Point Cloud Compression & Communication

Zhu Li

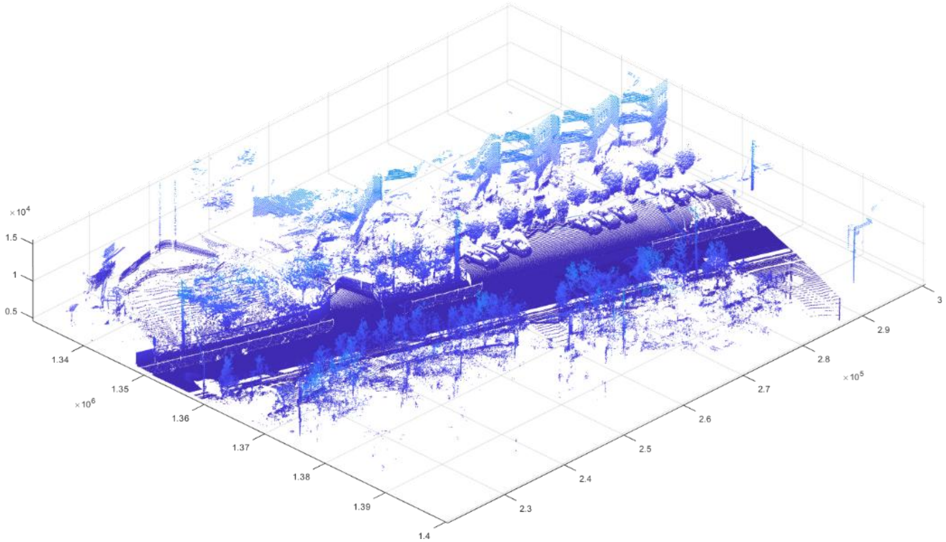
Director, UMKC NSF Center for Big Learning

Dept of Computer Science & Electrical Engineering

University of Missouri, Kansas City

Email: zhu.li@ieee.org, lizhu@umkc.edu

Web: http://l.web.umkc.edu/lizhu



Short Bio:

Research Interests:

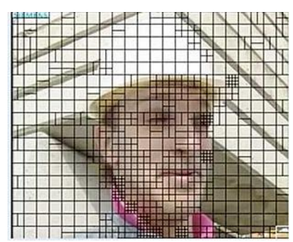
- Immersive visual communicaiton: light field, point cloud and 360 video coding and low latency streaming
- Low Light, Res and Quality Image Understanding
- What DL can do for compression (intra, ibc, sr, inter end2end)
- What compression can do for DL (compression, acceleration, distributed training)

NSF I/UCRC Center *for* Big Learning
Creating Intelligence

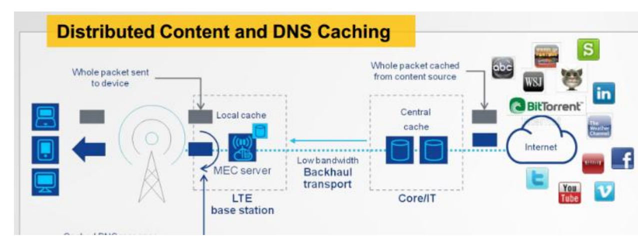
Multimedia Computing & Communication Lab
Univ of Missouri, Kansas City

signal processing and learning

image understanding



visual communication



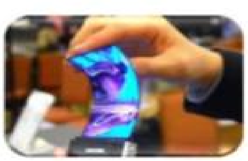
mobile edge computing & communication

Outline

- Short Self Intro & Research Highlights
- Point Cloud Capture and Applications
- Geometry Compression
- Graph Signal Processing and Attributes Compression
- UMKC PCC work
  - Static Geometry Compression: Plane Projection Approximation
  - Dynamic Geometry Compression: Kd-tree decomposition and residual coding
- Summary

Media Computing & Communication Horizon

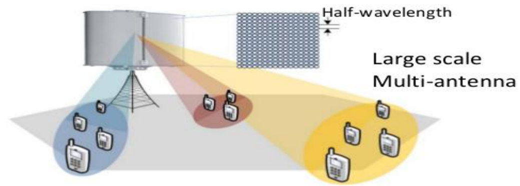
Devices



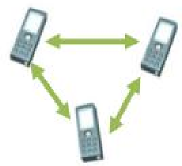
Networks

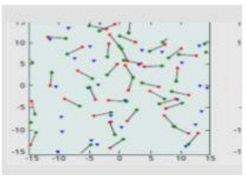
mmWave 5G

FD-MIMO



D2D





SDN/MEC

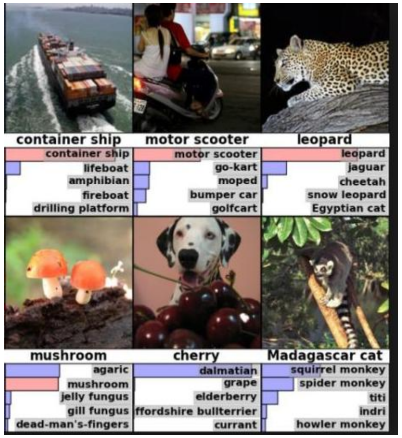
Akamai

Applications

4k/8k UHD Video

Free Viewpoint TV

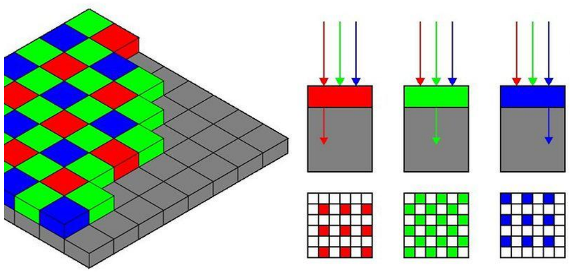
Samsung VR/AR



BIGDATA visual intelligence

Dark Image Enhancement

- To design network to denoise the low-light image in Bayer domain
- To use wavelet decomposition to divide and conquer the problem by learning sensor field sub images using separate netowks



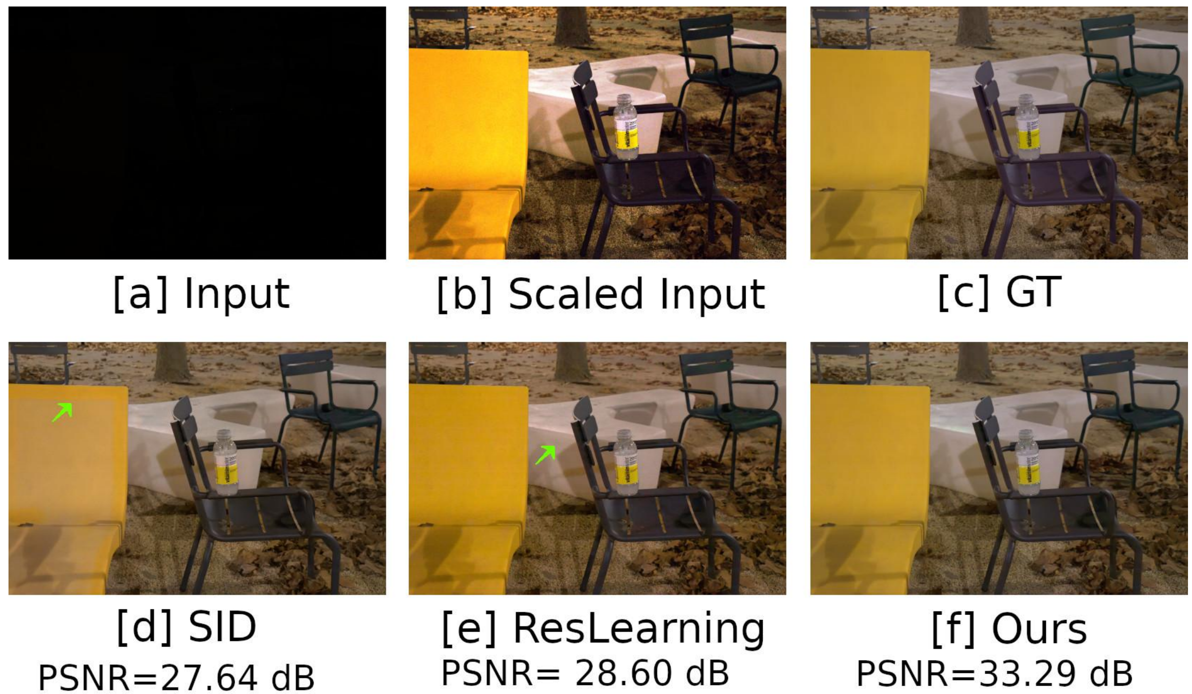


Figure 4: [a] Extreme low-light image from Sony a7S II exposed for 1/25 second . [b] 250x intensity scaling of image in [a]. [c] Ground truth image captured with 10 second exposure time. [d] Output from SID[]. SID introduced some artifacts around the edge of the chair as shown by green arrow. [e] Output from ResLearning[]. The white region as indicated by arrow in image is not properly reconstructed as white compared to that in ground truth image. [f] Our result.

Decomposition based residual learning from sensor field

- Decomposition of the target image via Wavelet
- Adaptive loss functions for different subbands to exploit strong texture prior

$$\mathcal{L}_{sturctural} = 1 - SSIM(\hat{x}, x) \qquad (2)$$

$$\mathcal{L}_{HF} = \mathcal{L}_1 + \alpha * \mathcal{L}_{sturctural} \qquad (3)$$

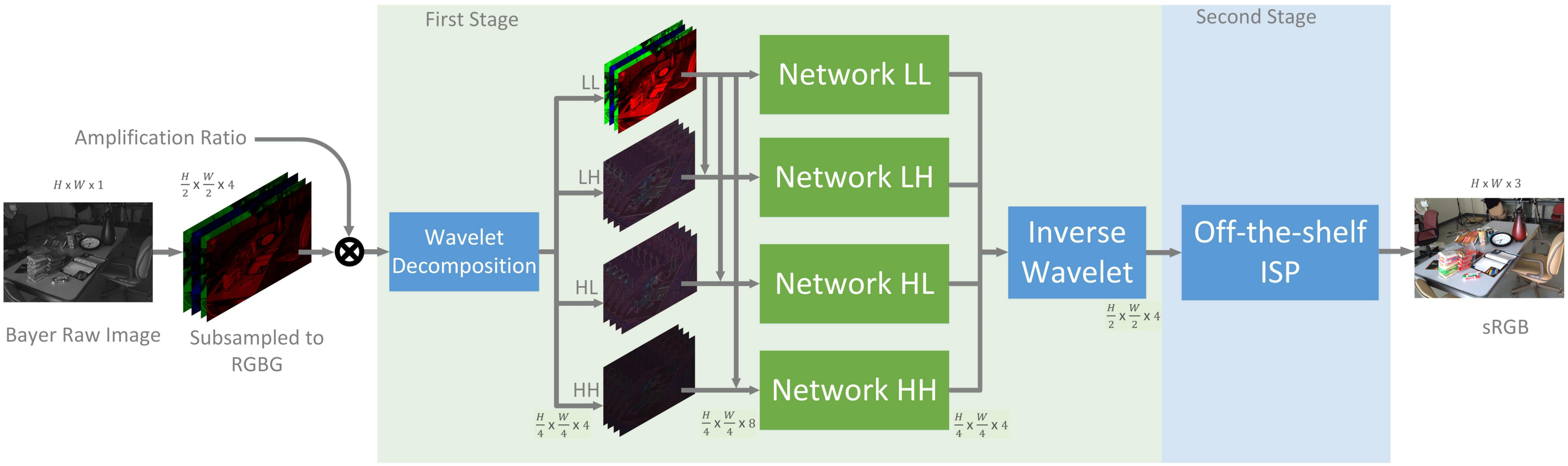


Figure 12: Overview of our wavelet decomposition based network. The first stage learns the decomposed image and used the inverse wavelet to reconstruct the denoised 4 channel image. The second stage uses the off-the-shelf ISP to enhances the image and converts into 3 channel sRGB image.

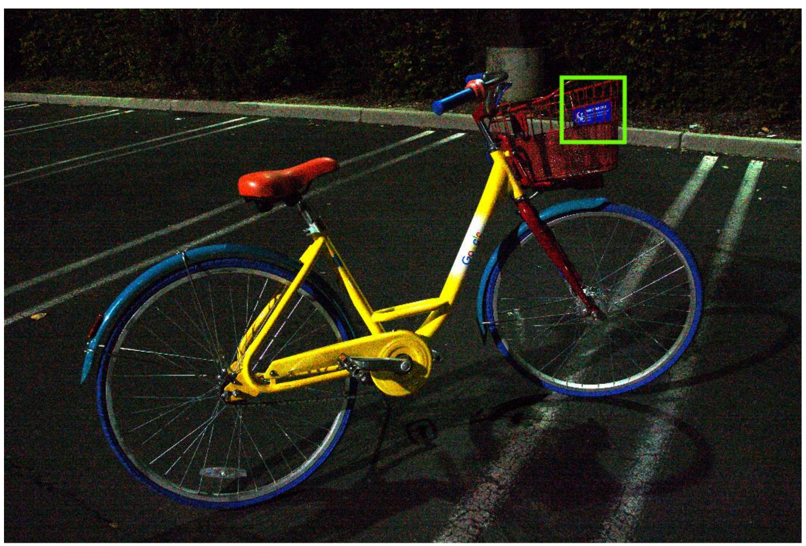
Experimental Results

Table 1. Comparison of our proposed method of denoising before ISP with the existing method of joint denoising and demosaicing.

Experiments	PSNR	RMSE	NIQE
SID[6]	28.97	0.03956	5.1904
ResLearning[21]	29.16	0.03926	5.8507
Ours	**30.02**	**0.03568**	**4.6166**

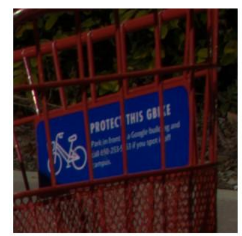
Table 2. Comparison of our proposed method of denoising before ISP with the existing method of joint denoising and demosaicing based on three subgroups of the dataset for 100x, 250x and 300x.

Experiments	100x	250x	300x
SID[6]	30.08	28.42	28.52
ResLearning[21]	30.53	28.78	28.38
Ours	**32.34**	**29.97**	28.22



Noisy Image

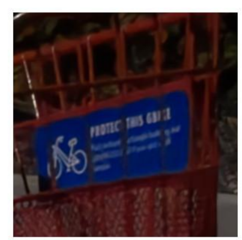
Noisy Image



GT

28.34 dB
BM3D

34.20 dB
SID

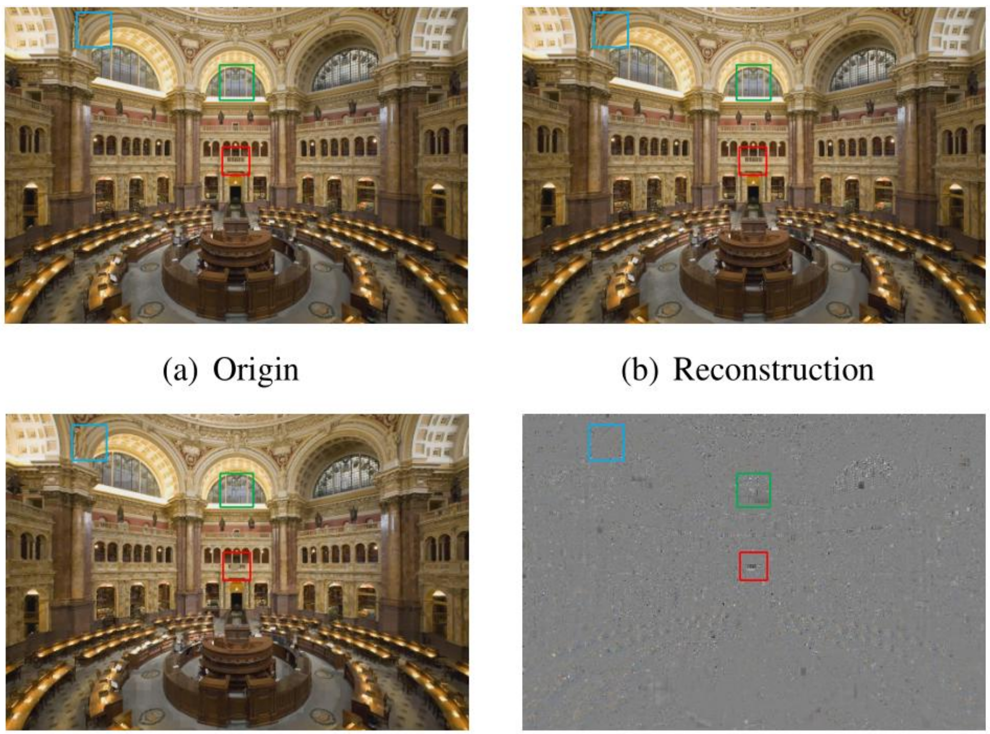


34.39 dB
ResLearning

35.04 dB
Ours

Deep Guided Filtering Deblocking

- The residual frame can be used as the guidance for the in-loop filter of the reconstructed frame
  - Larger residuals indicate larger reconstruction errors

