > 3DFC > 2DFC in PSNR
  - the quality drops as high as 12 dB in PSNR

Limitations of the current 2DFC method

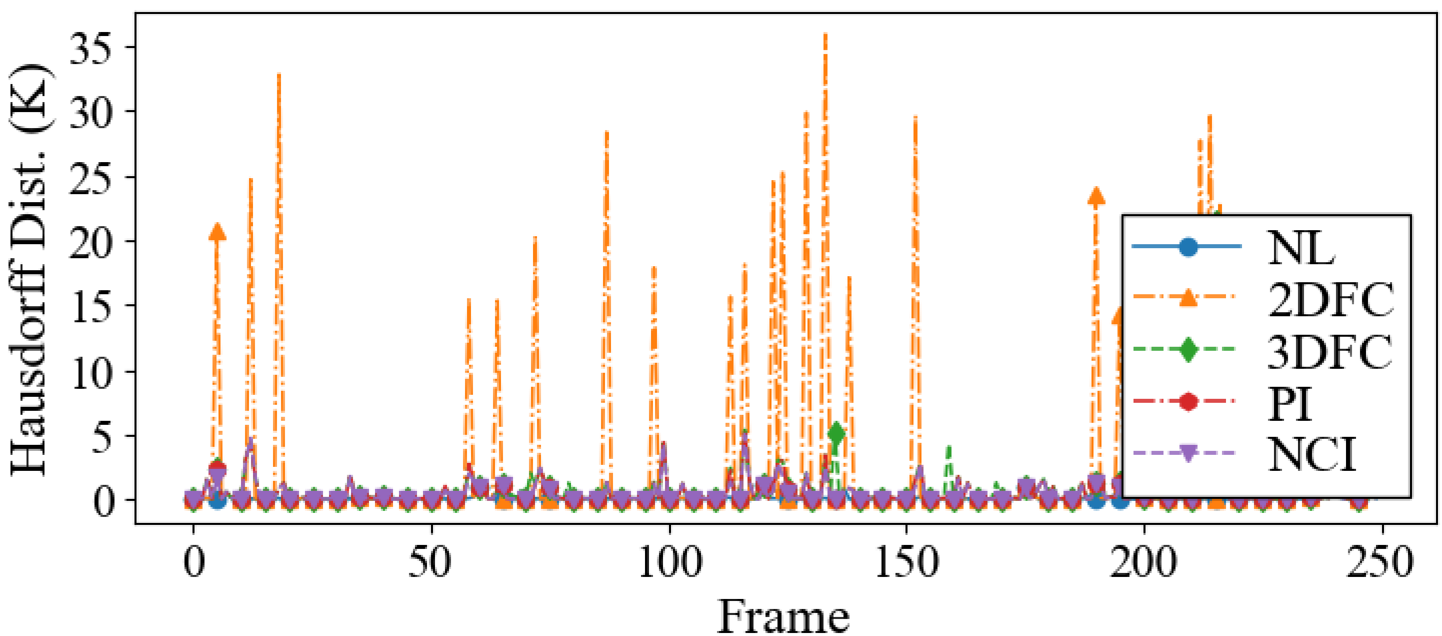


Per-Frame Line Figure

10% lost

- Key observations:
  - 2DFC > 3DFC > PI > NCI in Hausdorff distance
  - the quality surges as high as 35K in Hausdorff distance

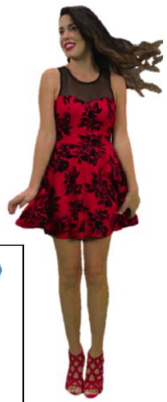
Limitations of the current 2DFC method

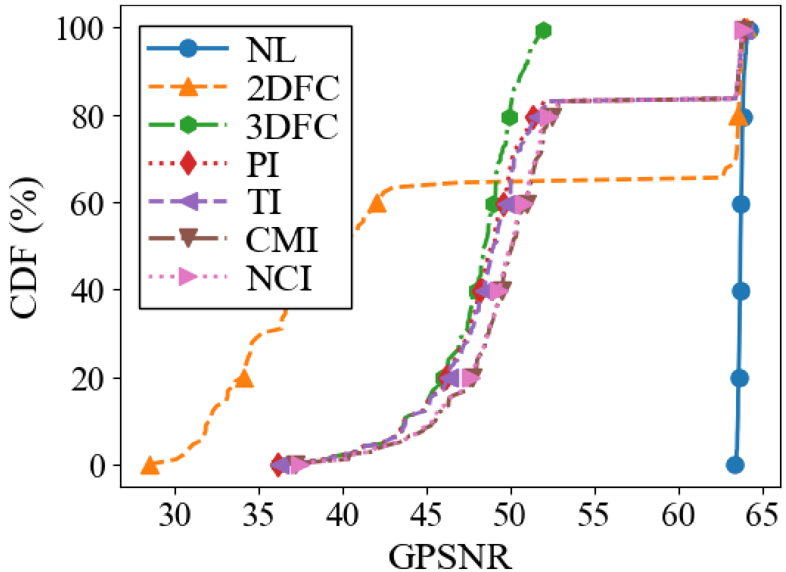


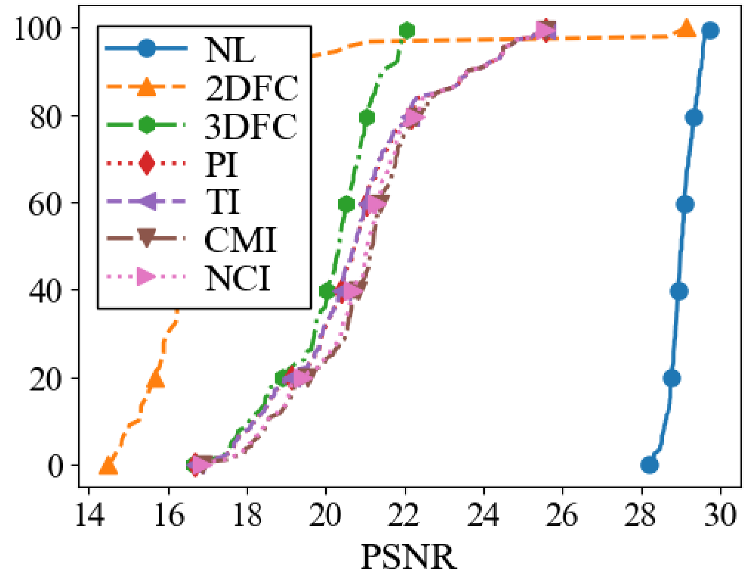
Cumulative Distribution Function (CDF)

- Key obserbations:
  - 2DFC still results low in as low as 30 dB in GPSNR
  - Clustered into (3DFC), (PI, TI), and (CMI, NCI)
  - 20% best performing of (CMI, NCI) is 52^{+} dB
  - 20% best performing of (3DFC) is 49^{+} dB

10% lost



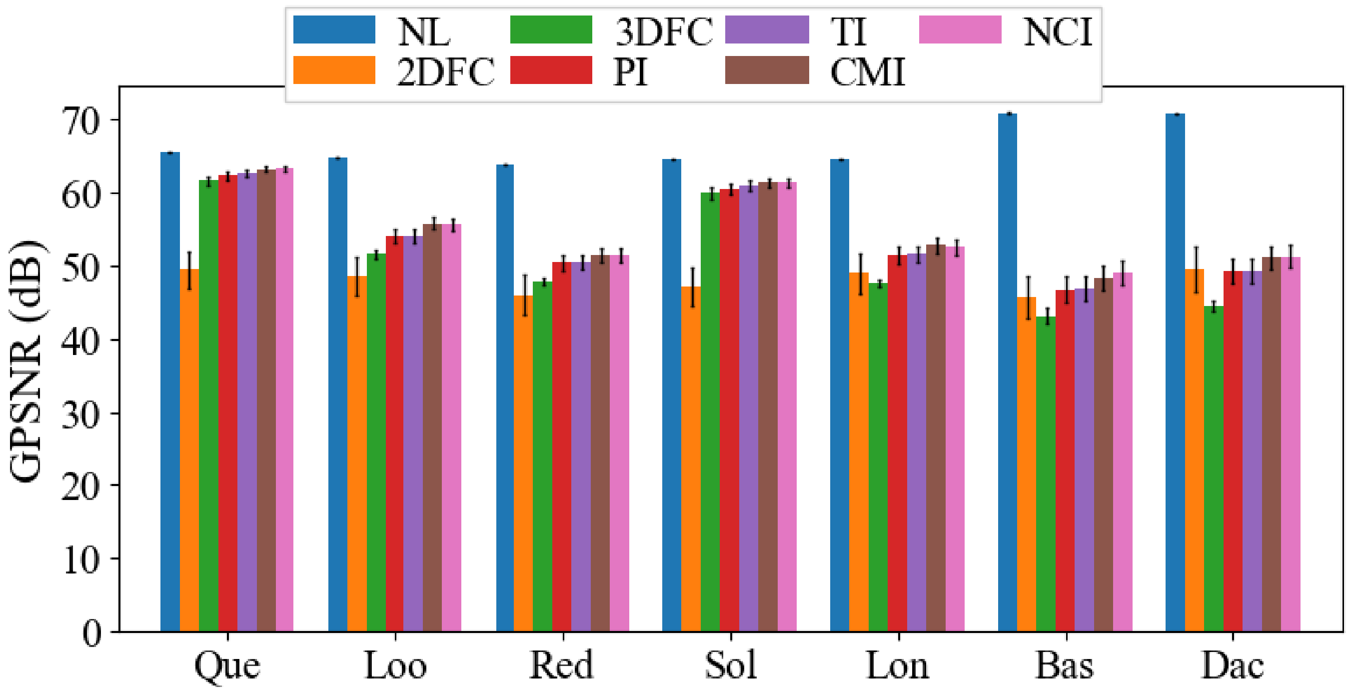




Overall Quality of GPSNR

- Our algorithms always outperform 3DFC
- CMI and NCI consistently outperform others