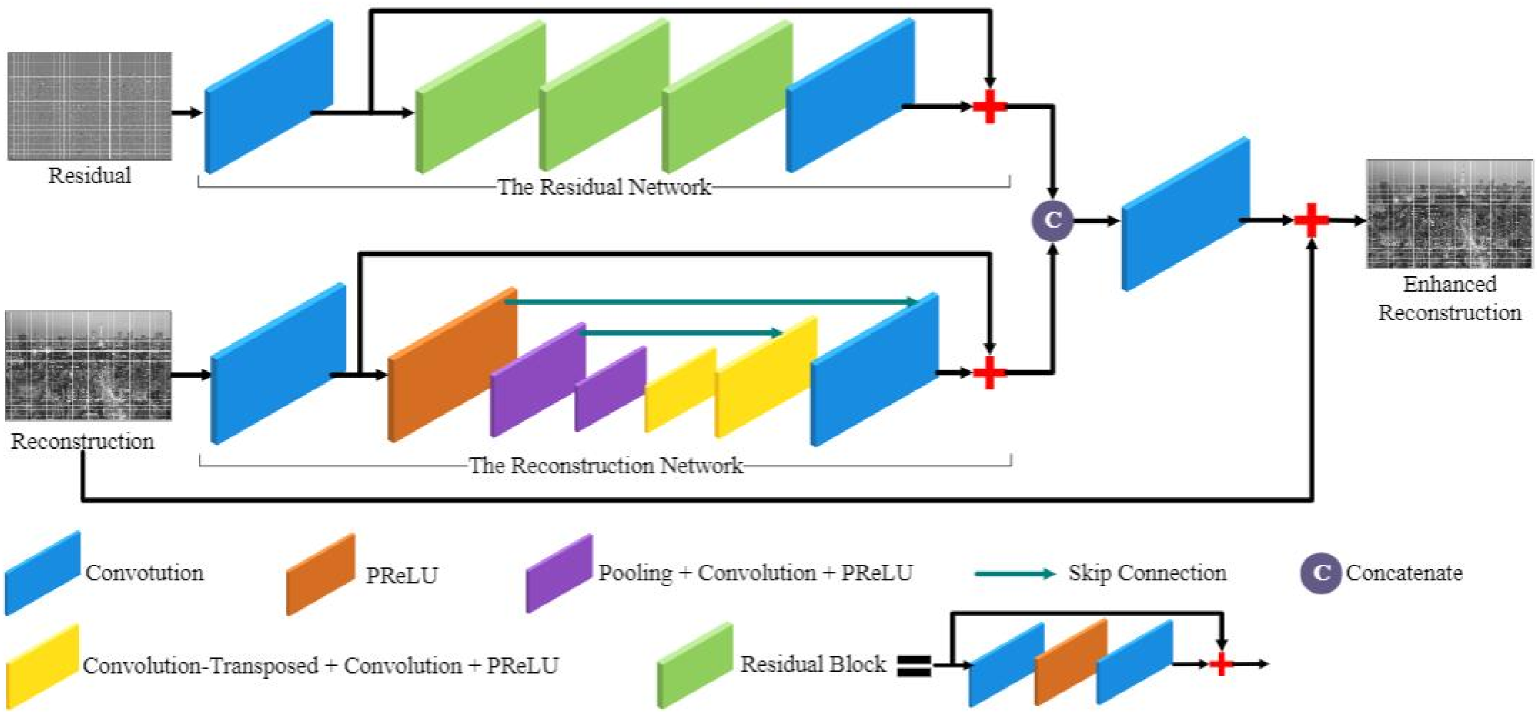


(a) Origin (b) Reconstruction

(c) Prediction (d) Residual

Coding-prior-based in-loop filter

- The residual frame is used as the additional input
- Specific networks for reconstruction and residual
  - Residual Network: residual blocks
  - Reconstruction Network: down-sampling and up-sampling



Experimental results

- Comparison with VRCNN

Intra: 2.1% improvement

Class	Sequence	VRCNN vs. HEVC	RRCNN vs. HEVC
A	Traffic	−8.1%	**-10.2%**
	PeopleOnStreet	−7.7%	**-9.4%**
B	Kimono	−5.9%	**-8.6%**
	ParkScene	−6.2%	**-8.1%**
	Cactus	−2.7%	**-5.8%**
	BasketballDrive	−5.2%	**-7.7%**
	BQTerrace	−2.9%	**-4.2%**
C	BasketballDrill	−10.6%	**-13.8%**
	BQMall	−7.3%	**-9.3%**
	PartyScene	−4.6%	**-5.6%**
	RaceHorses	−5.8%	**-7.1%**
D	BasketballPass	−7.6%	**-9.5%**
	BQSquare	−5.3%	**-6.3%**
	BlowingBubbles	−5.5%	**-6.7%**
	RaceHorses	−8.9%	**-10.2%**
E	FourPeople	−10.0%	**-12.8%**
	Johnny	−9.1%	**-12.5%**
	KristenAndSara	−9.4%	**-11.8%**
	Class A	−7.9%	**-9.8%**
	Class B	−4.6%	**-6.9%**
	Class C	−7.1%	**-8.9%**
	Class D	−6.8%	**-8.2%**
	Class E	−9.5%	**-12.4%**
Avg.	All	−6.8%	**-8.9%**

Inter: 0.7% improvement

Class	Sequence	VRCNN vs. HEVC	RRCNN vs. HEVC
A	Traffic	−5.0%	**-6.0%**
	PeopleOnStreet	−1.4%	**-1.6%**
B	Kimono	−1.9%	**-2.6%**
	ParkScene	−2.7%	**-3.4%**
	Cactus	−3.2%	**-3.9%**
	BasketballDrive	−1.4%	**-1.9%**
	BQTerrace	−5.2%	**-5.8%**
C	BasketballDrill	−3.1%	**-4.3%**
	BQMall	−2.0%	**-2.5%**
	PartyScene	−0.5%	**-1.0%**
	RaceHorses	−1.3%	**-1.4%**
D	BasketballPass	−0.7%	**-0.9%**
	BQSquare	−1.4%	**-2.1%**
	BlowingBubbles	−1.8%	**-2.4%**
	RaceHorses	−1.5%	**-1.6%**
E	FourPeople	−8.2%	**-9.5%**
	Johnny	−7.6%	**-10.2%**
	KristenAndSara	−6.9%	**-7.6%**
	Class A	−3.2%	**-3.8%**
	Class B	−2.9%	**-3.5%**
	Class C	−1.7%	**-2.3%**
	Class D	−1.4%	**-1.7%**
	Class E	−7.6%	**-9.1%**
Avg.	All	−3.1%	**-3.8%**

Deep Radar Signal Learning for Privacy Preserving Fall Detection

- Use case:Seniors assisted living - Fall Detection
- Approach:
  – 77Ghz portable radar array sensor set up: horizontal and vertical scanning, 4x2 Tx/Rx
  – Radar Signal Low Dimension Embedding + LSTM action recognition

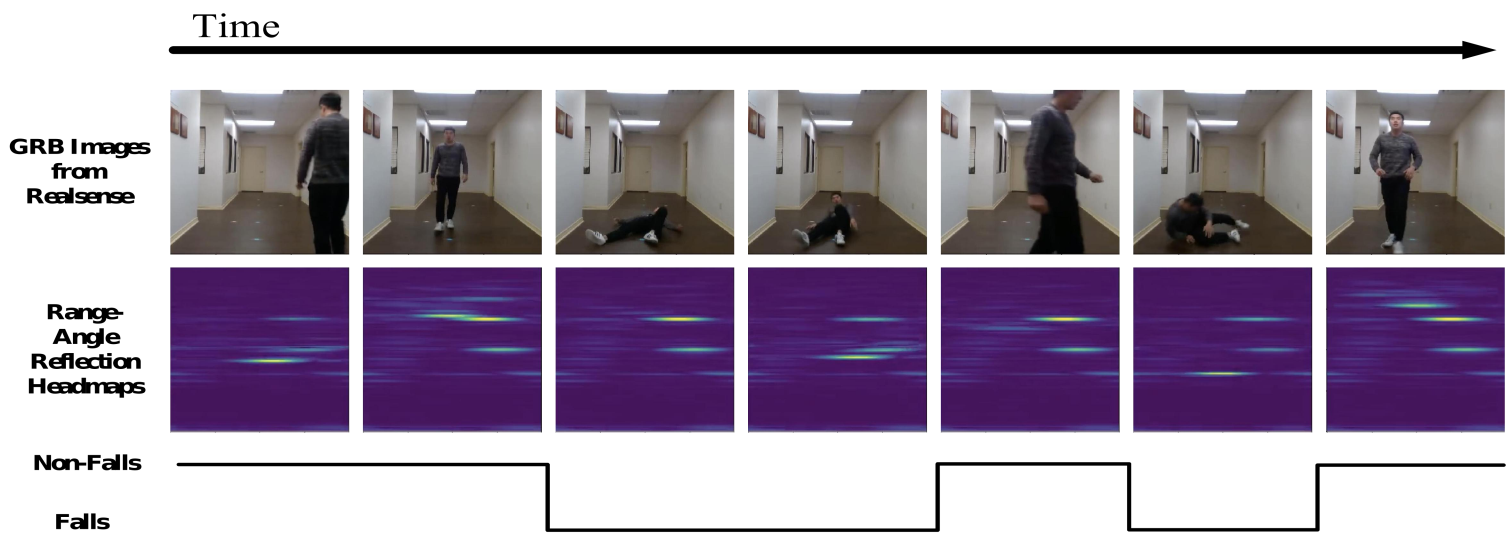


Figure 1. mmWave Radar based Fall Detector

Neural network processing

- Human activities are continuous dynamic patterns that can be recognized in both spatial and temporal dependencies. We use successive radar reflection heatmaps as the representative of human activities.
  - PCA is adopted as RLDE algorithm to project reflection heatmaps $\{H_t, V_t\}$ to a low-dimension subspace P as the elimination of spatial redundancies,
  - The proposed RNN with LSTM units utilizes the changes of motion at the temporal domain. The softmax layer operates as a classifier. The cross-entropy function is adopted as the objective function.

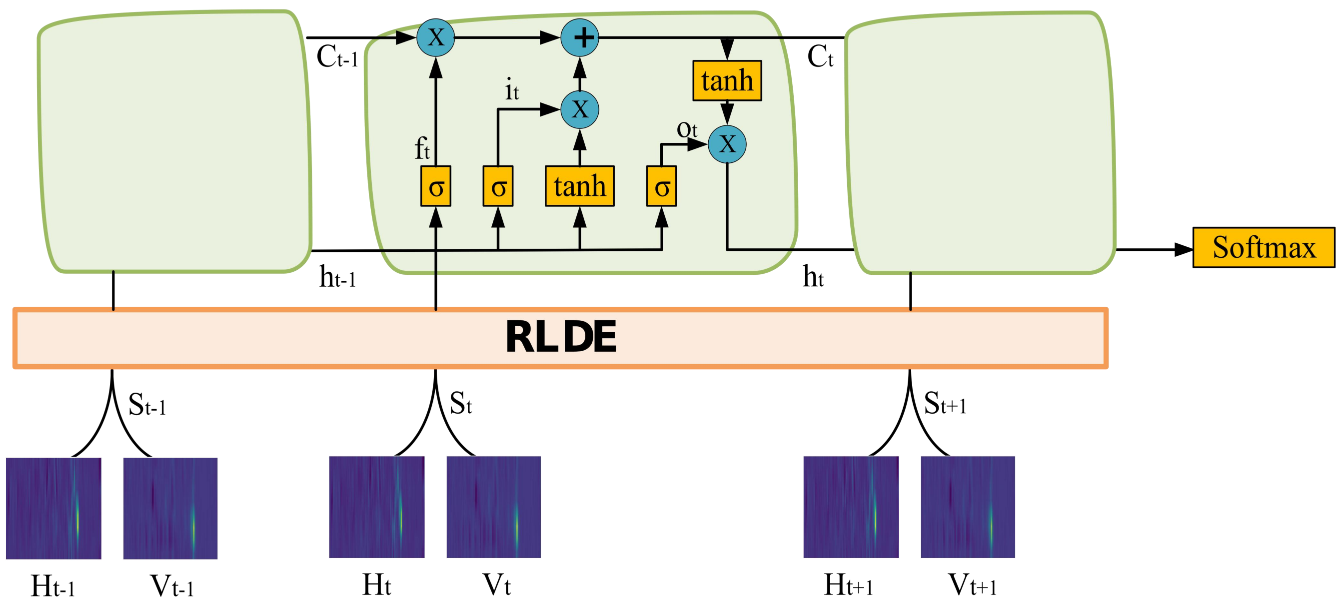


Figure 3. Architecture of RNN with LSTM units

Extensive experiment

- Multiple human activities detections: 7 categories of human activities are labeled: Boxing, Falling, Jogging, Jump, Pick up, Stand up & Walking.

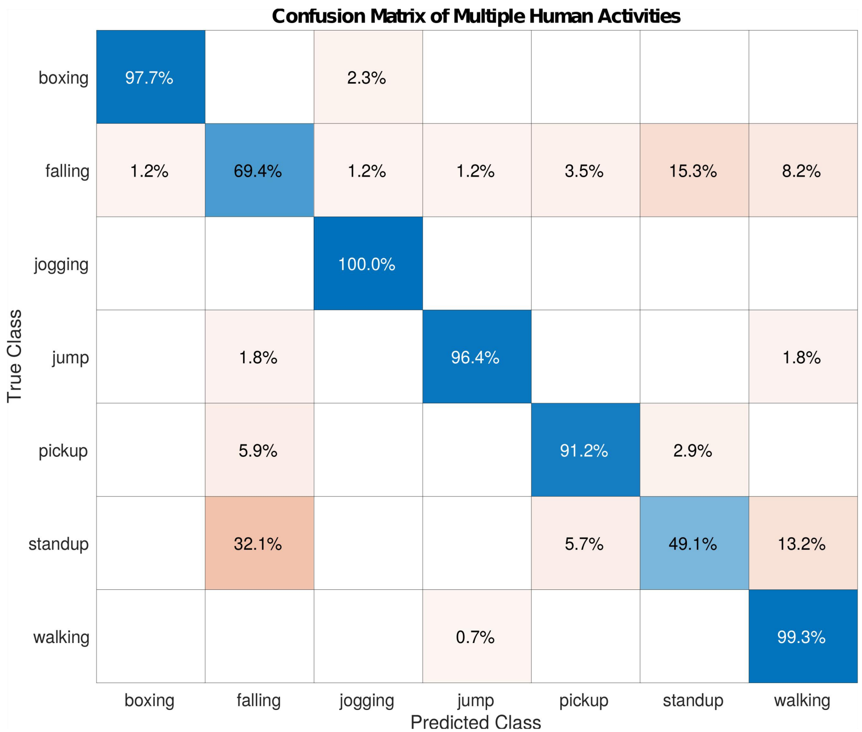


Figure 4. Accuracy of Multiple Human Activities Detecting

Average Inference Time Complexity:
RLDE + LSTM: 0.06042 sec
3DCNN: 7.336 sec

Outline

- Short Self Intro & Research Overview
- Point Cloud Capture and Applications
- Geometry Compression
- Graph Signal Processing and Attributes Compression
- UMKC PCC work
- Summary

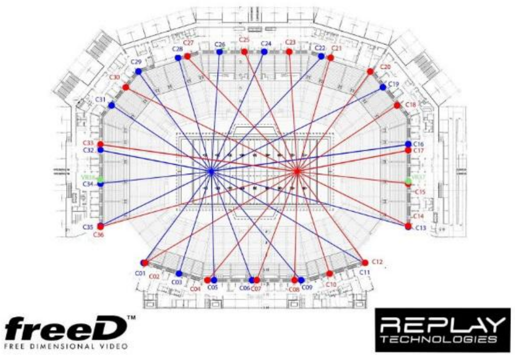
What is Point Cloud

- A collection of Un-ordered points with
  - Geometry: expressed as [x, y, z]
  - Color Attributes: [r g b], or [y u v]
  - Additional info: normal, timestamp, …etc.
- Key difference from mesh: no order or local topology info

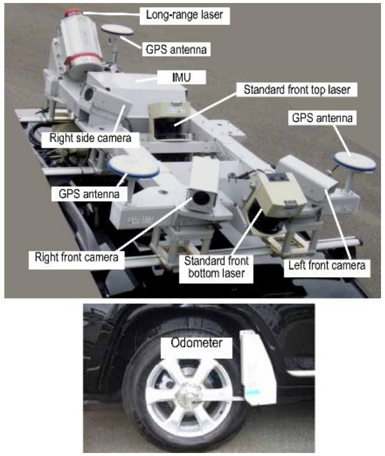


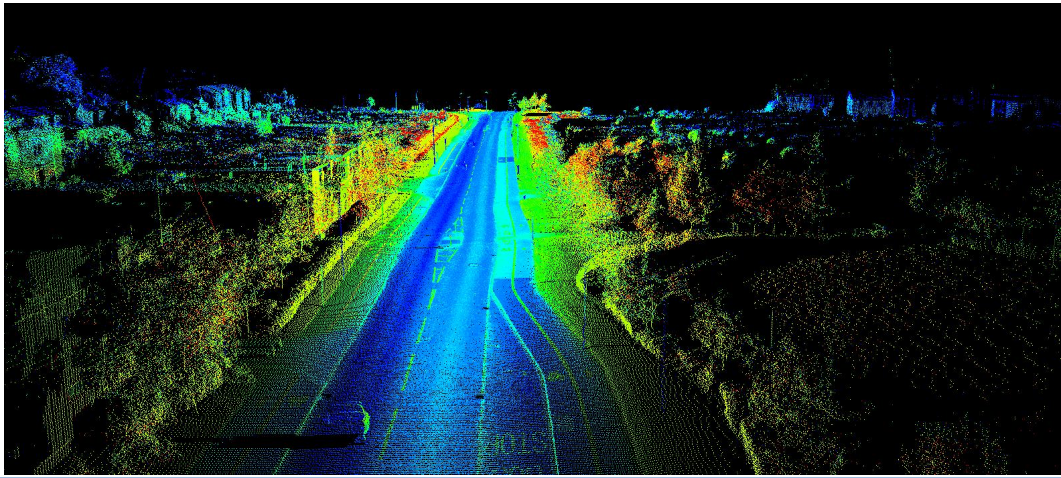
Point Cloud Capture

- Passive: Camera array stereo depth senso



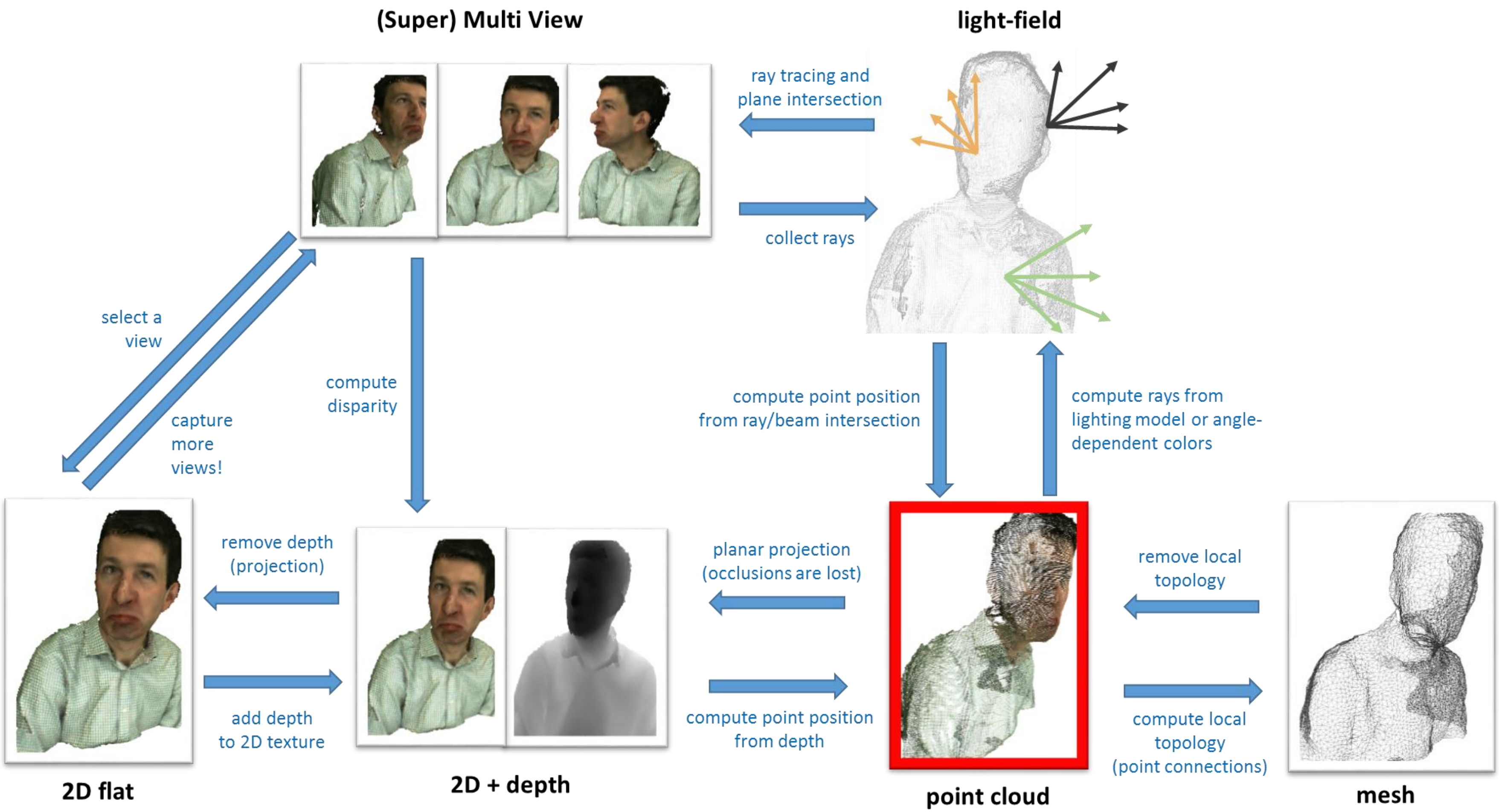
- Active: LiDAR, mmWave, TOF sensors





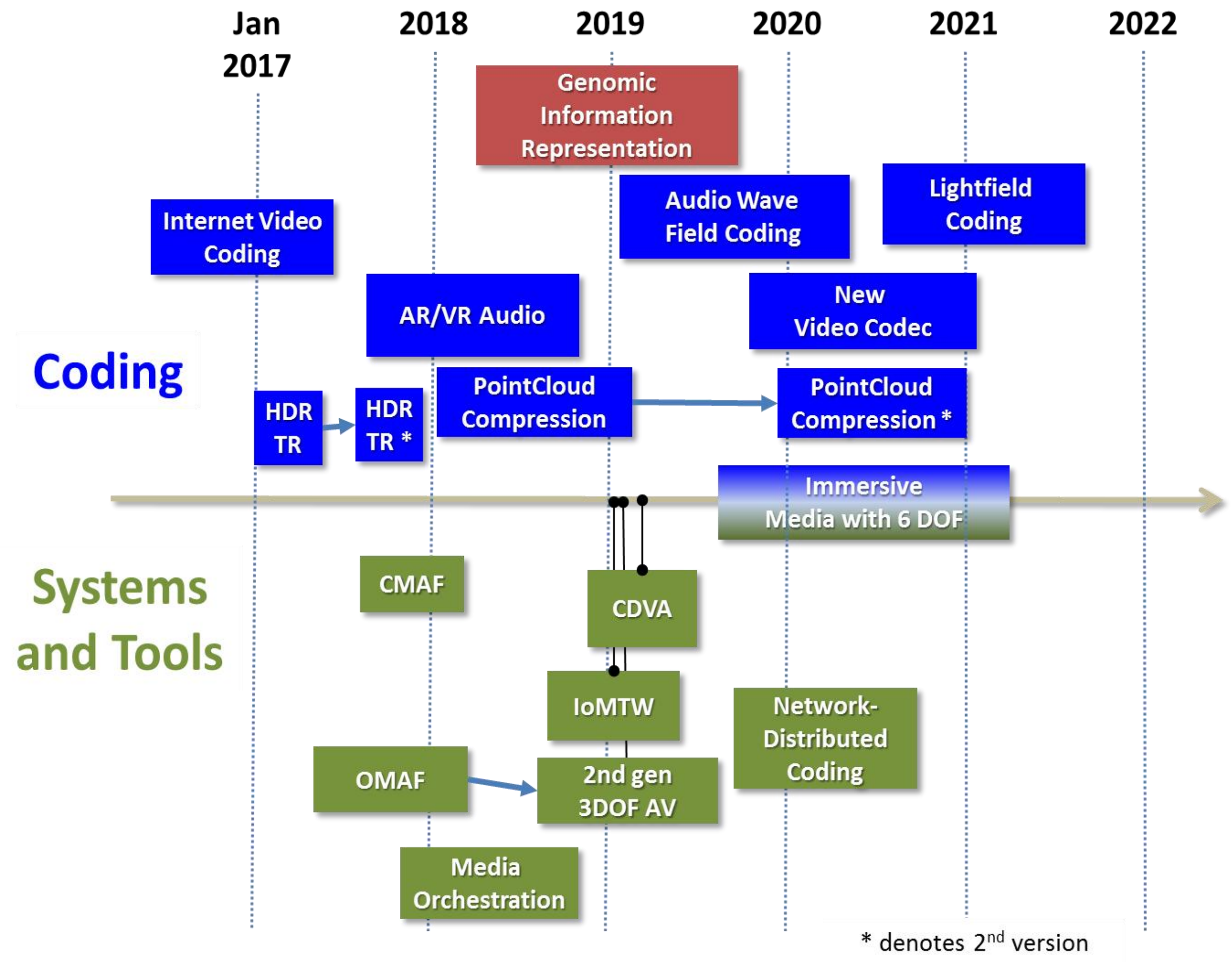
Point Cloud Inter-Operability with Other Formats

- Provide true 6-DoF Content capacity



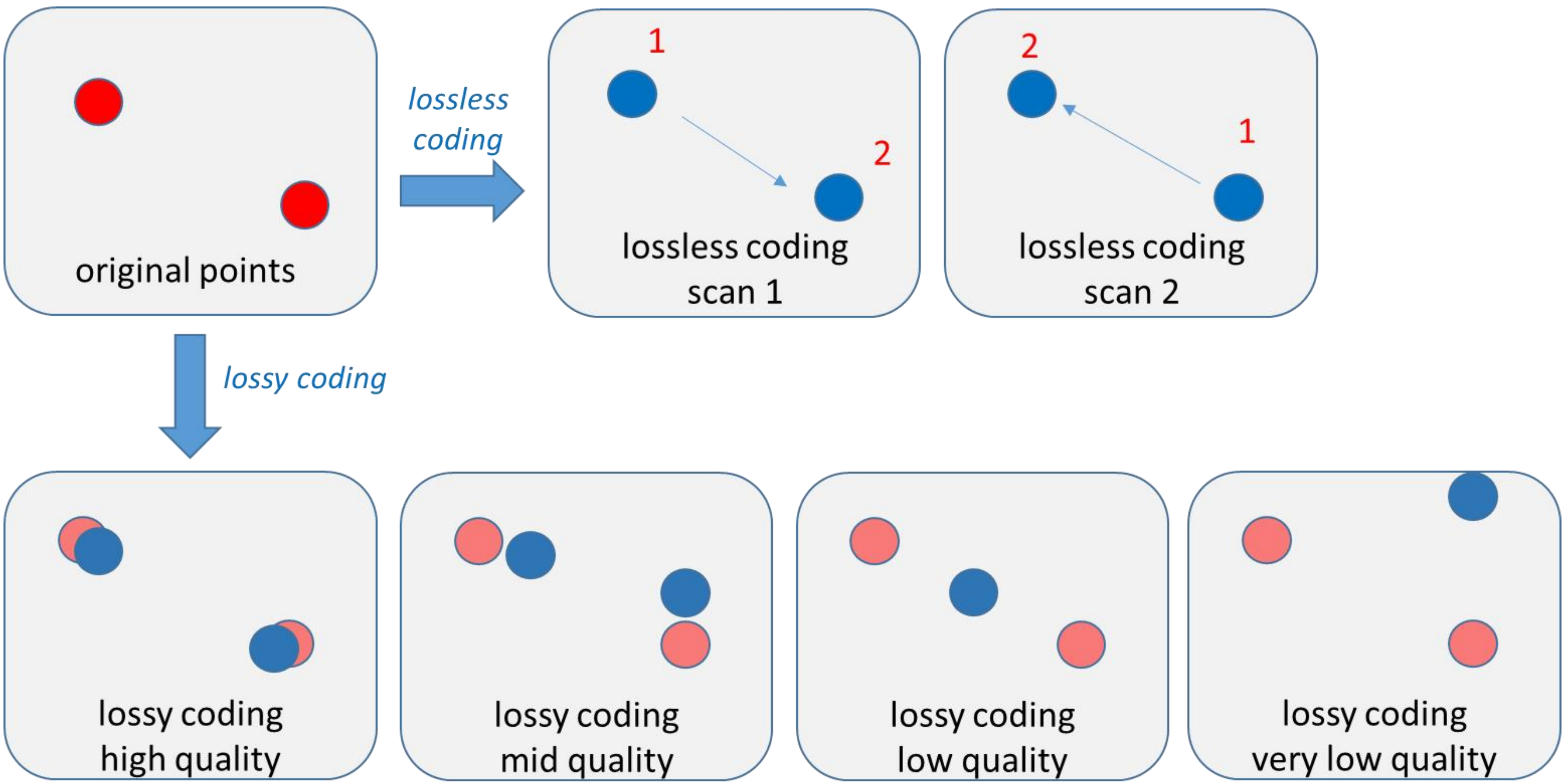
PCC in MPEG

- Part of the MPEG-Immersive grand vision



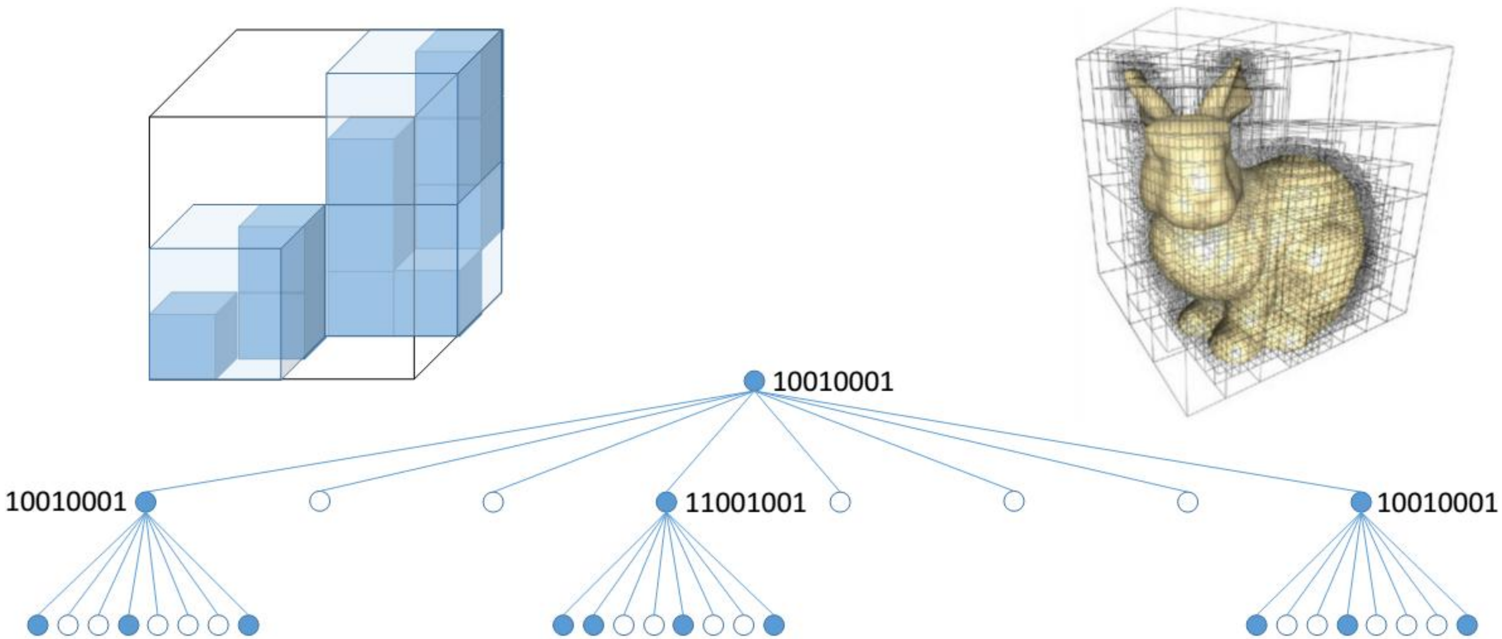
Point Cloud Compression - Geometry

- Can be lossy in both quantization and samples



Octree Based Point Cloud Compression

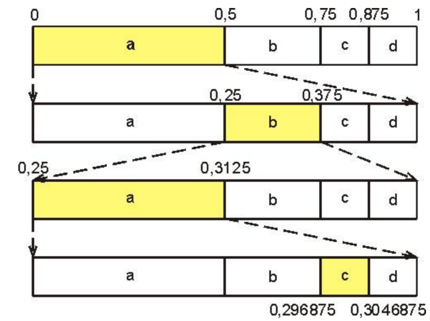
- Octree is a space partition solution
  - Iteratively partition the space into sub-blocks.
  - Encoding: 0 if empty, 1 if contains data points
  - Level of the tree controls the quantization error