Best: +7 dB in Dancer



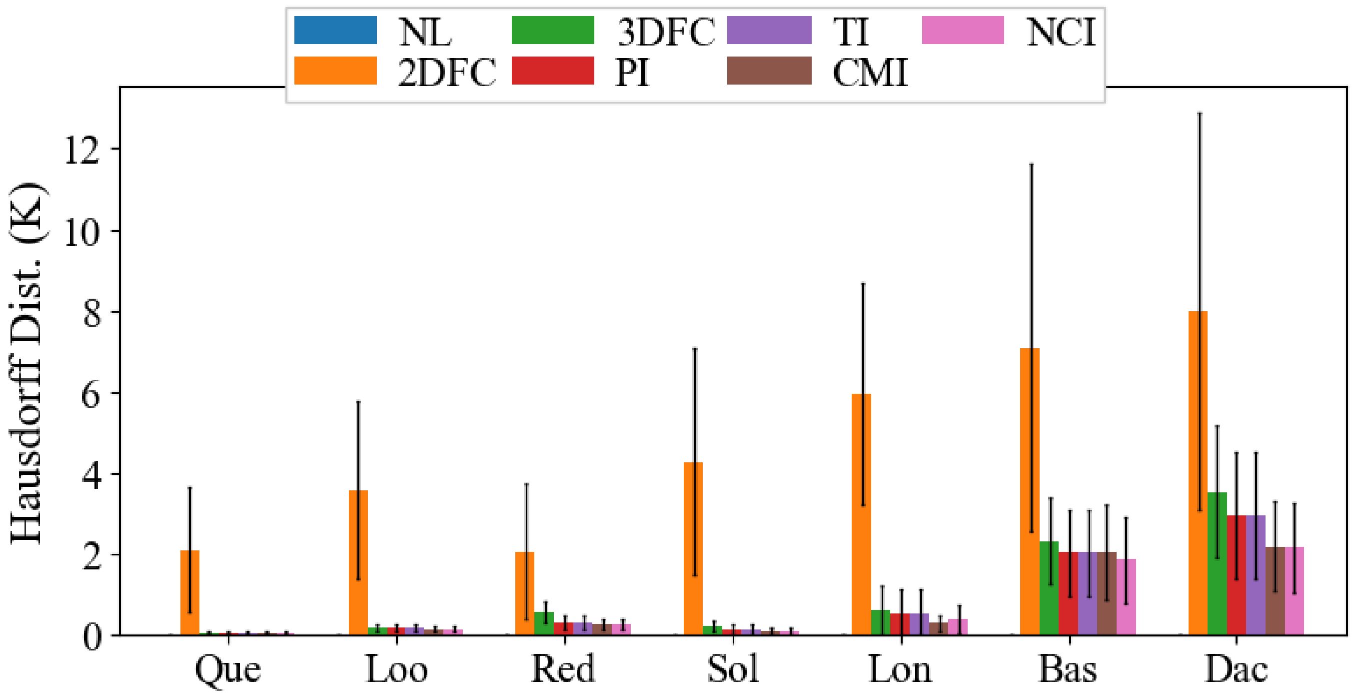
10% lost

Overall Quality of Hausdorff Distance

- Our algorithms always outperform 3DFC
- CMI and NCI consistently outperform others

Best: -1.5 K in Dancer

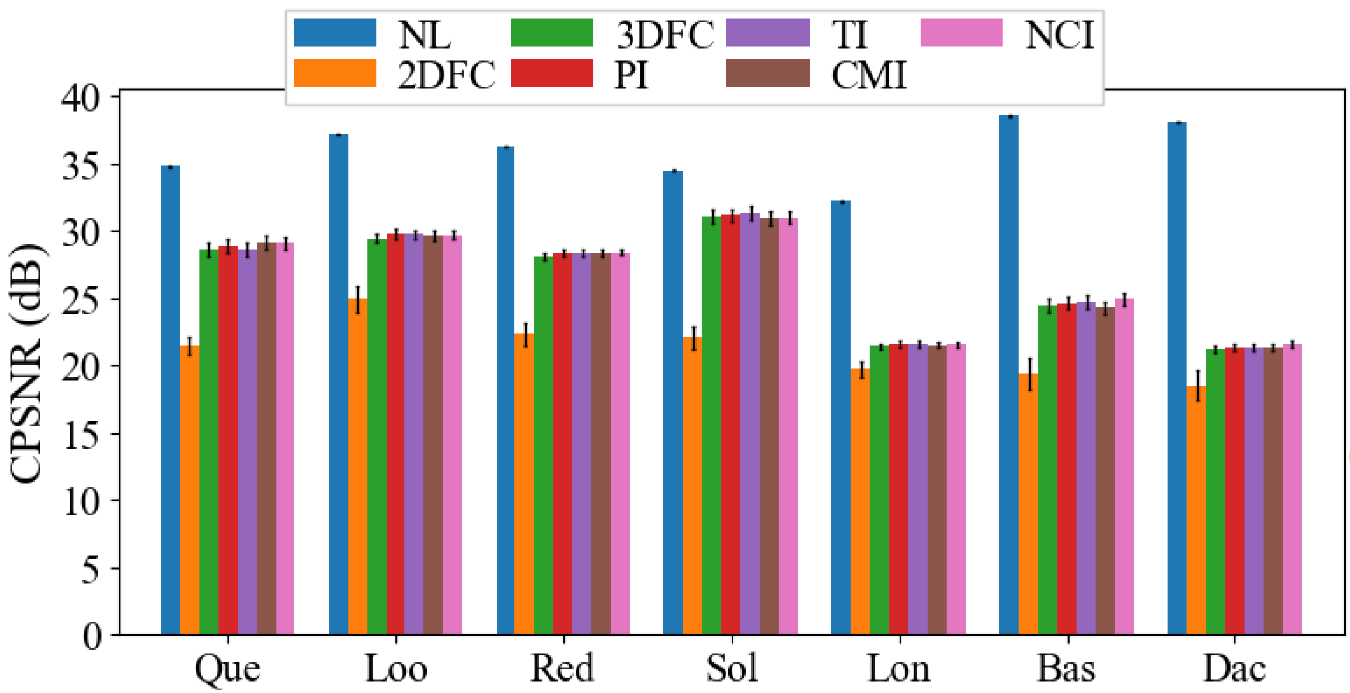
10% lost



Overall Quality of CPSNR

- Our algorithms may not outperform others
- CMI and NCI may not outperform others

Best: +1 dB in Queen Worst: Soldier and Basket

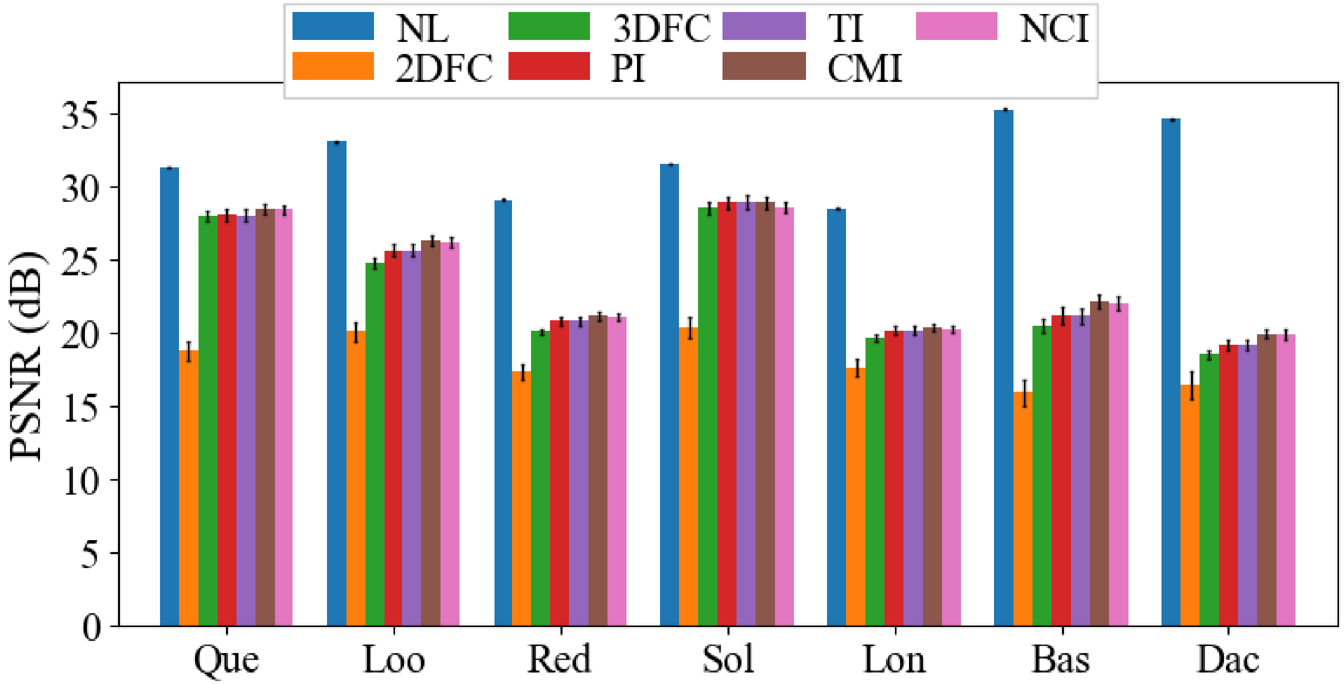


10% lost

Overall Quality of PSNR

- Our algorithms outperform 3DFC in PSNR in most cases
- NCI may not outperform others