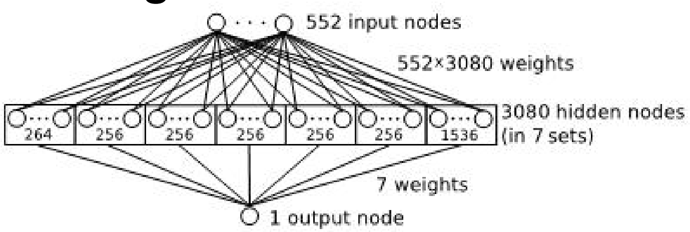
Credit: Phil Chou, PacketVideo 2016

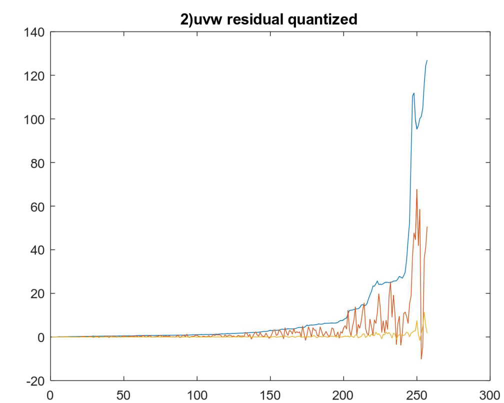
Lossless Compression of the Octree with Neural Network driven CABAC

- Tree Structure:
  - DFS scanning of the Octree node byte to have a byte stream
  - Compression of the byte stream via Arithmetic Coding, or shallow neural network PAQ coding



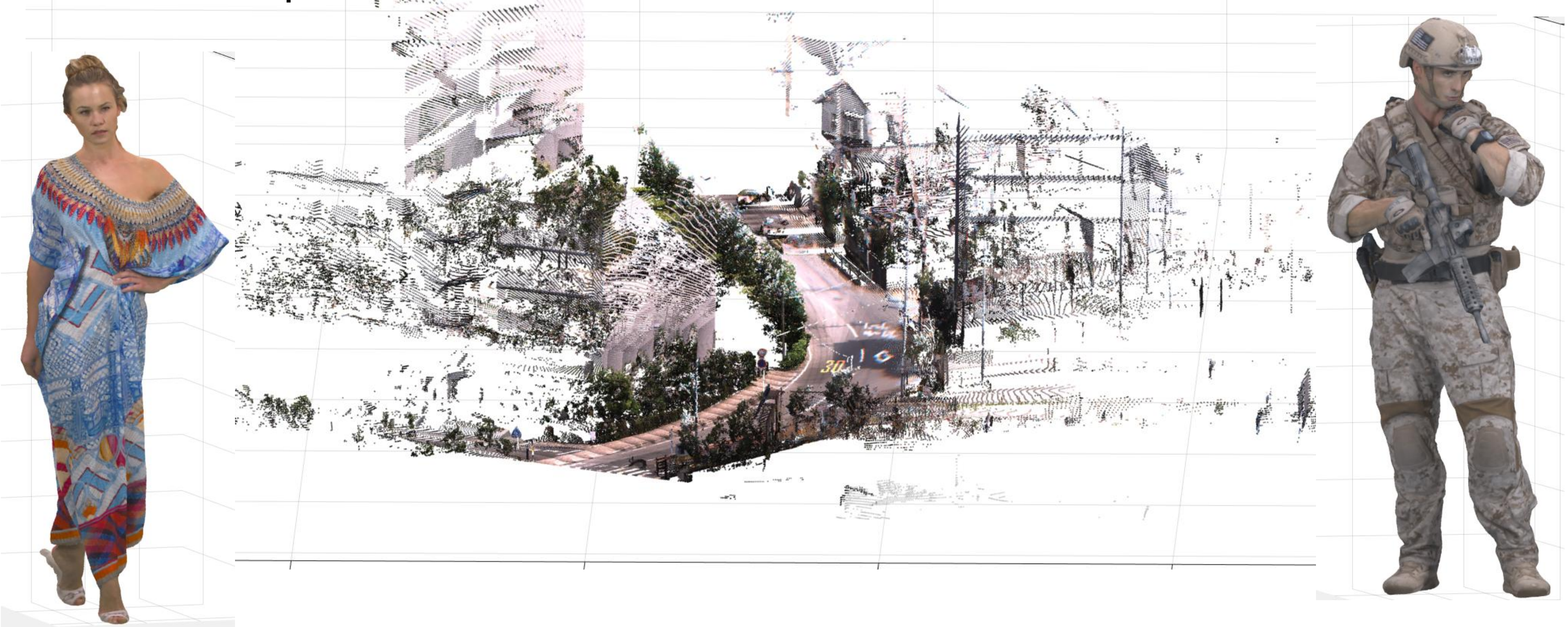
- Residual Coding:
  - Range coding: coding the residual against a ref point (eg., centroids of octree leaf node centroids)
  - Plane/Surface approximation coding: compute the projection distances to a surface, surface can be polynomial or planar.





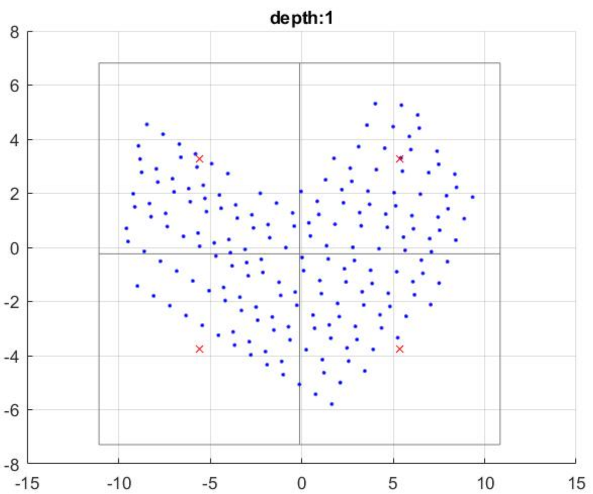
Scalable Point Cloud Geometry Coding

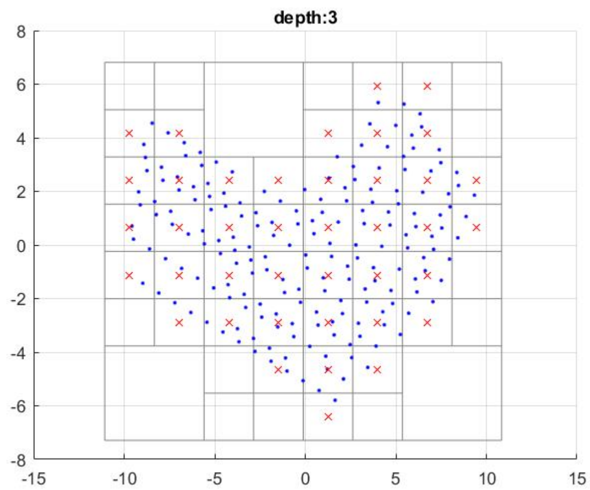
- Binary Tree embedded Quadtree (BTQT) coding:
  - Binary tree partition to have lossy geometry approximation
  - Refine each leaf node with Quadtree/Octree to offer scalable details upto near lossless

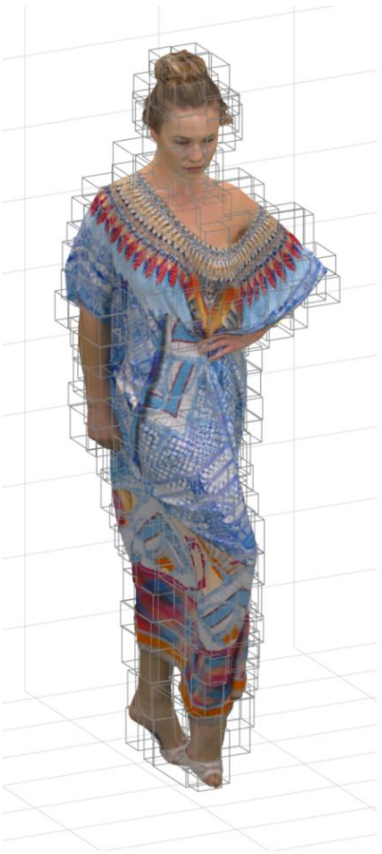


Scalable Geometry Coding with BTQT

- Construct Binary Tree of Point Cloud
  - $R_1 = (2^L-1)*(2+K) + 6*K$, cost of signalling for resolution *K* bits and binary tree depth *L*
- *Intra-Coding* i.e. either *Quadtree* (flat surface) or *Octree*(not flat)
  - QT case overhead: $R_2 = 3*p + 3*q$ bits, for singalling norma at p bits and point at q bits. q < K *proportional* to the leaf node size.







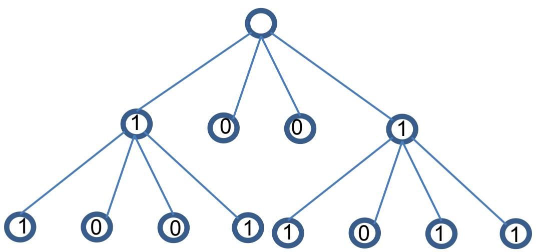
Quadtree/Octree Mode Decision

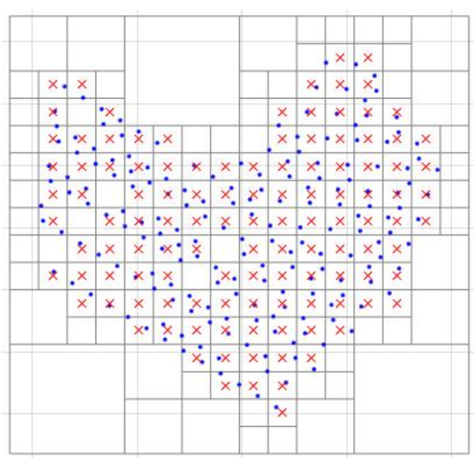
Quadtree | Octree

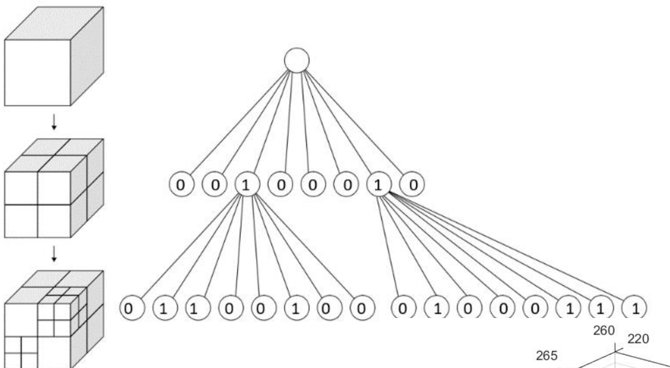
Flatness Criterian:

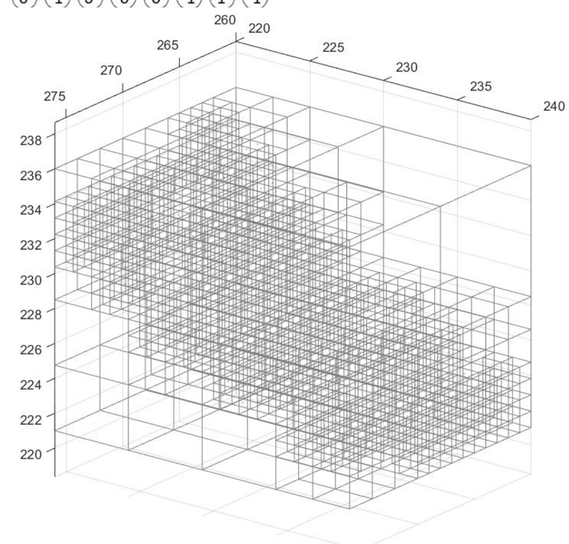
$$\theta = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$$