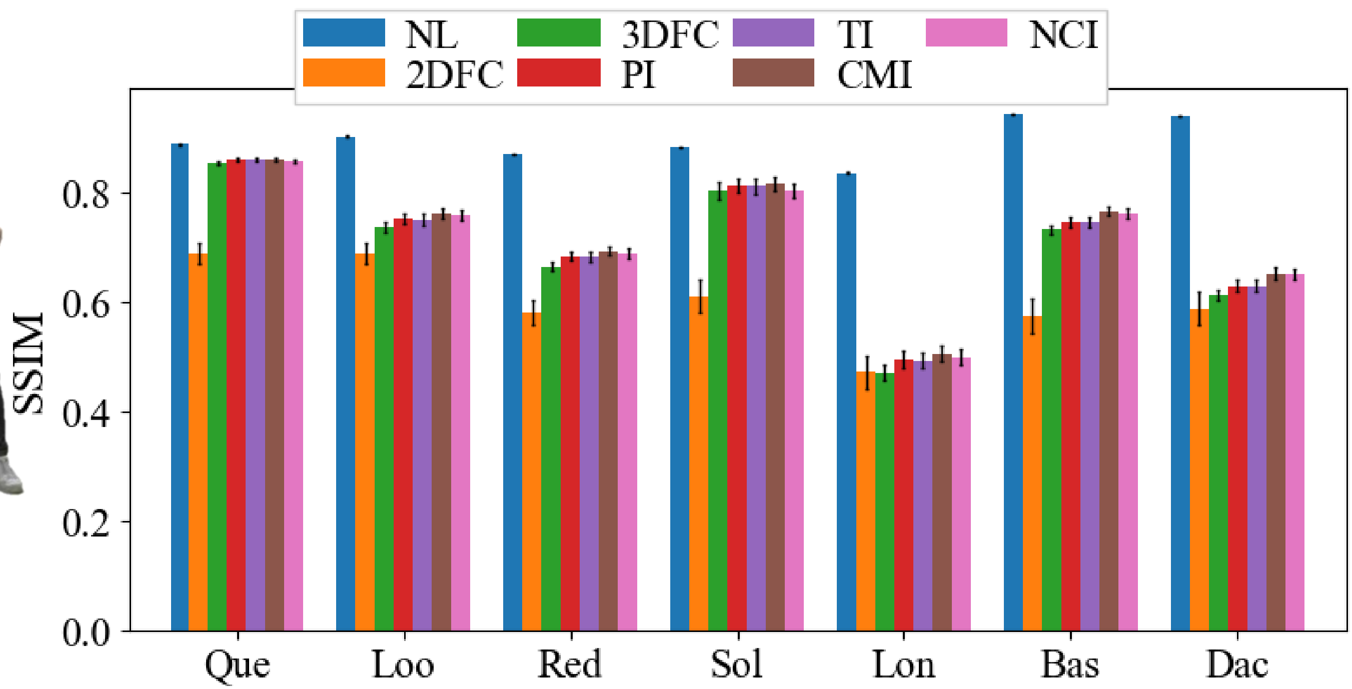
Best: +2 dB in Loot Worst: Soldier

Overall Quality of SSIM

- Our algorithms outperform 3DFC in SSIM in most cases
- NCI may not outperform others

Best: +0.05 dB in Dancer   Worst: Soldier

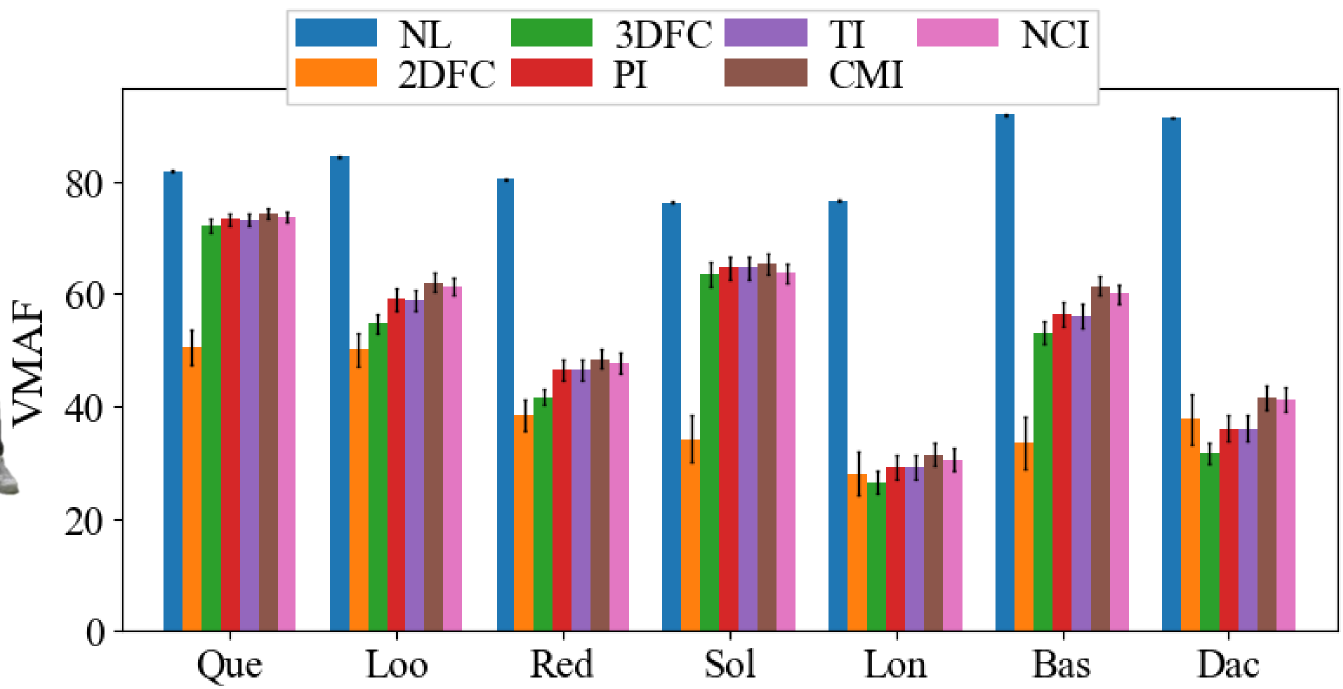


Overall Quality of VMAF

- Our algorithms outperform 3DFC in VMAF in most cases
- NCI may not outperform others

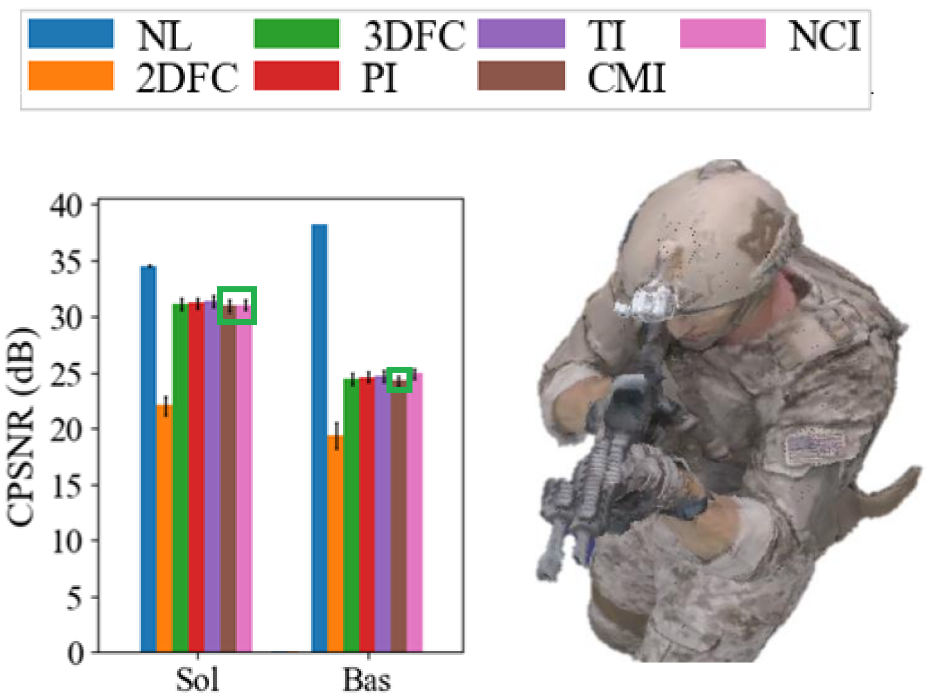
Best: +9 in Dancer

Worst: Soldier



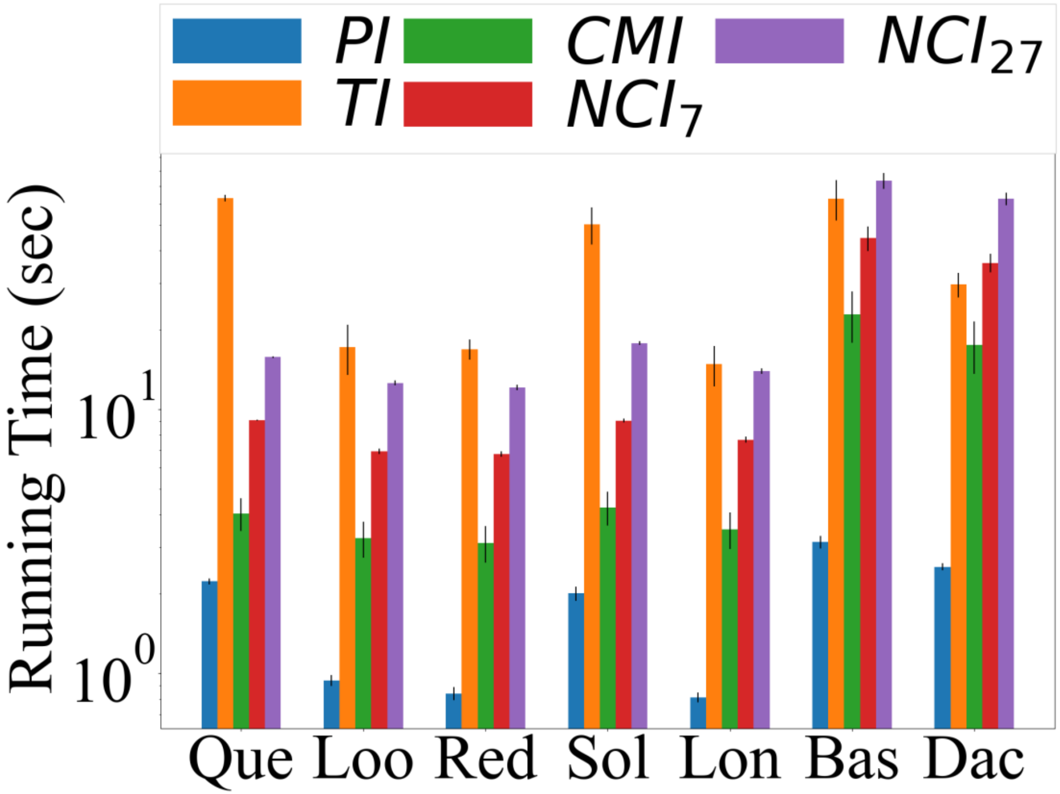
Sequences with Inferior Quality

- Example of artifacts from **Soldier** and **Basketball** sequences with CMI and NCI algorithms



Per-frame Running Time

- Select 24 random point cloud frames from total 250 frames
- PI run the fastest
- CMI and NCI runs slower on high-complexity sequences
- NCI runs slower on Dancer and Basketball player sequences
- Absolute running time is still long

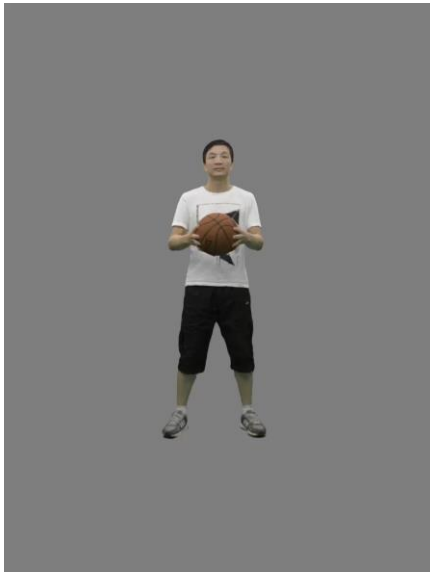


SUBJECTIVE RESULTS

Experimental Setup

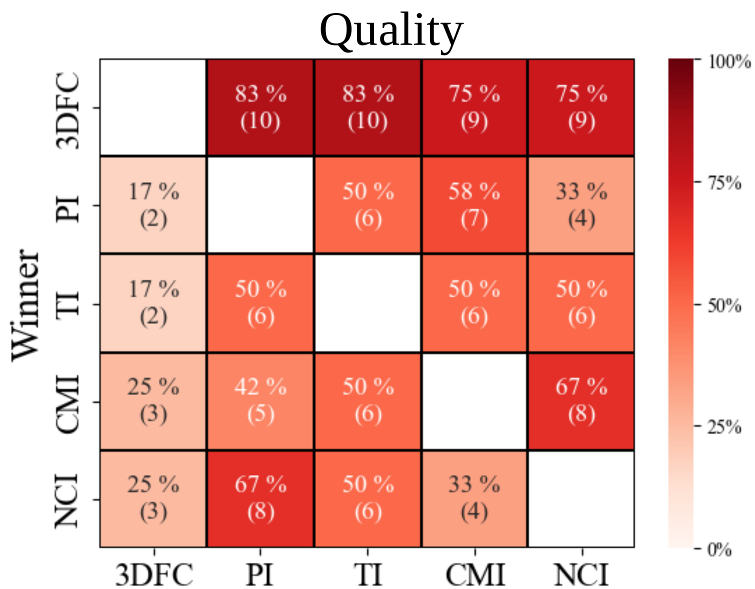
- Head-to-head video comparison
- No. subjects: 12
- Three questions:
  - *Which video was smoother?*
  - *Which video had better image quality?*
  - *Which video did you prefer?*
- Derive head-to-head comparison to MOS between 0 to 1
  - Transform by Plackett-Luce model[6]
  - Normalize to [0, 1]

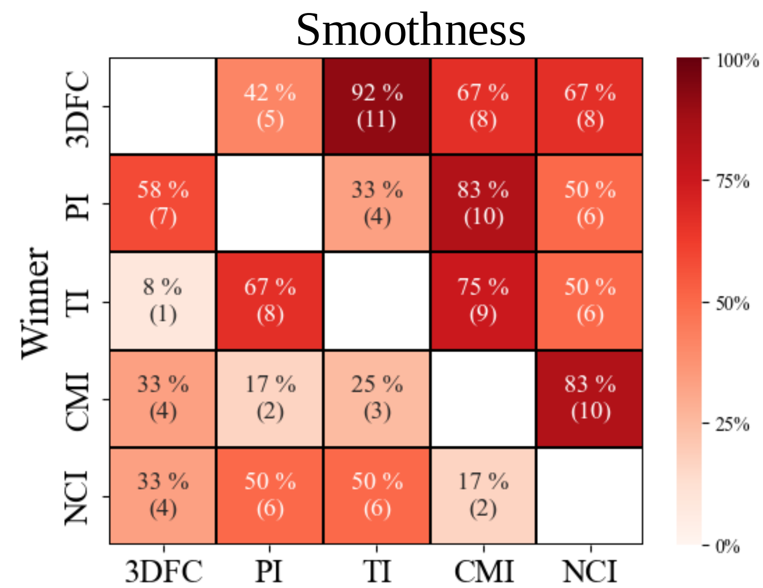
250 frames, 20 fps

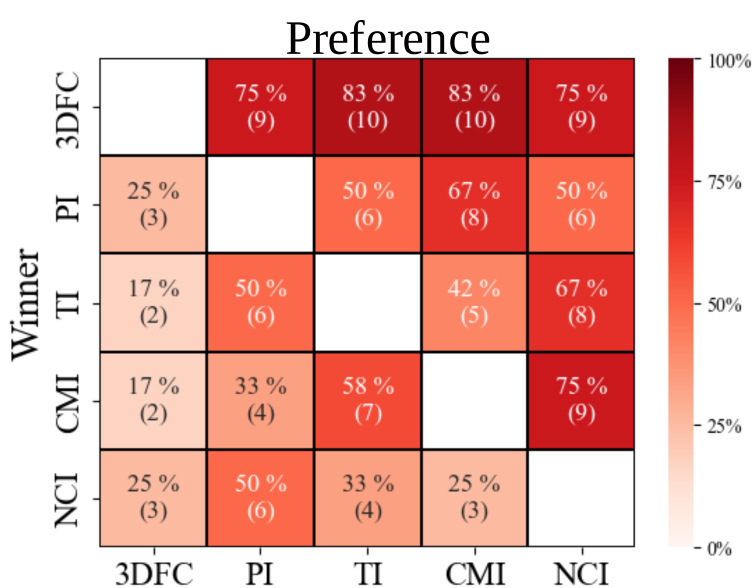


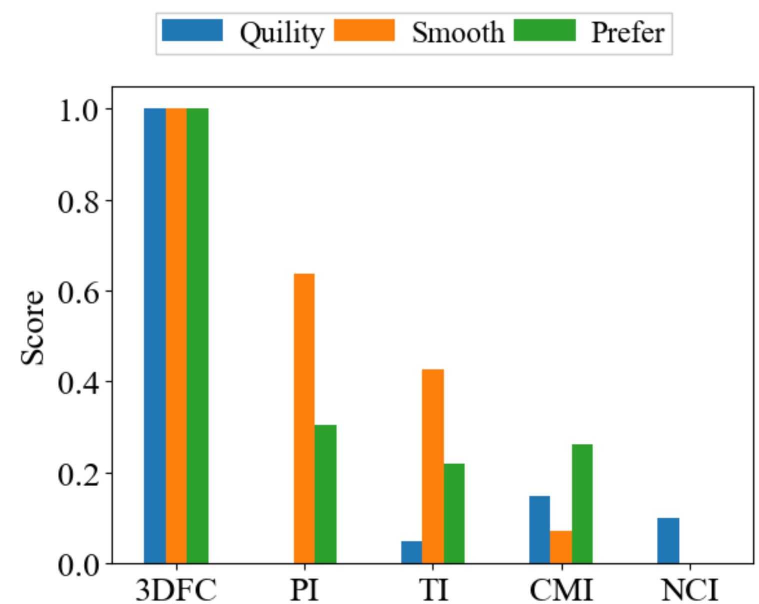
[6] H. L. Turner, J. van Etten, D. Firth, and I. Kosmidis, "Modelling rankings in R: The PlackettLuce package," Computational Statistics, pp. 1–31, 2020.