$$\lambda = eig(cov(X))$$



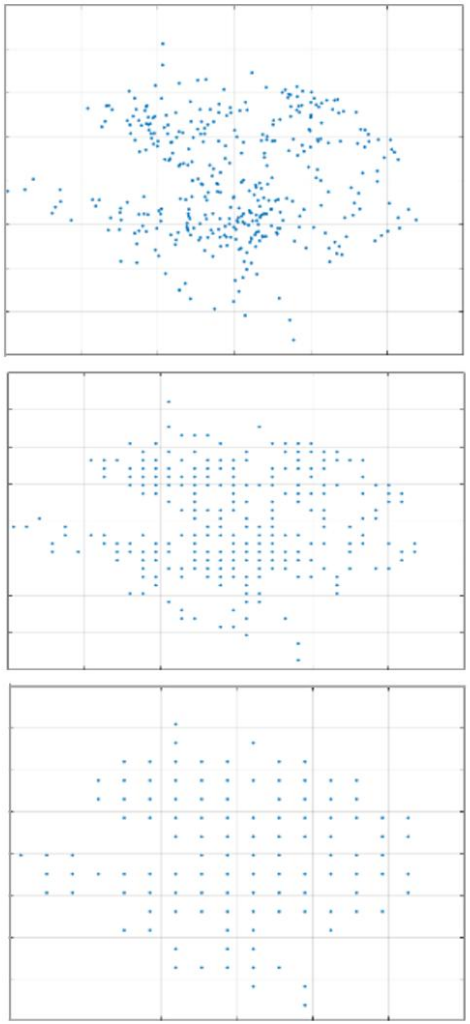




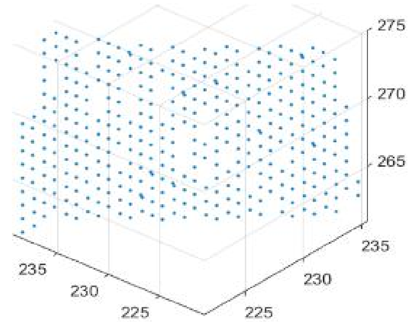


Scalable Coding with Quadtree/Octree

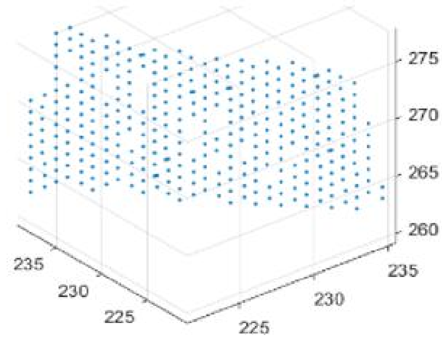
Quadtree



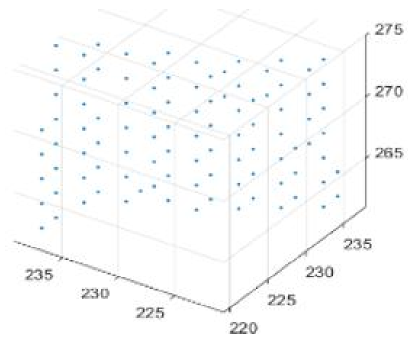
Octree



L = L1



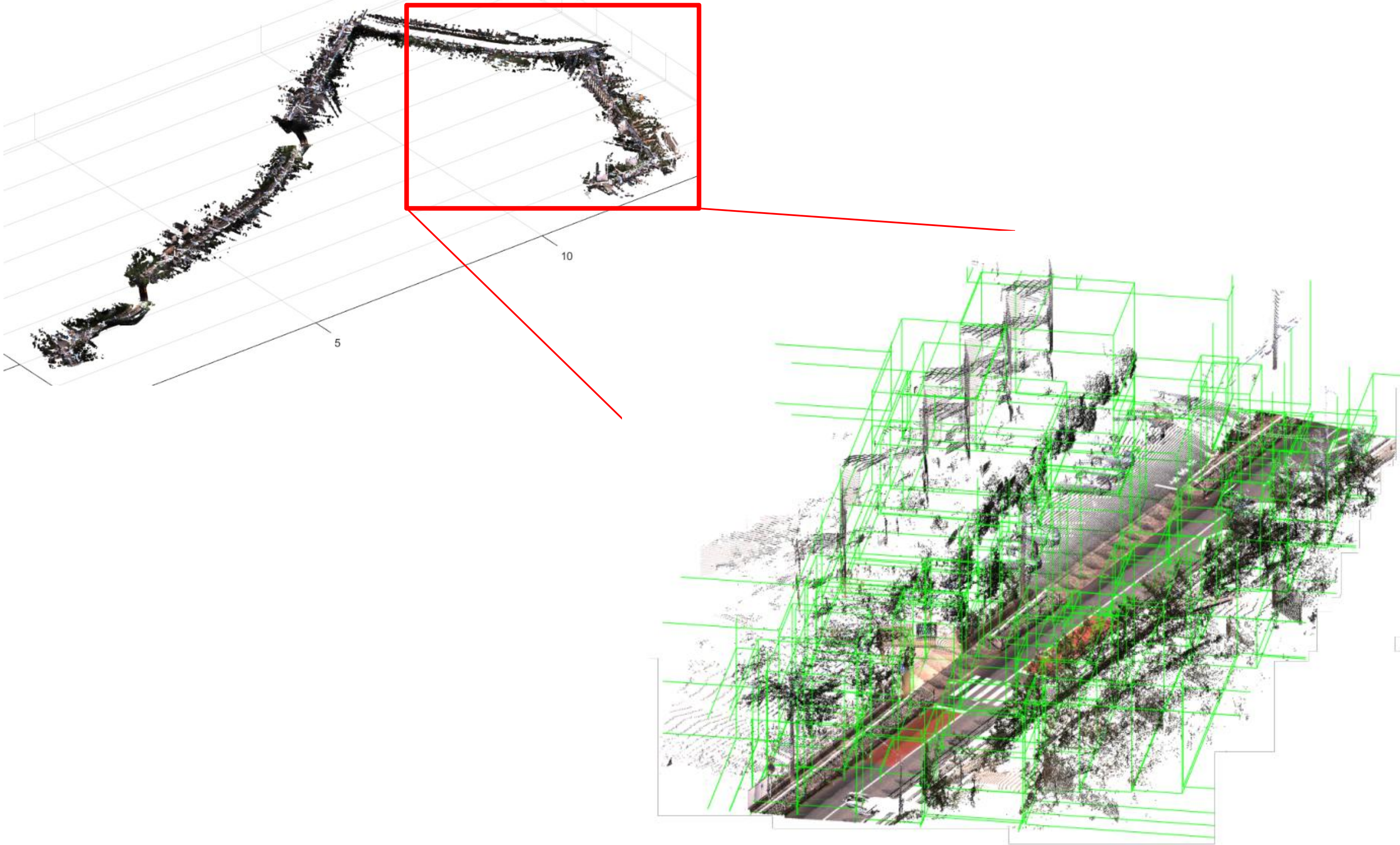
L = L2 < L1



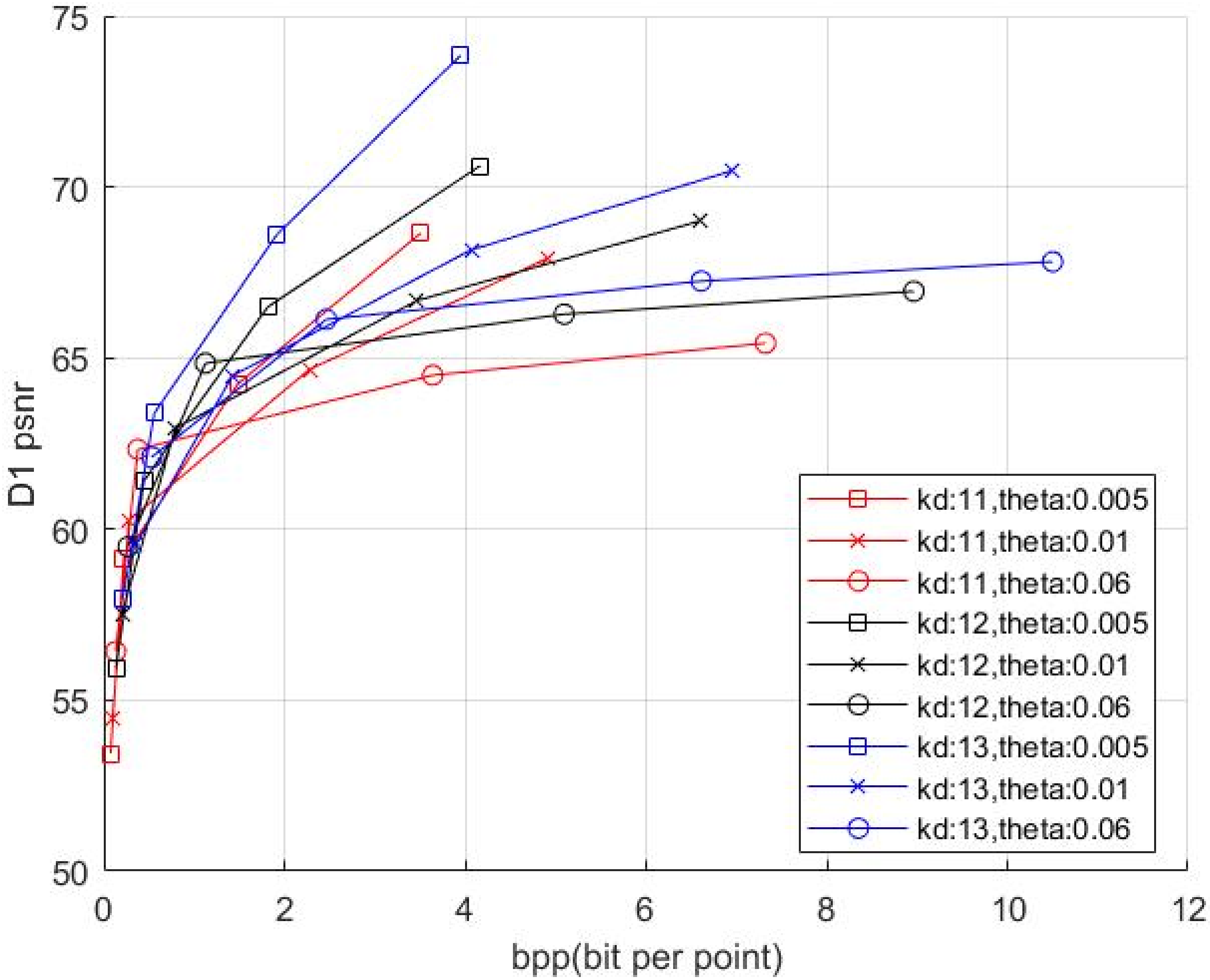
L = L3 < L2

Point cloud Visualization

- *citytunnel* dataset (MERL) – 1.5 km long section of a

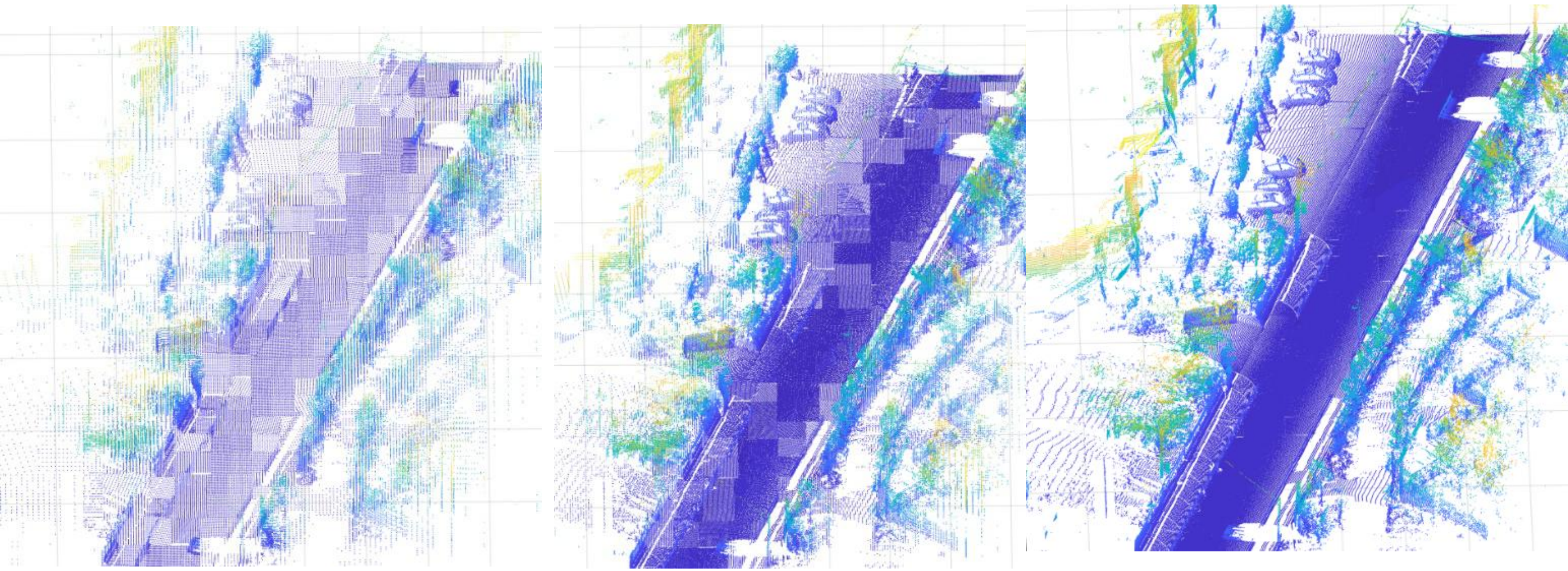


Result: Category 1 Geometry Coding Efficiency



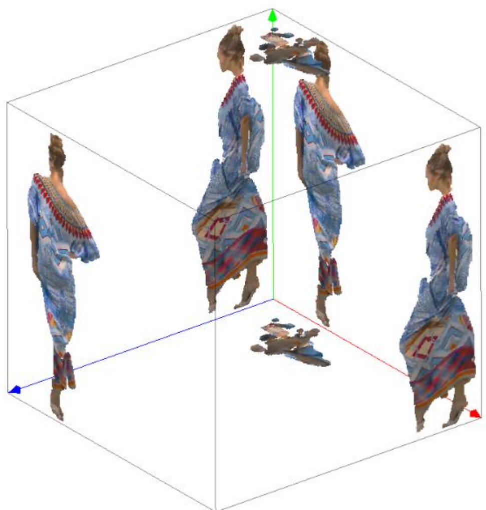
Reconstructed-Zoomed

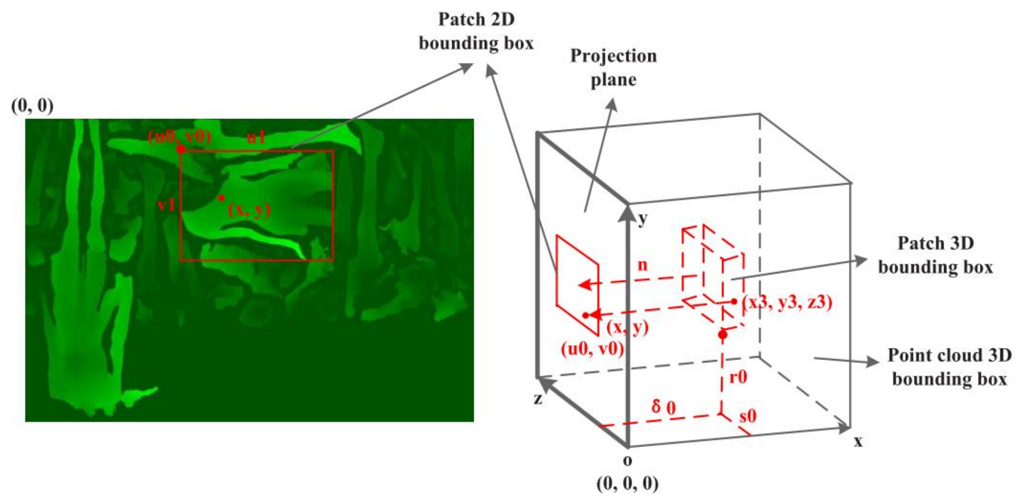
- Various reconstruction accuracy:



Video-based point cloud compression

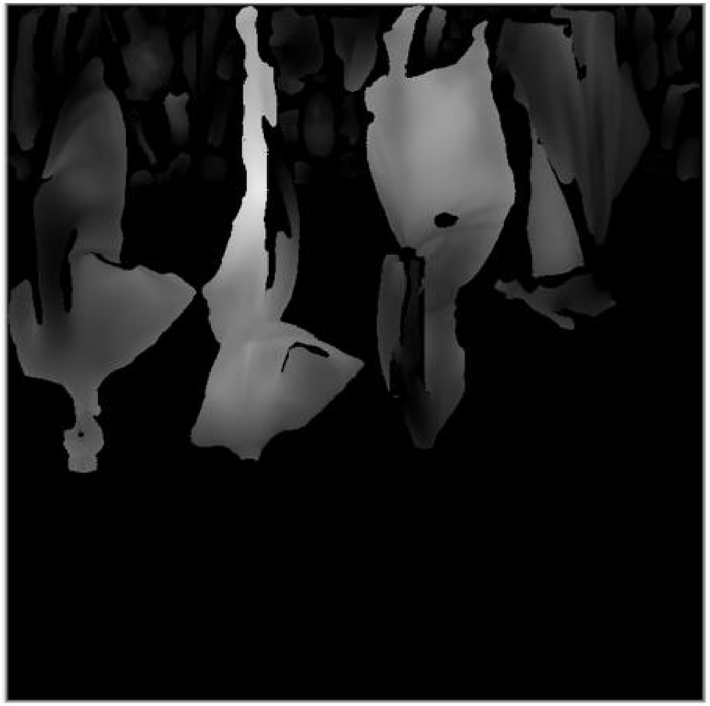
- Basic steps
  - Normal-based projection, frame packing, and frame padding
- Normal-based projection
  - Organize the points with similar normal into a patch
  - Project each patch to the 3D point cloud bounding box





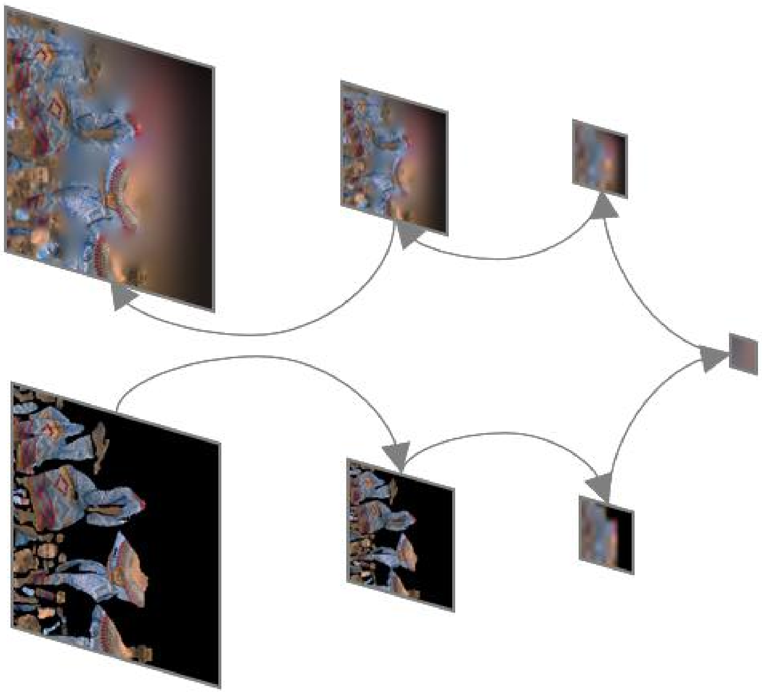
Video-based point cloud compression

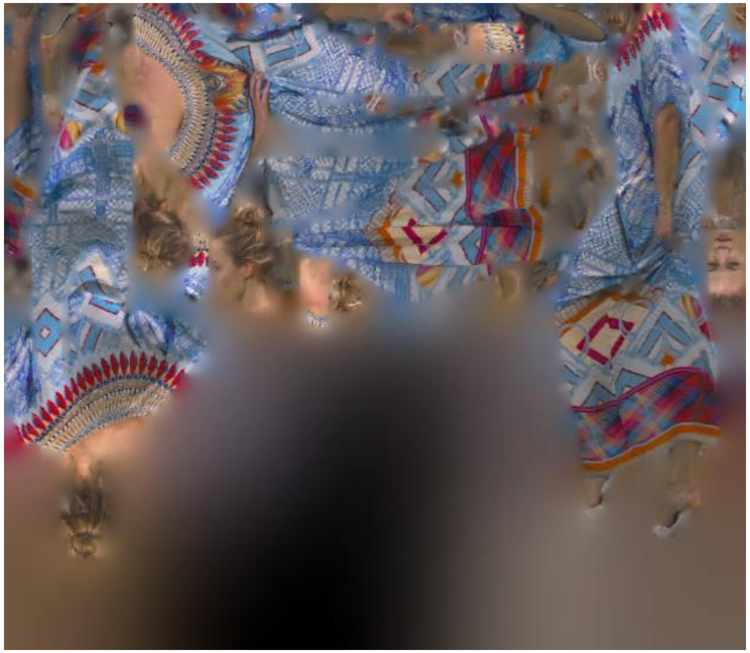
- Frame packing: pack the patches into frames
  - Exhaustive search empty space for the current patch
  - Patch rotation is supported
  - Introduced a lot of sharp edges