Subjective Results - Basketball

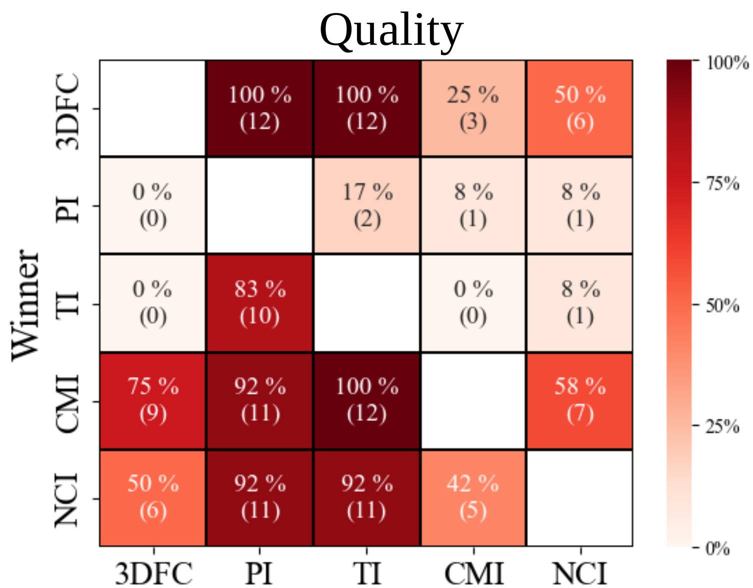


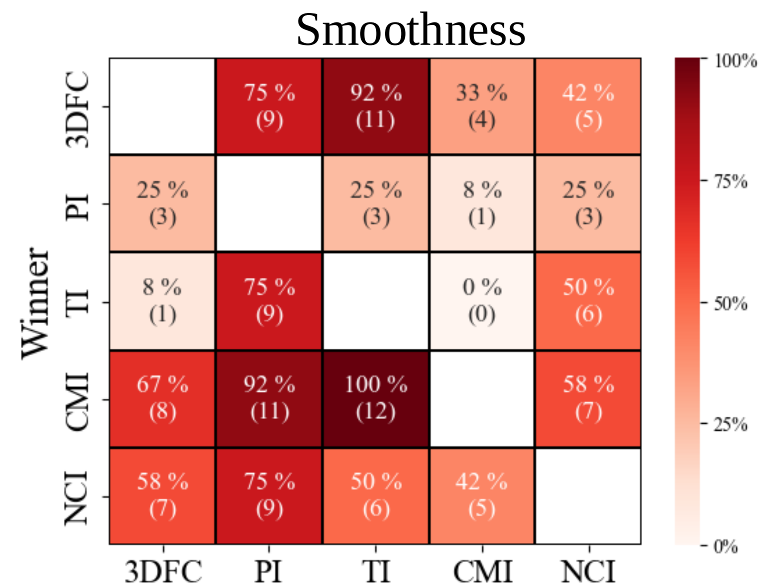


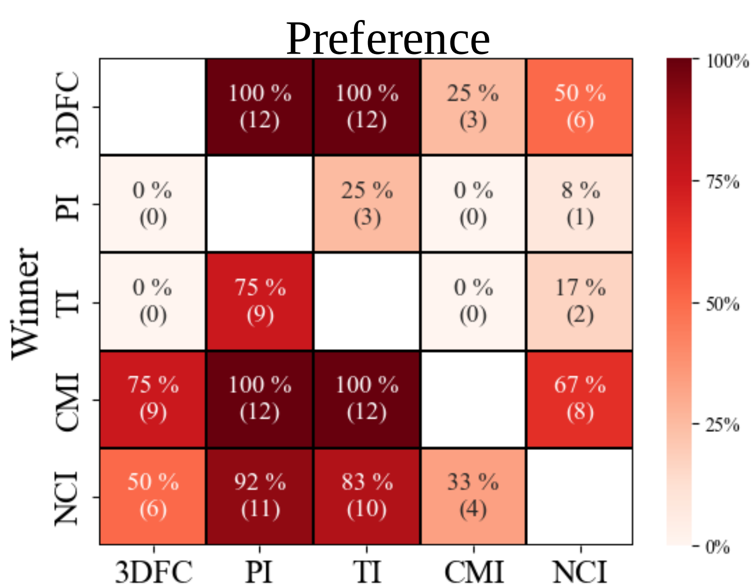


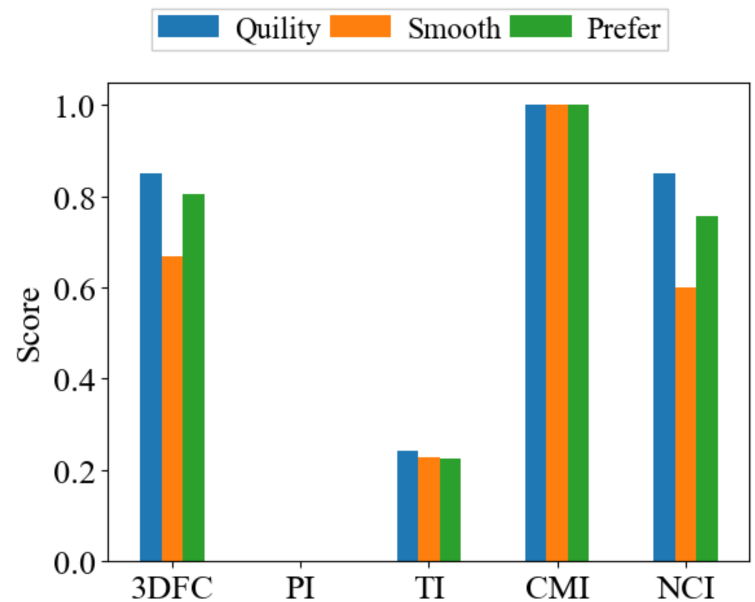


Subjective Results - Queen

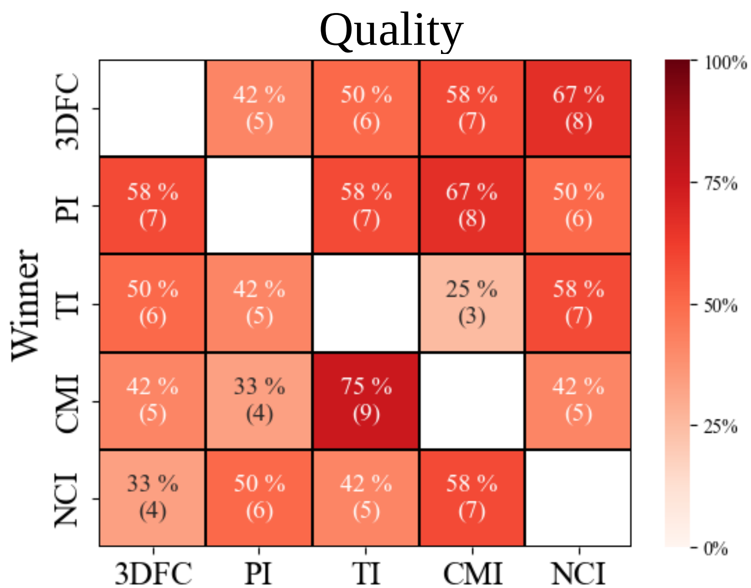


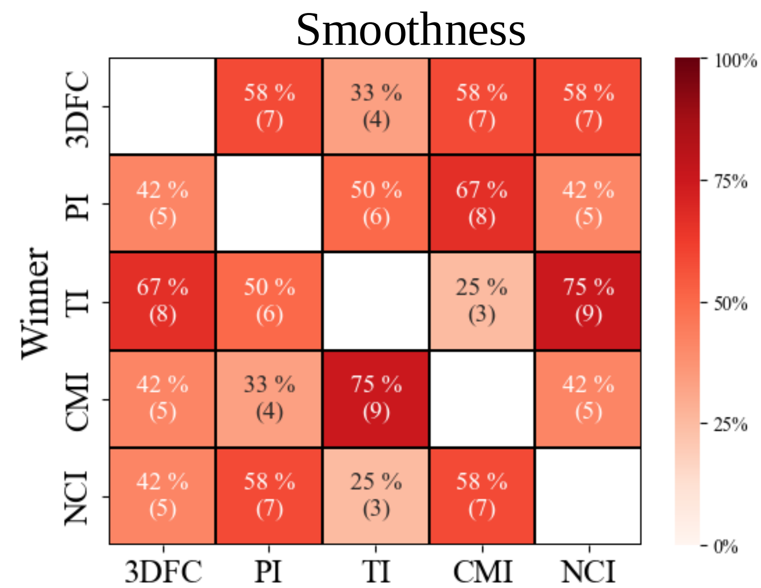


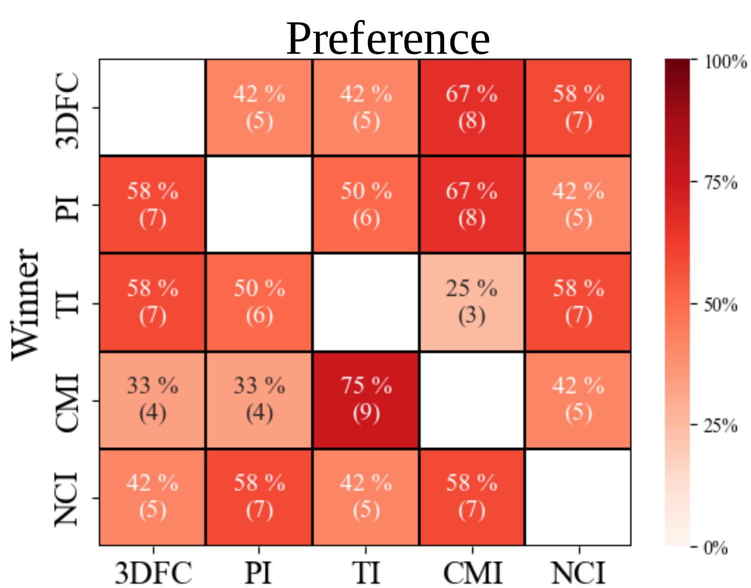


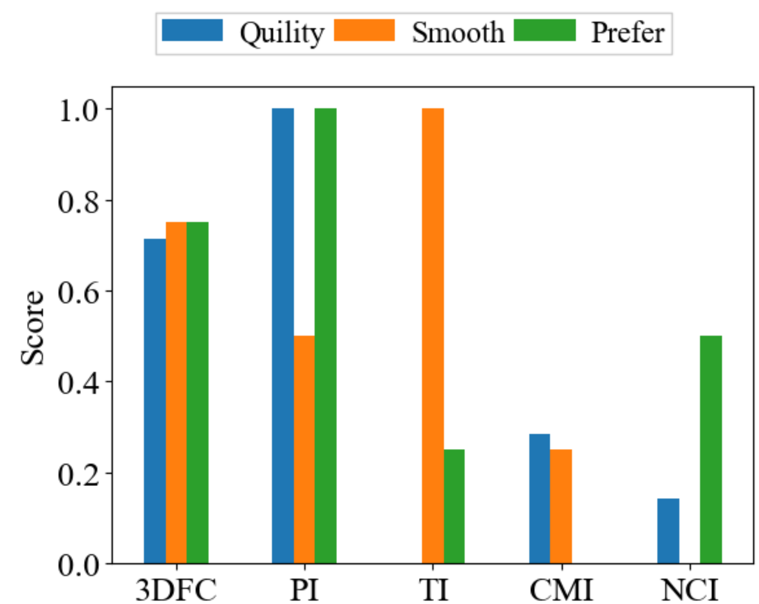


Subjective Results - Soldier









Need more investigations for most suitable algorithms on each sequences

CONCLUSION

Conclusion

- Studied the uninvestigated problem of error concealment for 3D point cloud streaming
- Proposed error concealment algorithms
  - PI, TI, CMI, and NCI
- Significantly outperform the baseline 2DFC and usually outperform 3DFC
  - Report computational time for tradeoff
- CMI and NCI usually outperform others except for Basketball and Soldier sequences
  - Issue for avatars carrying objects
- 3DFC performs well in the user study
  - Hypothesis: Subjects are accustomed to stalls rather than cracks