VPCC - Texture Padding

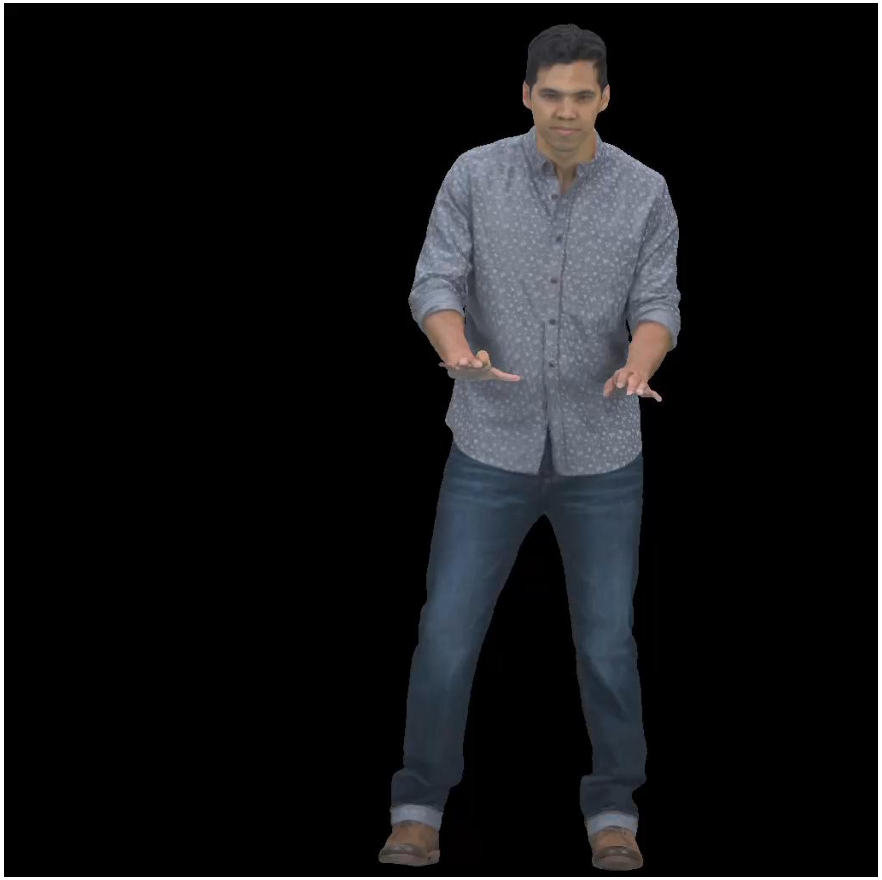
- Texture padding: a number of methods are proposed to minimize the bitrate of the unoccupied pixels
- Using push-pull algorithm as an example, like dilation



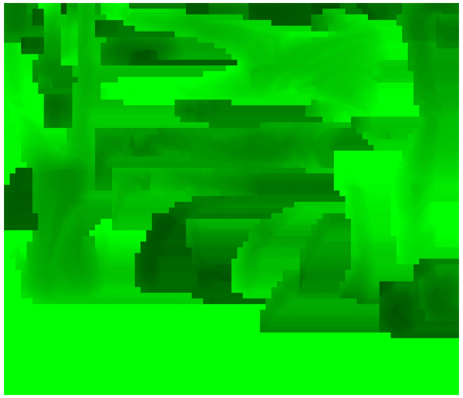


Video-based point cloud compression

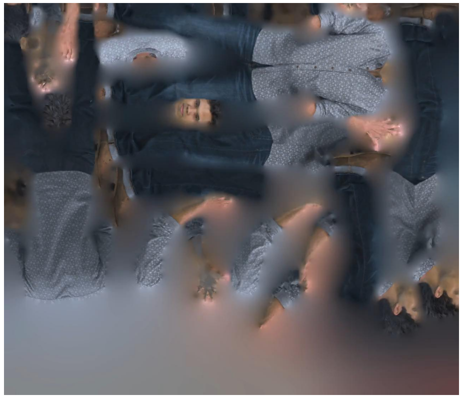
- Basic idea: project a point cloud to a 2-D video for an efficient compression



Geometry



Attribute



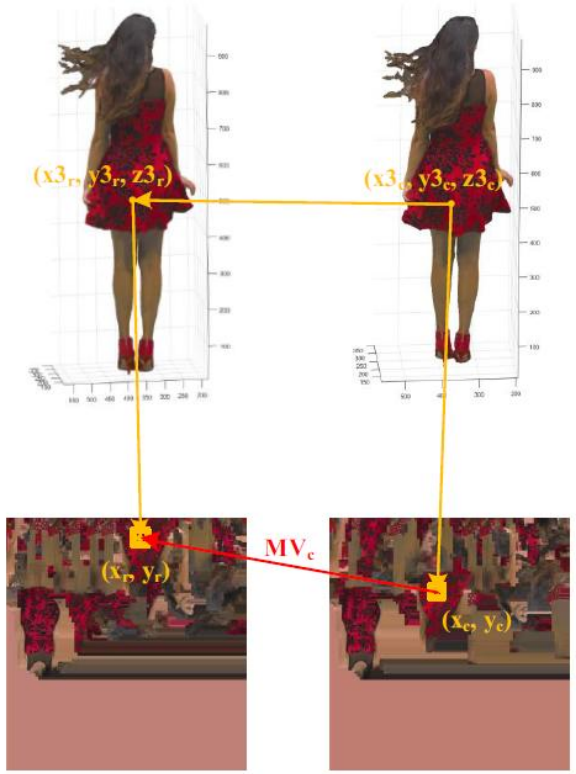
VPCC Motion Model

- The corresponding patches may be put in totally different positions in various frames (Green squares)
  - The current video codec may be unable to find a good motion vector for each block in this case
  - The geometry is encoded before the attribute, we can use the geometry to derive a better motion vector for attribute

General 3D to 2D motion model

- Given the 3D motion and the 3D to 2D correspondence, we can derive the 2D motion
  - g(), f(): 3D to 2D projection in reference and current frames

$$\boldsymbol{MV_c} = \boldsymbol{g}(x3_r, y3_r, z3_r) - \boldsymbol{f}(x3_c, y3_c, z3_c)$$

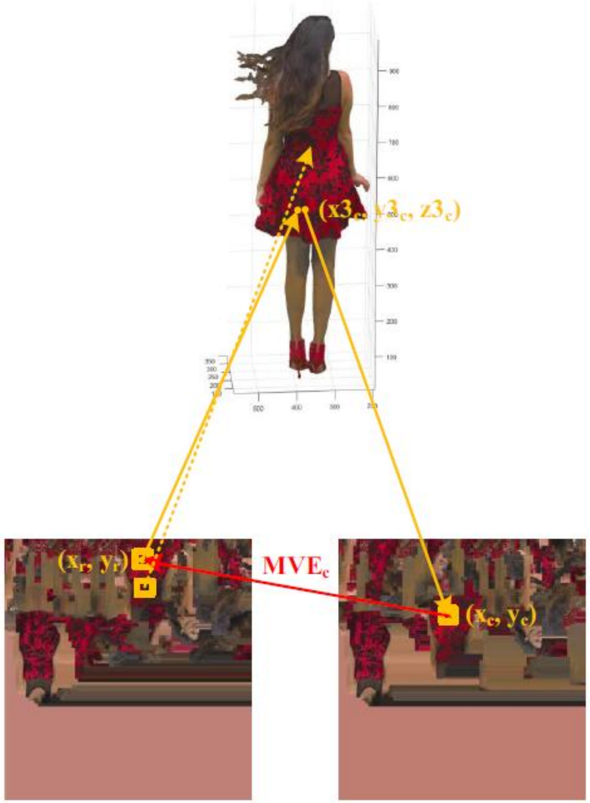


Geometry-based motion prediction

- In the V-PCC, we know the 3D-to-2D correspondence but do not know the 3D motion
- We assume the current frame and the reference frame will not change dramatically

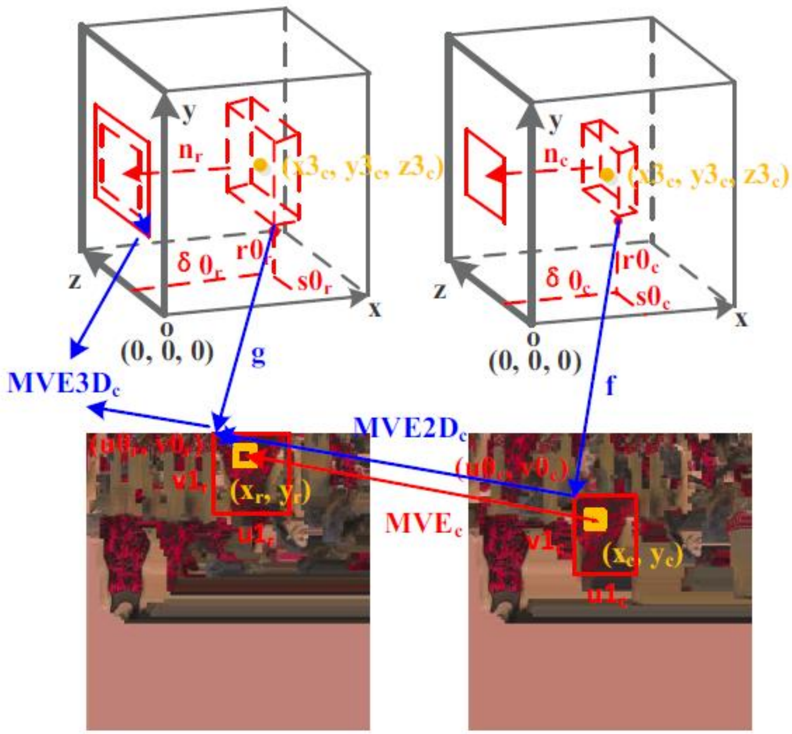
$$\boldsymbol{MVE_c} = \boldsymbol{g}(x3_c, y3_c, z3_c) - \boldsymbol{f}(x3_c, y3_c, z3_c)$$

- The problem is that (x3c,y3c,z3c) may not have a corresponding 2D point in the reference frame
  - We perform motion estimation which will increase the encoder and decoder complexity



Auxiliary information based motion prediction

- The previous method has the following two disadvantages
  - The high encoder and decoder complexity
  - It can only apply to the attribute
- The auxiliary information based motion prediction
  - The auxiliary information basically provides the coarse geometry
  - We use the 3D offset plus the 2D offset



Experiments setup

- The proposed algorithm is implemented in the V-PCC reference software and the corresponding HEVC reference software
- We test the all the dynamic point clouds defined in the common test condition including loot, redandblack, soldier, queen, longdress
- For the geometry, both point-to-point is point-to-plane are used
- For the attribute, the qualities of the luma, Cb, and Cr are considered

Experimental results on the overall scheme

- Overall scheme results

TABLE III
PERFORMANCE OF THE GEOMETRY-BASED MOTION PREDICTION COMPARED WITH THE V-PCC ANCHOR

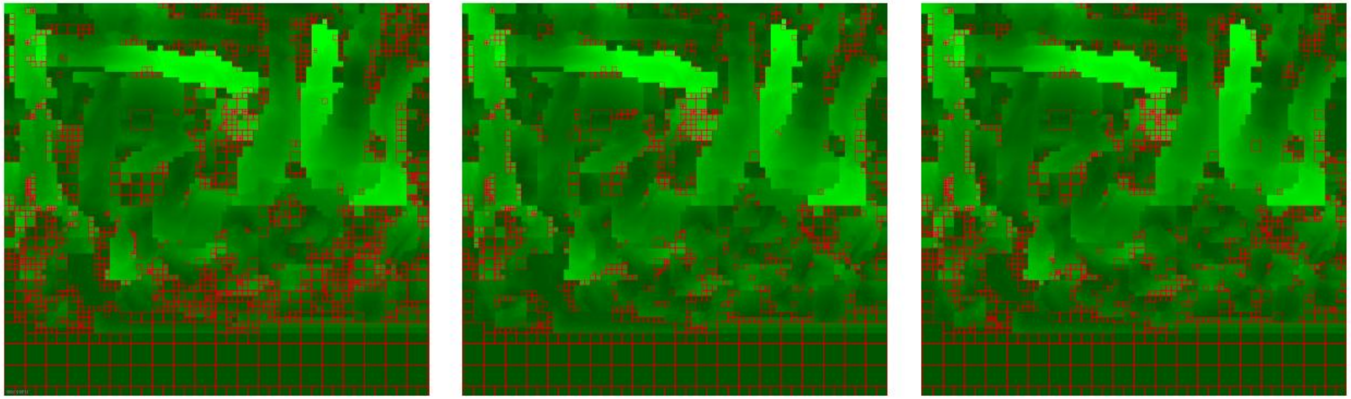
Test	Geom.BD-GeomRate		Attr.BD-AttrRate			Geom.BD-TotalRate		Attr.BD-TotalRate		
point cloud	D1	D2	Luma	Cb	Cr	D1	D2	Luma	Cb	Cr
Loot	0.0%	0.0%	−18.1%	−31.4%	−30.4%	−3.4%	−6.1%	−8.4%	−17.7%	−16.9%
RedAndBlack	0.0%	0.0%	−16.3%	−25.0%	−15.9%	−4.6%	−4.6%	−8.8%	−15.4%	−8.4%
Solider	0.0%	0.0%	−33.4%	−42.5%	−43.2%	−8.2%	−8.2%	−17.2%	−26.3%	−27.0%
Queen	0.0%	0.0%	−13.7%	−20.5%	−19.2%	−3.5%	−3.6%	−7.8%	−12.7%	−11.6%
LongDress	0.0%	0.0%	−9.8%	−13.5%	−12.3%	−3.7%	−3.7%	−6.4%	−9.5%	−8.4%
Avg.	0.0%	0.0%	−18.2%	−26.6%	−24.2%	−4.7%	−4.7%	−9.7%	−16.3%	−14.5%
Enc. time self	97%									
Dec. time self	98%									
Enc. time child	486%									
Dec. time child	337%									

TABLE IV
PERFORMANCE OF THE AUXILIARY-INFORMATION-BASED MOTION PREDICTION COMPARED WITH THE V-PCC ANCHOR UNDER THE NORMATIVE SOLUTION

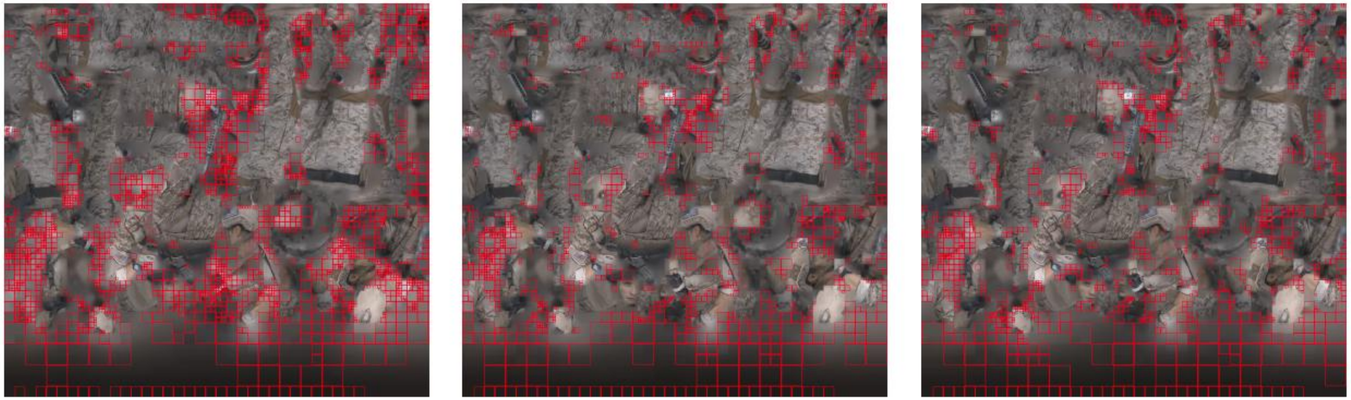
Test	Geom.BD-GeomRate		Attr.BD-AttrRate			Geom.BD-TotalRate		Attr.BD-TotalRate		
point cloud	D1	D2	Luma	Cb	Cr	D1	D2	Luma	Cb	Cr
Loot	−4.0%	−3.9%	−16.3%	−26.4%	−28.5%	−6.3%	−6.2%	−9.6%	−16.7%	−17.9%
RedAndBlack	−1.0%	−1.1%	−12.2%	−18.9%	−10.9%	−4.0%	−4.1%	−7.2%	−12.1%	−6.2%
Solider	−8.0%	−7.9%	−31.3%	−41.4%	−40.4%	−13.6%	−13.4%	−19.8%	−28.7%	−28.1%
Queen	−5.9%	−5.9%	−11.8%	−17.0%	−15.7%	−7.3%	−7.3%	−9.1%	−12.9%	−11.8%
LongDress	−1.1%	−1.1%	−8.3%	−11.2%	−10.2%	−3.8%	−3.6%	−5.7%	−8.2%	−7.3%
Avg.	−4.0%	−4.0%	−16.0%	−23.0%	−21.1%	−7.0%	−6.9%	−10.3%	−15.7%	−14.3%
Enc. time self	100%									
Dec. time self	100%									
Enc. time child	98%									
Dec. time child	99%									

Performance Analysis

- Intra blocks reduce significantly, resulting in taking adv of inter coding efficiency