Future Work

- Exploit parallelization of Graphic Processing Unit (GPU)
- Improve matching for cubes across frames
  - Formulate problem of motion estimation
  - Consider the rotation
- Address issue for avatars with extra items

- Implement real streaming system
- Implement spatial concealment

Publications and Cooperators

- **T. Hung**, I. Huang, S. Cox, W. Ooi, and C. Hsu, “Error Concealment of Dynamic 3D Point Cloud Streaming,” In submission of *ACM International Conference on Multimedia* (MM’22), **Under Review.**
- C. Wu, C. Hsu, **T. Hung**, C. Griwodz, W. Ooi, and C. Hsu, “Quantitative comparison of point cloud compression algorithms with PCC Arena,” *IEEE Transactions on Multimedia*, pages 1–16, February 2022. Accepted to Appear.
- I. Huang, S. Cox, Y. Shi, **T. Hung**, W. Ooi, and C. Hsu, “Trajectory-driven error concealment of dynamic 3D point cloud streaming for interactive applications,” In preparation for *IEEE Transactions on Multimedia*.

- I-Chun Huang, *National Tsing Hua University*
- Sam Cox, *National Tsing Hua University*
- Young Shi, *National Tsing Hua University*
- Chen-Hao Wu, *Synopsys*
- Wei Tsang Ooi, *National University of Singapore*
- Carsten Griwodz, *University of Oslo*
- Chih-Fan Hsu, *National Yang Ming Chiao Tung University*

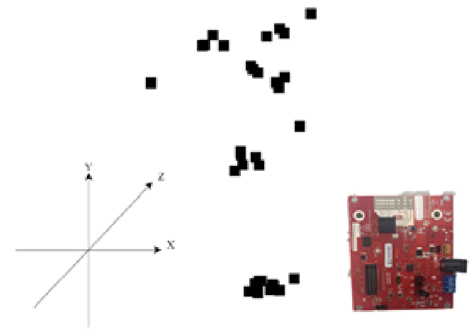
Thank you for listening

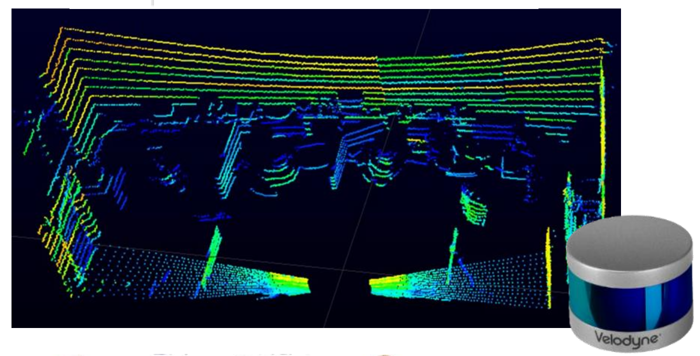
Q&A

Applications on Point Clouds

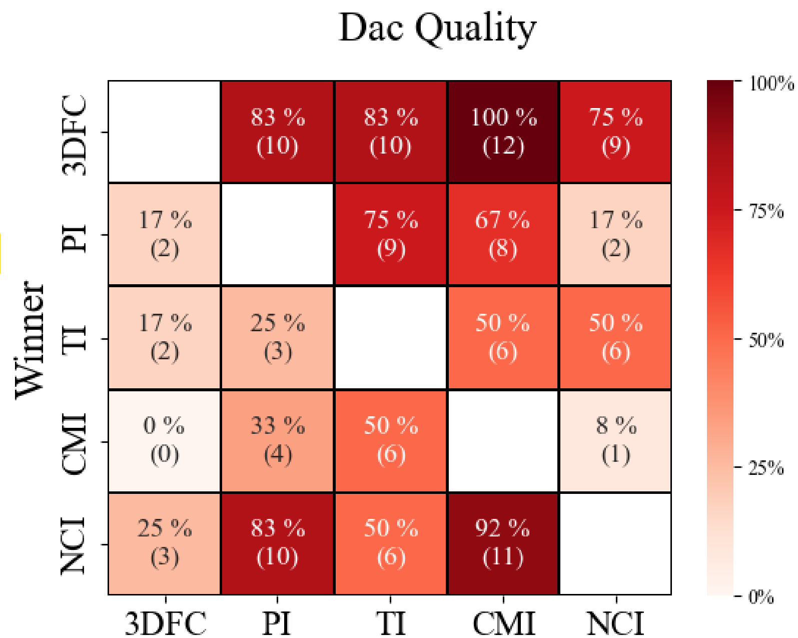
- Sparse Point Clouds
  - Human activity analysis
  - Fall detection
- Cylindrical Point Clouds
  - Civil engineering inspection
  - Obstacle detection
- Dense Point Clouds
  - Entertainment
  - Teleconference

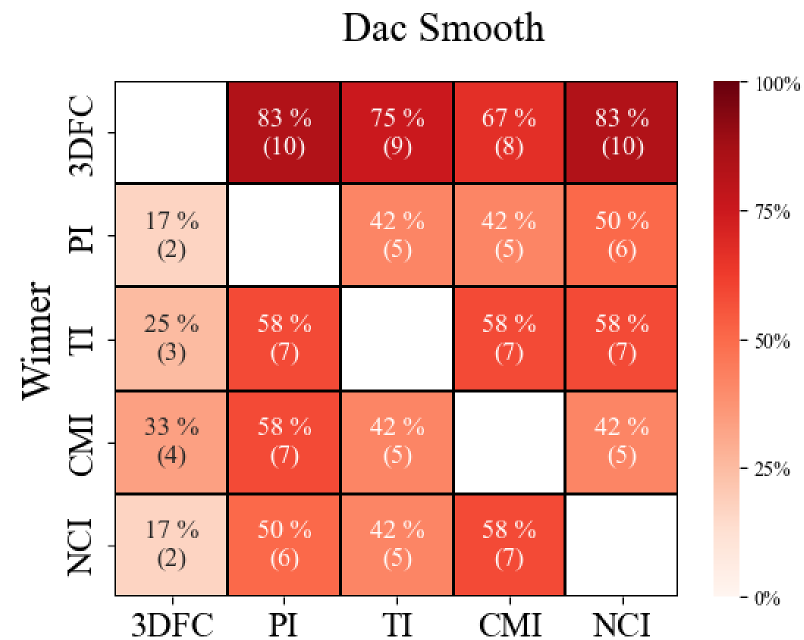
Our usage scenario

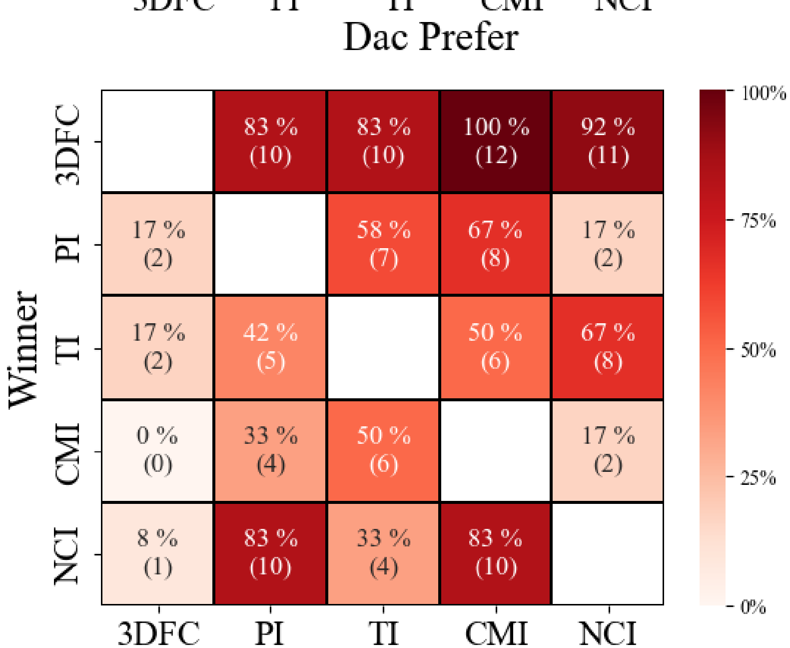


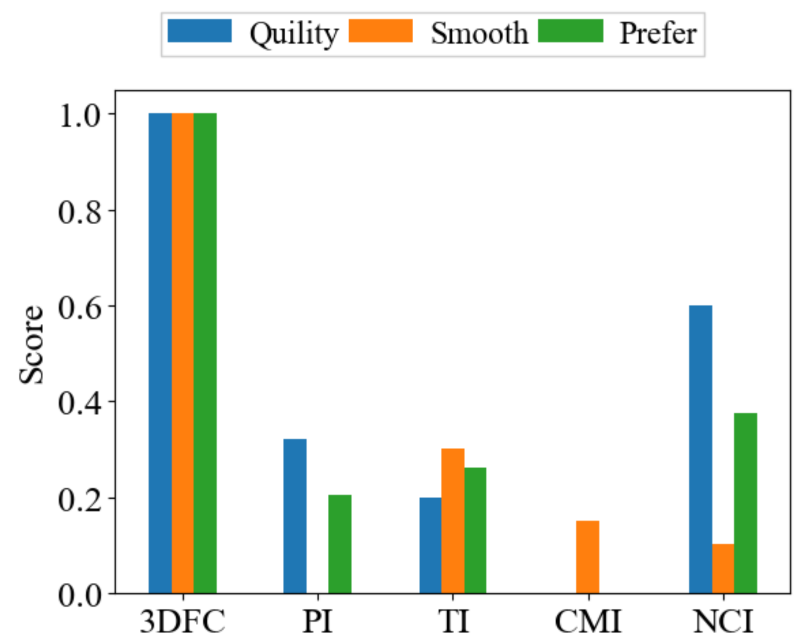


Subjective Results - Dancer

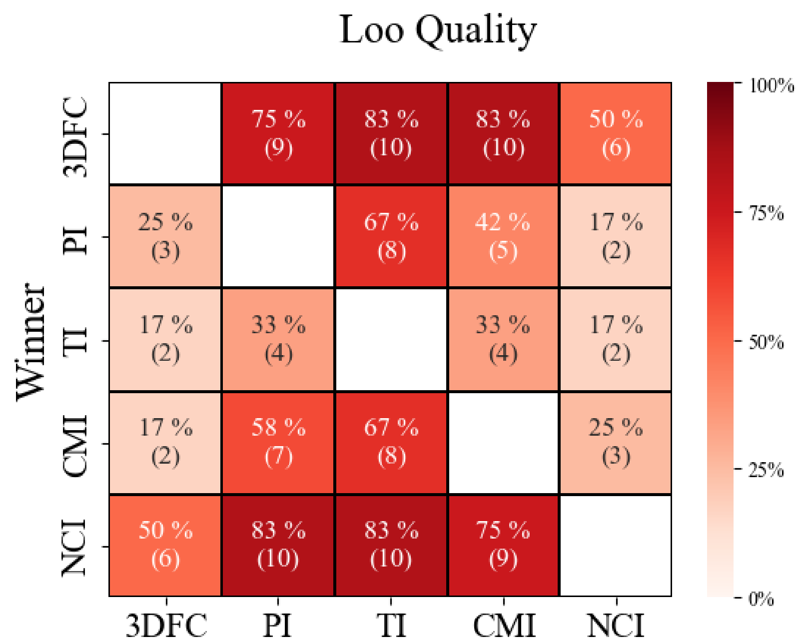


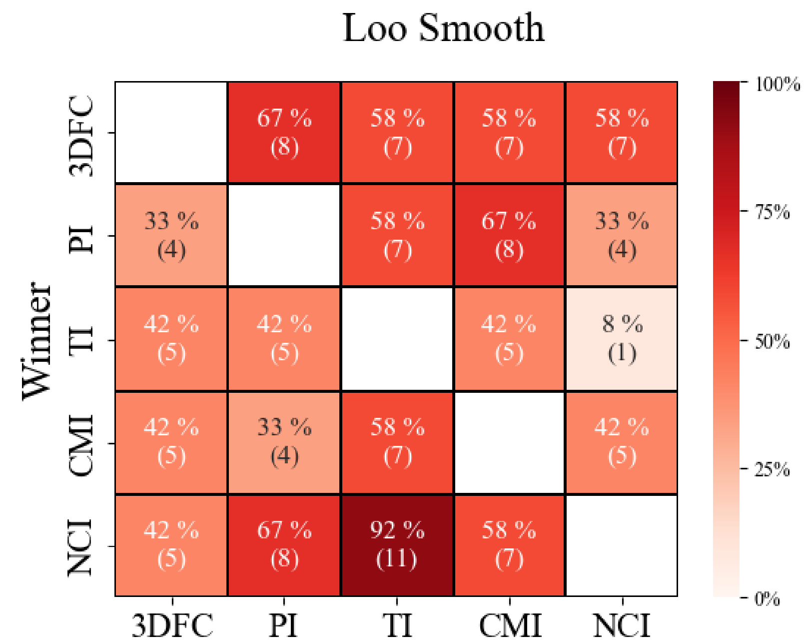


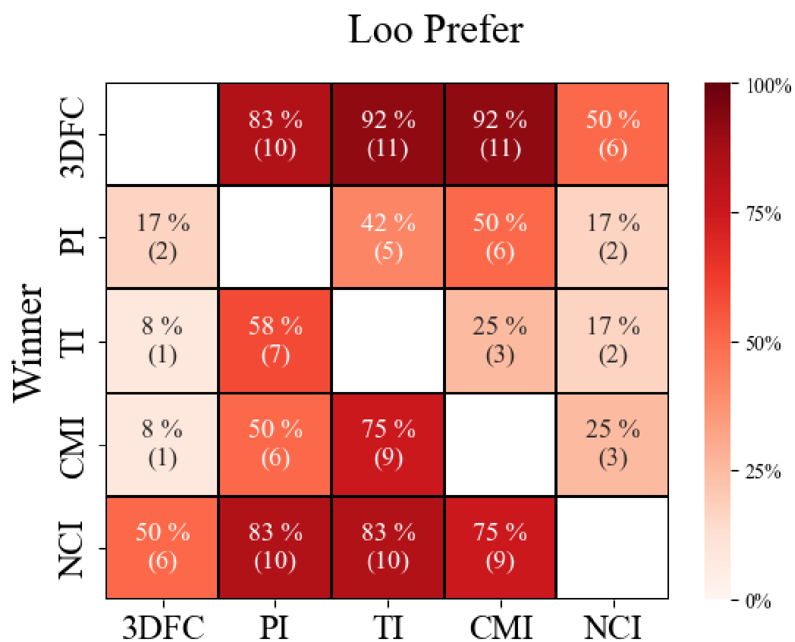


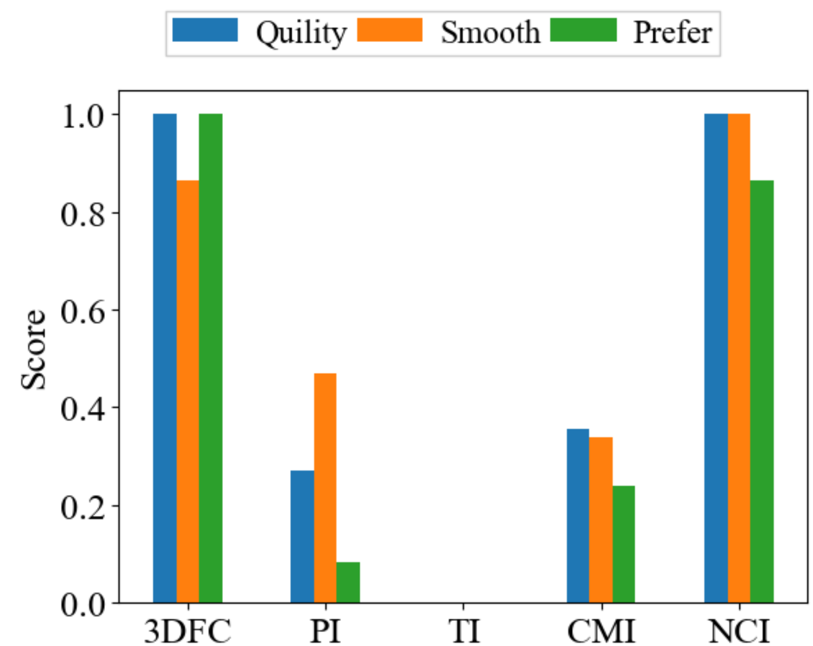


Subjective Results - Loot

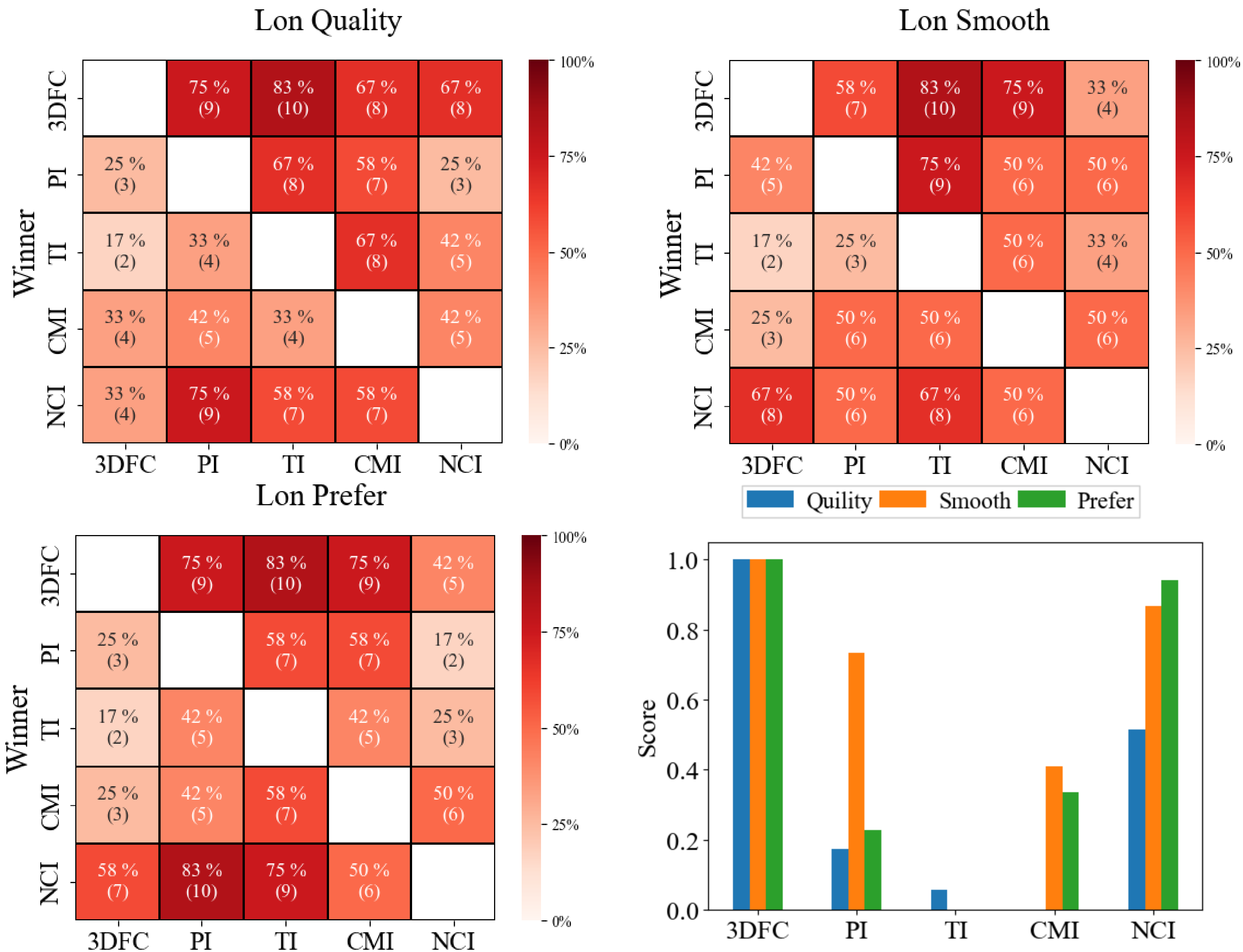






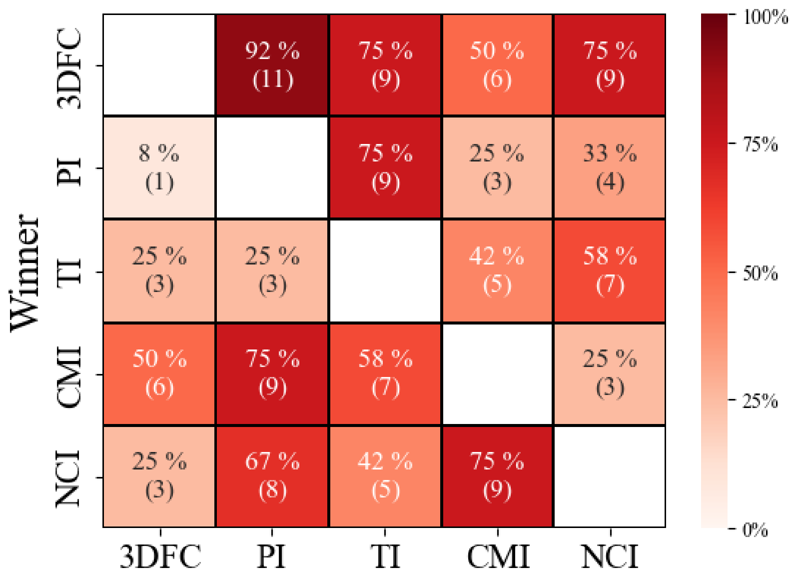


Subjective Results - Longdress

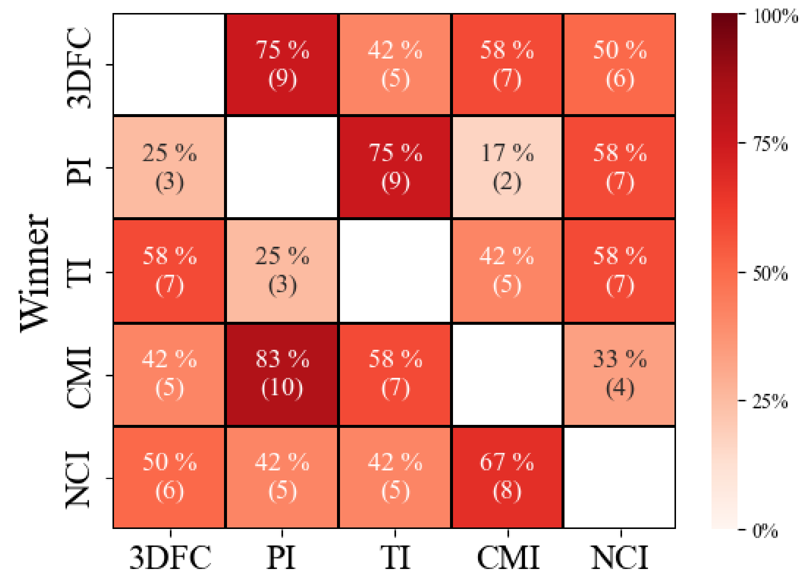


Subjective Results - Redandblack

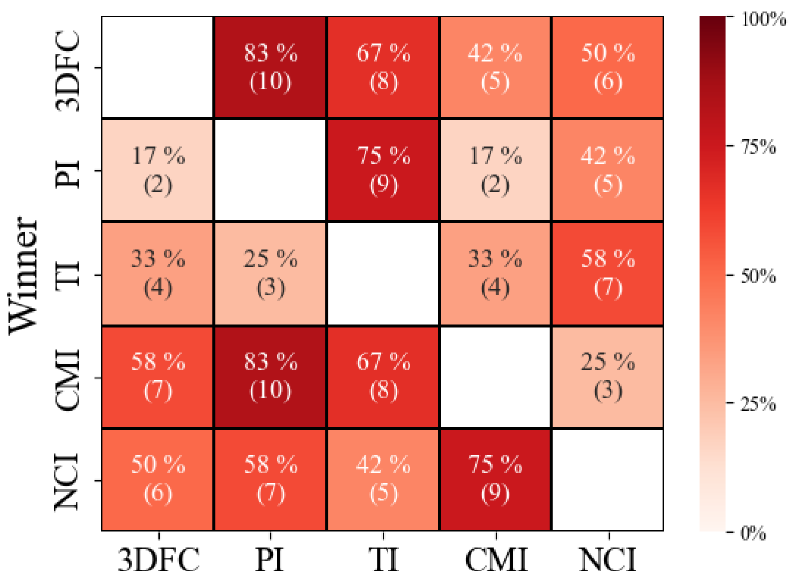
Red Quality

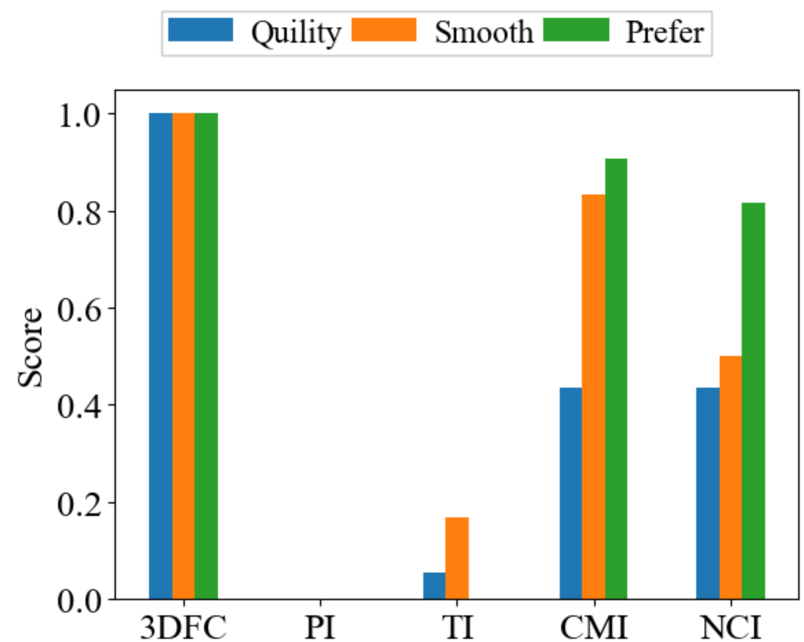


Red Smooth



Red Prefer





Why don’t we use ML

- NN-based PCC algorithms runs at least 10 times slower than SP-based ones

C.-H. Wu, C.-F. Hsu, T.-K. Hung, C. Griwodz, W. T. Ooi, and C.-H. Hsu. Quantitative comparison of point cloud compression algorithms with PCC Arena. IEEE Transactions on Multimedia, pages 1–16, February 2022. Accepted to Appear

Why use V-PCC as the ref SW

- Proposed by a well known ISO/IEC standards organization group: MPEG