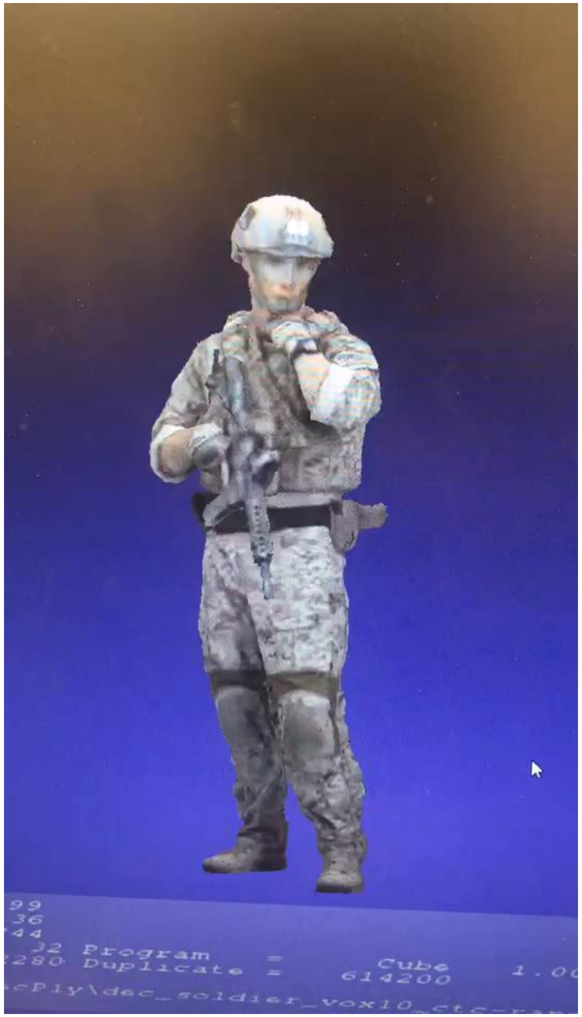


(a) Soldier Geometry Anchor (b) Soldier Geometry Normative (c) Soldier Geometry Non-normative

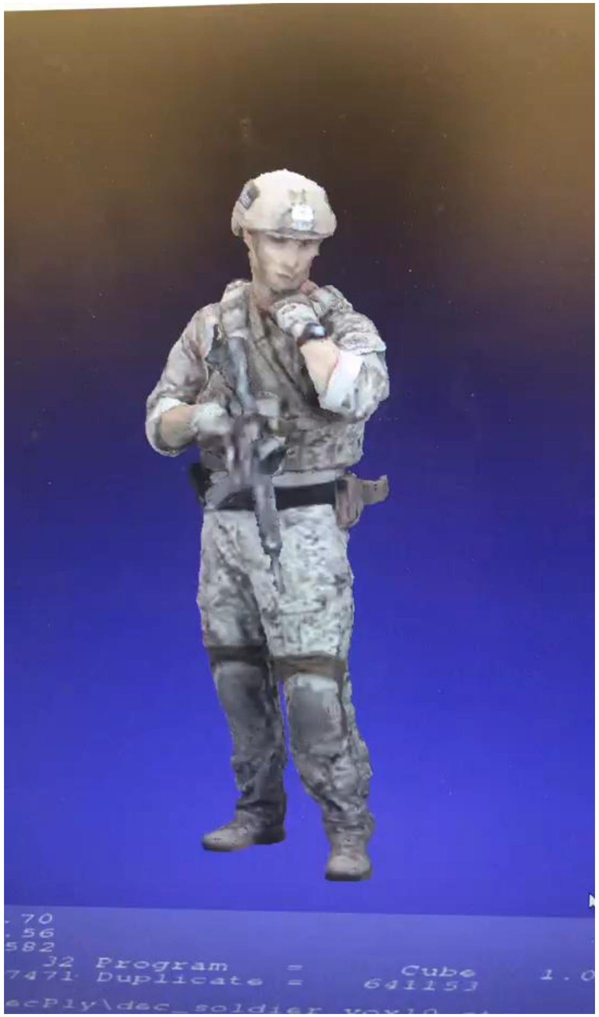


(d) Soldier Geometry Anchor (e) Soldier Geometry Normative (f) Soldier Geometry Non-normative

Subjective quality



Anchor



Proposed

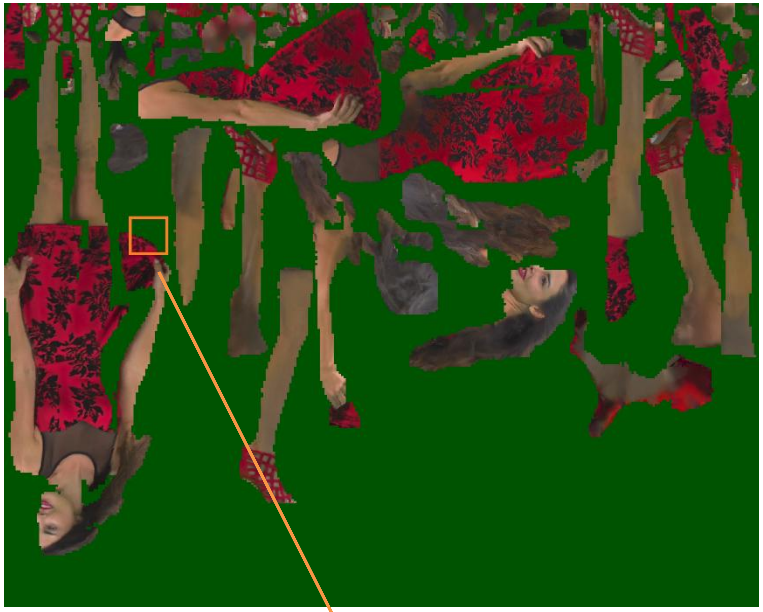
Occupancy Map Driven Rate-Distortion Optimization

- The current rate distortion optimization process in a video encoder such as HM is not handling the unoccupied pixels in a proper way

$$\min_{P} J = \sum_{i=1}^{N} D_i + \lambda R$$

$$\Downarrow$$

For a block with both occupied and unoccupied pixels, all the pixels are treated as equal importance



Proposed occupancy-map-based RDO

- The unoccupied pixels are not beneficial for the reconstructed quality of the point cloud at all

- In the proposed solution, a distortion mask is added in the RDO to handle the unoccupied pixels

$$\min_{P} J = \sum_{i=1}^{N} D_i \times M_i + \lambda R$$

where M_i is 1 when the current pixel is occupied, M_i is 0 when the current pixel is unoccupied

- This method is applied to intra/inter prediction and SAO

Intra prediction

- The RDO in intra prediction can be divided into three steps
  - INTRA Mode (Direction) Decision
    - The occupancy-map-based RDO is **not applied** as the residue bits are not counted in the bit cost

$$\min_{P} J = \sum_{i=1}^{N} SATD_i + \lambda R_{dir}$$

  - Precise mode decision and residue Quadtree decision
    - The occupancy-map-based RDO is **applied** as the residue bits are counted in the bit cost

$$\min_{P} J = \sum_{i=1}^{N} D_i \times M_i + \lambda R$$

Inter prediction

- The inter mode can be divided into merge 2Nx2N and the other inter modes
  - Merge 2Nx2N/modes comparison
    - The occupancy-map-based RDO is **applied** as the residue bits are counted in the bit cost

$$\min_P J = \sum_{i=1}^{N} D_i \times M_i + \lambda R$$

  - Other inter modes in Integer and fractional motion estimation processes or merge estimation
    - The occupancy-map-based RDO is **not applied** as the residue bits are not counted in the bit cost

$$\min_P J = \sum_{i=1}^{N} SAD_i / SATD_i + \lambda R_{motion}$$

Simulation setup

- We implement the proposed algorithm in V-PCC (TMC2-3.0) and the corresponding HEVC reference software to verify the performance of the proposed algorithm
- Follow the common test condition
  - Random access case and all intra case
- Test point cloud

Test point cloud	Frame rate	Number of points	Geometry precision	Attributes
Loot	30	~780000	10bit	RGB
RedAndBlack	30	~700000	10bit	RGB
Soldier	30	~1500000	10bit	RGB
Queen	50	~1000000	10bit	RGB
Longdress	30	~800000	10bit	RGB

Experimental results

- Random access case

Test point cloud	Geom.BD-Rate		Attr.BD-Rate		
	D1	D2	Luma	Cb	Cr
Loot	-16.3%	-16.4%	-24.3%	-18.2%	-19.3%
RedAndBlack	-6.6%	-7.2%	-12.2%	-9.8%	-12.3%
Soldier	-15.8%	-16.0%	-16.8%	-9.4%	-9.0%
Queen	-13.4%	-13.2%	-15.7%	-11.2%	-10.5%
Longdress	-7.5%	-7.8%	-7.9%	-7.7%	-7.2%
Avg.	-11.9%	-12.1%	-15.4%	-11.3%	-11.7%
Enc. self	101%				
Dec. self	99%				
Enc. child	88%				
Dec. child	88%				

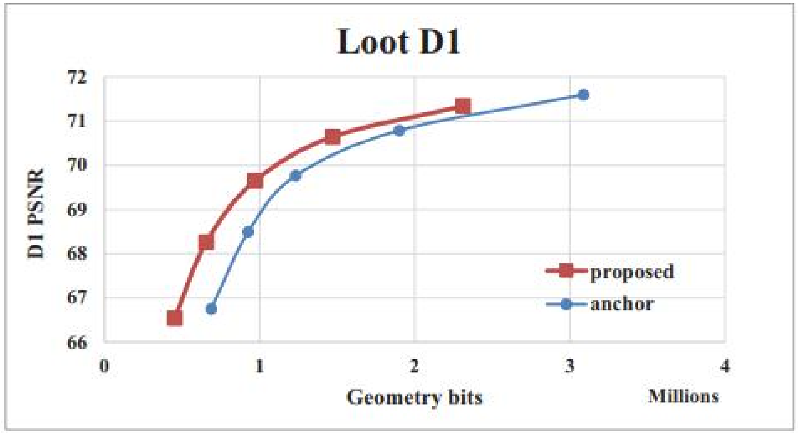
Experimental results

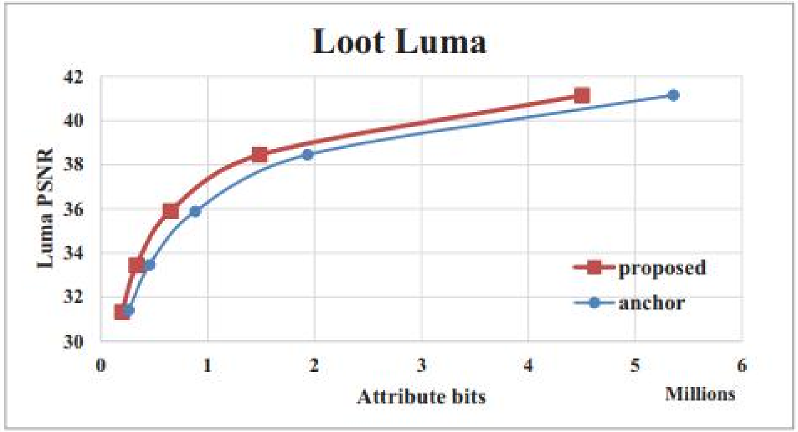
- All intra case

Test point cloud	Geom.BD-Rate		Attr.BD-Rate		
	D1	D2	Luma	Cb	Cr
Loot	-3.4%	-3.5%	-1.4%	-0.5%	-0.9%
RedAndBlack	-2.7%	-3.1%	-1.1%	-0.9%	-1.4%
Soldier	-2.9%	-3.2%	-1.1%	0.7%	1.2%
Queen	-2.6%	-2.5%	-1.2%	-1.3%	-2.0%
Longdress	-2.7%	-2.9%	-0.7%	-0.7%	-0.8%
Avg.	-2.9%	-3.0%	-1.1%	-0.5%	-0.8%
Enc. self	101%				
Dec. self	98%				
Enc. child	94%				
Dec. child	88%				

Experimental results

- Examples of R-D curves in random access case





Experimental results analysis

- Under the proposed occupancy-map-based RDO, the unoccupied pixels will be encoded with much larger distortions, and therefore we can save the bitrate

(a) Occupancy-map-based RDO

(b) Original RDO

VPCC Based Plenoptic Coding

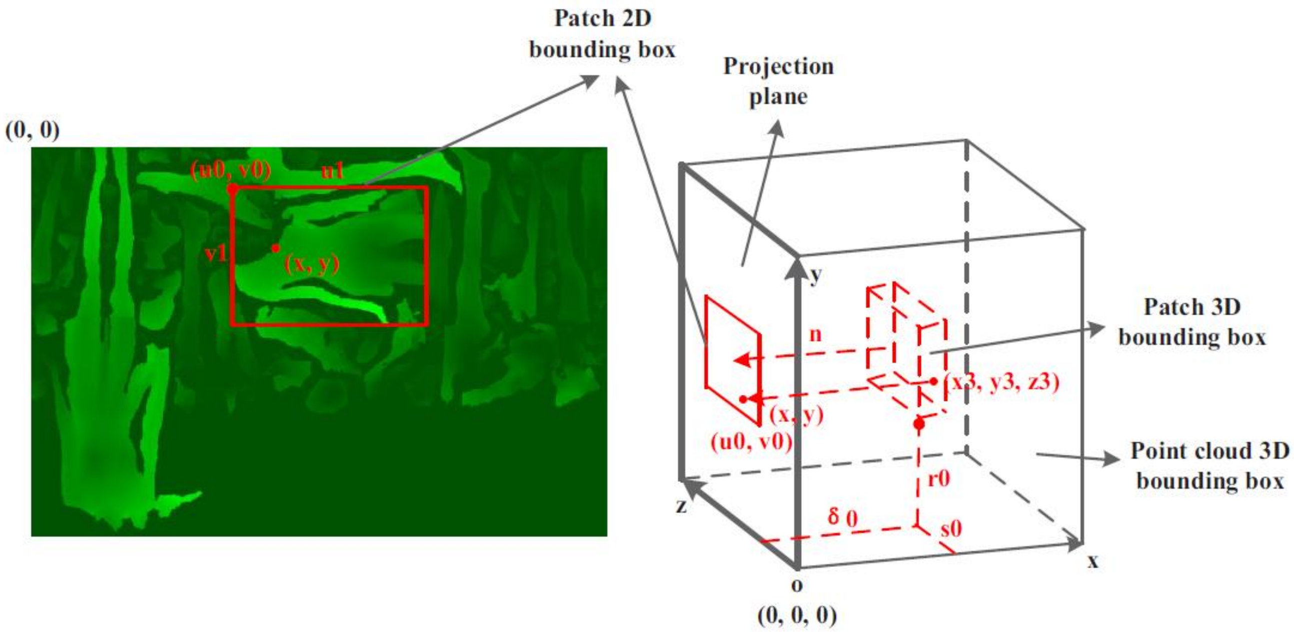
- Each point of a general point cloud is associated with one single color
  - This format is not realistic since the colors of the real world objects may vary along with the change of the view angles
  - Example, the colors of the points in the wet floor or the car surface will vary when the viewing point changes



- The plenoptic point cloud with multiple colors per point is a more complete 3-D representation and needs to be compressed efficiently, but no video-based solution yet

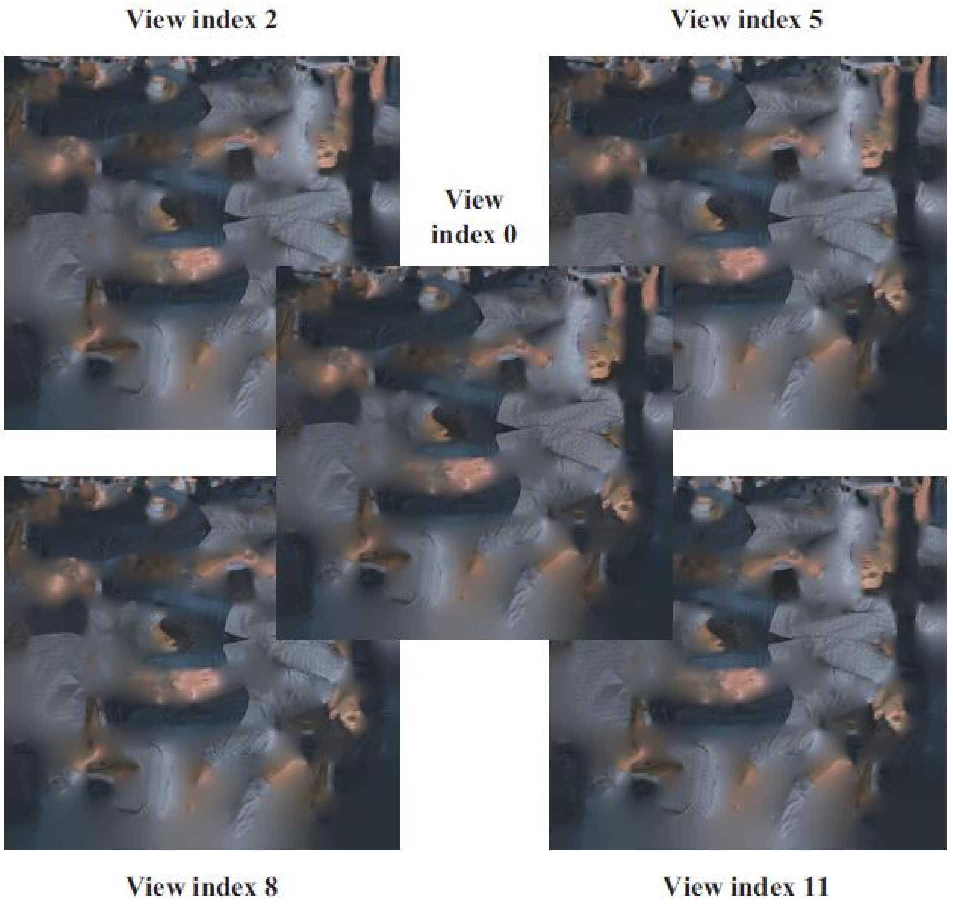
Proposed Multiview-video Compression Framework