- SP-based PCC algorithm
- Suitable for point cloud videos
- Well documented

S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivoku´ca, S. Lasserre, Z. Li et al., “Emerging MPEG standards for point cloud compression,” IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 133–148, 2018.

PC MB/s

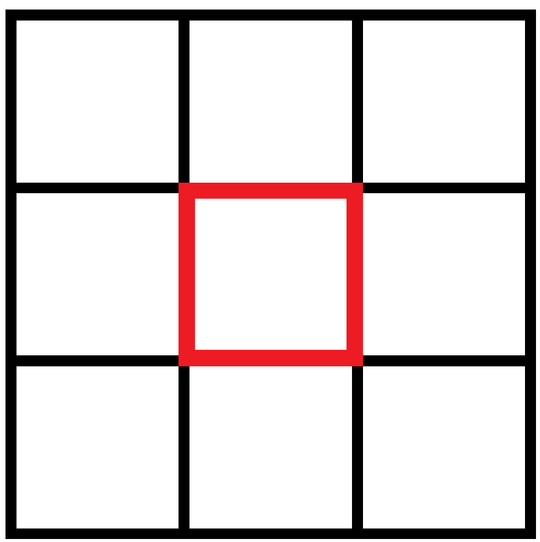
	queen	longdress	loot	redandblack	soldier
Average number of points (in 300 frames)	1,005,000	834,000	794,000	727,000	1,076,000
Bitrates for transmitting uncompressed video (Mbytes/s)	514.47	542.22	490.61	448.21	681.96

C. Cao, M. Preda, and T. Zaharia, “3D point cloud compression: A survey,” ACM International Conference on 3D Web Technology (Web3D’19), pages 1–9, July 2019.

NCI

$$f_2' = \Sigma_{i=1}^{27}(\mathrm{M}_i/V_i) \;/\; \Sigma_{i=1}^{27}(1/\mathrm{V}_i)$$

where $M_i = (x_i, y_i, z_i)$, $Vi = |x_i| \times |\mathrm{y}_i| \times |\mathrm{z}_i|$



GPSNR

$$\mathrm{ACD}(\mathbf{P}_1, \mathbf{P}_2) = \frac{1}{|\mathbf{P}_1|} \sum_{p \in \mathbf{P}_1} \min_{p\prime \in \mathbf{P}_2} \|p - p\prime\|_2^2$$