- We first use a similar method as in V-PCC to project the plenoptic point cloud into a video
  - Patch projection; patch packing; patch padding
  - The main difference is that multiple attribute videos will be generated instead of one attribute video



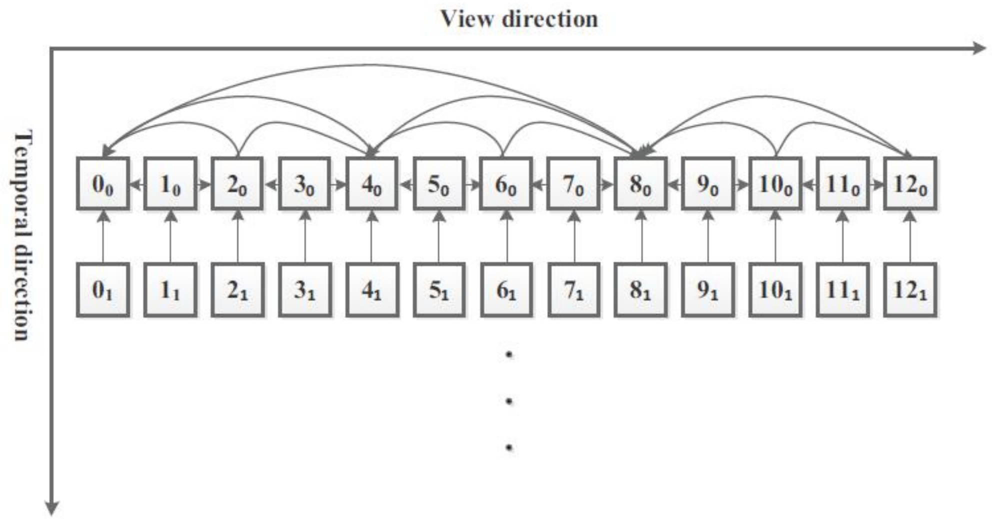
Proposed Multiview-video Compression Framework

- Examples of projected multiple attribute videos
  - Different view angles are very similar despite some pixel differences



Proposed Multiview-video Compression Framework

- Using Multiview HEVC to compress the videos efficiently
  - Encoding structures using 13 views as an example

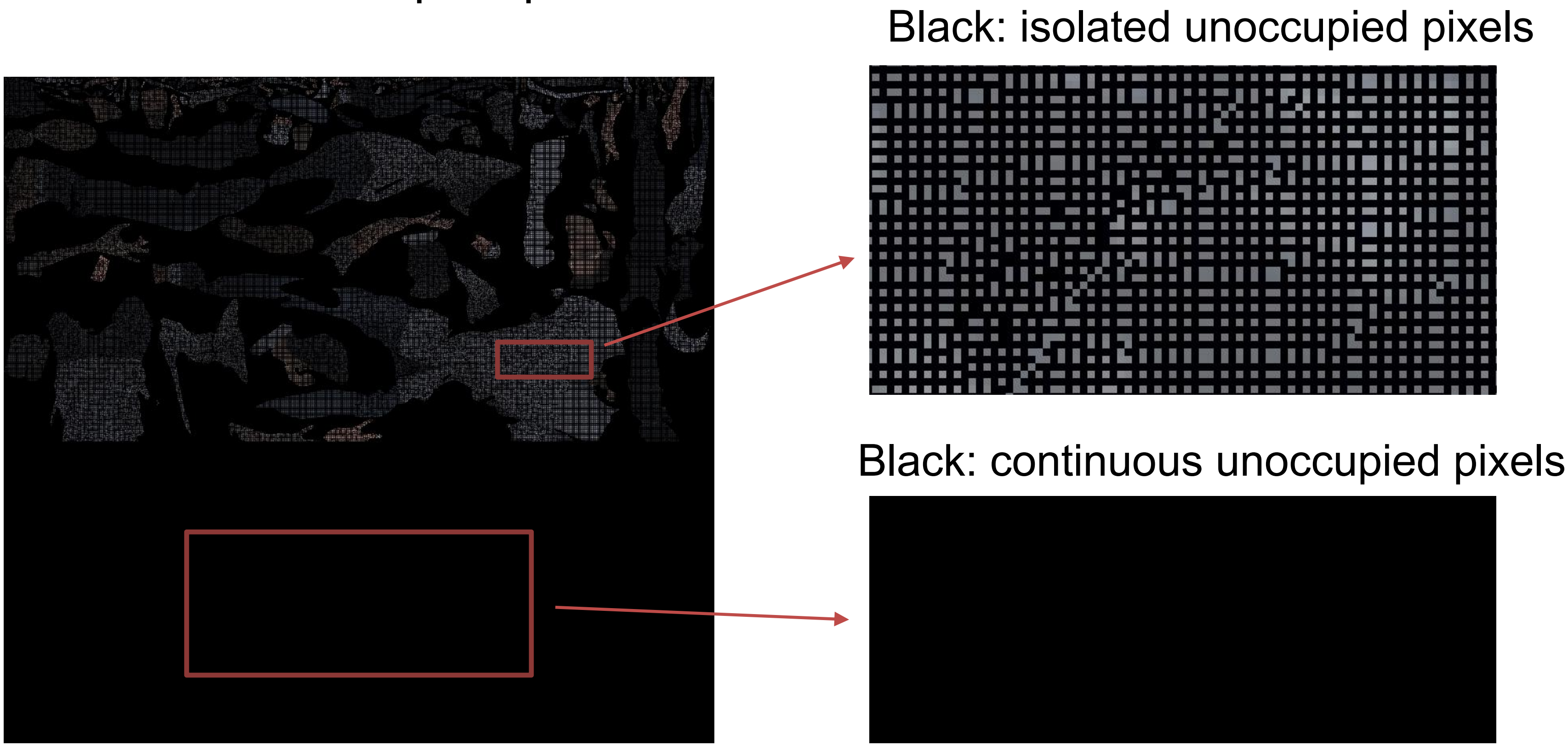


  - Bit allocation process

hierarchical level	frame 0	frame 1
0	QP_I+1	QP_I+4
1	QP_I+2	QP_I+5
2	QP_I+3	QP_I+6
3	QP_I+4	QP_I+7

Block-based group padding

- The unoccupied pixels can be divided into two groups
  - Continuous unoccupied pixels
  - Isolated unoccupied pixels



Block-based group padding

- The continuous unoccupied pixels are proposed to be padded as the average of all the unoccupied pixels across N views

$$f_{i,j} = \sum_{k=0}^{N-1} (f_{0,k} + f_{1,k})/(2N), i \in 0, 1, j \in 0, N-1$$

- The isolated unoccupied pixels should not be padded since it may destroy the spatial continuity of a block containing both occupied and unoccupied pixels

- Block-based padding decision: the unoccupied pixel is padded only when a KxK block including the current pixel as the center pixel is unoccupied

Occupancy-based RDO

- The block-based group padding can only deal with the continuous unoccupied pixels instead of the isolated ones

- The occupancy-map-based RDO is applied to handle both the continuous and isolated unoccupied pixels
  - A mask is added to the RDO when calculating the R-D cost

$$\min_P J = \sum_{i=1}^{N} D_i(P) \times M_i + \lambda R(P)$$

- This formula is applied to intra prediction, inter prediction and sample adaptive offset processes

Experimental results

- Comparison with the state-of-the-art methods
  - RAHT-KLT: **27.0%** on average
    - Significant bitrate savings in low/medium bitrate cases
    - Slight performance losses in high bitrate case
  - V-PCC: **74.4%** on average
    - Consistent performance for all the plenoptic point clouds

Name	Y	Cb	Cr
Boxer	−62.4%	−67.1%	−69.2%
Loot	−67.1%	−71.8%	−73.3%
Soldier	−73.6%	−75.1%	−76.1%
Thaidancer	−82.6%	−83.5%	−83.2%
Longdress	−86.5%	−86.6%	−86.5%
Redandblack	−78.1%	−78.1%	−79.1%
Average	−74.4%	−76.8%	−77.7%

Name	RAHT-KLT		Multiview-video		Y
	Color bits	Y-PSNR	Color bits	Y-PSNR	BD-rate
Boxer	534974	36.58	800592	37.42	1.0%
	1102667	38.51	1469168	39.37	
	2506516	41.02	2646496	41.41	
	4144398	42.77	5030344	43.40	
	7624336	45.08	9660168	45.22	
Loot	505156	36.47	639904	37.71	−31.2%
	1036214	38.57	1214984	40.35	
	2252251	41.16	2183896	42.74	
	3576056	42.91	4010464	44.80	
	6210303	45.21	7266152	46.51	
Soldier	1193244	34.15	1088120	35.37	−26.9%
	2361547	36.60	2077560	37.87	
	3514995	38.24	3741072	40.13	
	7227865	41.62	6740544	42.07	
	11973133	44.15	12000880	43.78	
Thai	434126	28.46	515368	31.29	−42.7%
	1719585	33.63	959752	34.09	
	3058823	36.63	1744408	36.67	
	4292715	38.52	3242152	38.86	
	5599587	40.03	6150648	40.89	
Long	519371	28.01	942000	33.03	−35.4%
	2081546	33.01	1639144	35.37	
	3770193	36.19	2798792	37.40	
	5245716	38.36	4972816	39.13	
	9214122	42.67	9176816	41.17	
Red	224020	31.82	744512	36.61	−16.9%
	903125	35.90	1305184	38.80	
	1736193	38.43	2203496	40.73	
	3313844	41.59	3970184	42.44	
	6081458	45.08	7318392	44.05	
Average	–	–	–	–	−27.0%

Experimental results

- Block-based group padding logic (isolated patch not padded)
  - An extra 13.3% performance improvements on average

Name	Y	Cb	Cr
Boxer	–18.7%	–13.8%	–16.5%
Loot	–16.5%	–15.7%	–15.0%
Soldier	–9.6%	–7.7%	–7.4%
Thaidancer	–13.3%	–12.2%	–12.6%
Longdress	–8.1%	–8.2%	–8.2%
Redandblack	–13.6%	–13.6%	–14.1%
Average	–13.3%	–11.5%	–13.6%

  - Different influences of block size K
    - K = 4 shows the best R-D performance on average

Name	Y BD-rate				
	K = 1	K = 2	K = 4	K = 8	K = 16
Boxer	105.2%	–17.9%	–18.7%	–18.8%	–18.0%
Loot	62.2%	–16.0%	–16.5%	–16.1%	–15.3%
Soldier	60.1%	–9.2%	–9.6%	–9.2%	–8.1%
Thaidancer	–13.4%	–13.3%	–13.3%	–12.8%	–11.9%
Longdress	–1.6%	–8.3%	–8.1%	–7.7%	–7.2%
Redandblack	64.4%	–12.6%	–13.6%	–13.6%	–12.7%
Average	42.5%	–13.0%	–13.3%	–12.9%	–12.1%

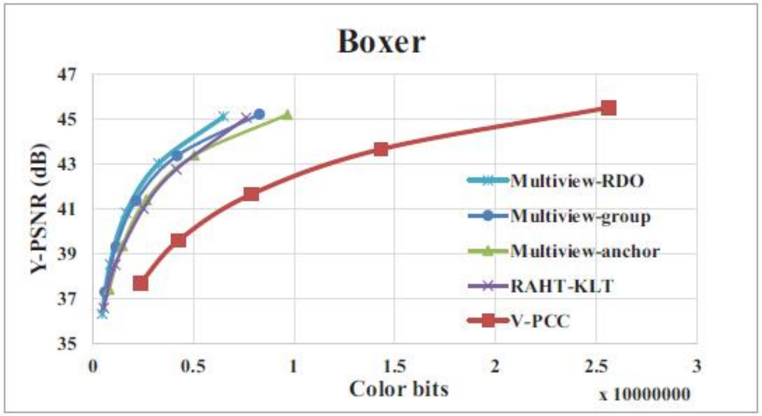
Experimental results - Extra gain

- Occupancy-map-based RDO compared with MV-HEVC
  - 19.5% on average

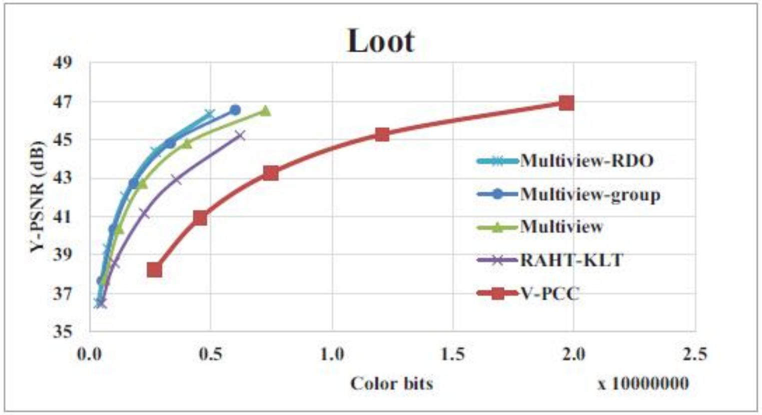
Name	Y	Cb	Cr
Boxer	–26.4%	1.6%	2.4%
Loot	–21.0%	–1.1%	3.5%
Soldier	–16.7%	13.2%	13.6%
Thaidancer	–15.8%	–14.8%	–14.8%
Longdress	–17.6%	–7.6%	–7.9%
Redandblack	–23.1%	–15.4%	–18.6%
Average	–19.5%	–1.7%	–0.5%

Experimental results

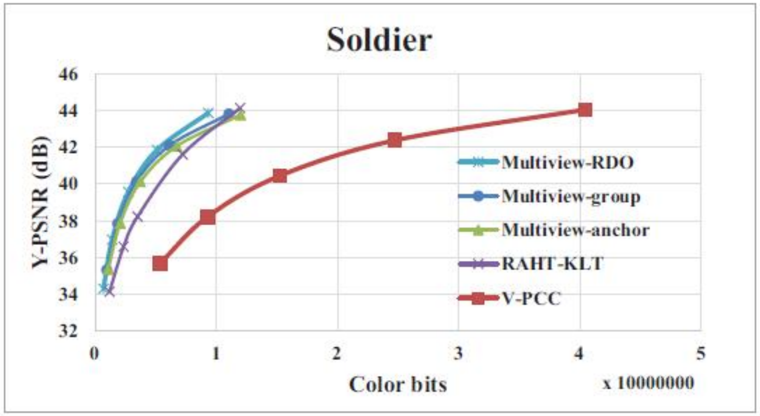
- R-D curves



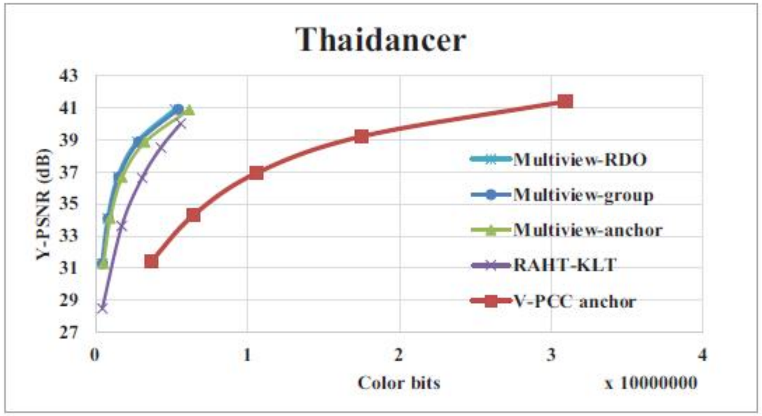
(a)



(b)



(c)



(d)

VPCC Work Summary

- For 3D sensing/Auto-driving, geometry is the key, BTQT is a good framework with room for entropy coding optimization (LSTM), and RDO
- vPCC deals with immersive content, current MPEG vPCC has many in-efficiency, we introduced advanced motion model, occupancy map based RDO to significantly improve the over all performance
- Introduced plenoptic (multi-attributes) point cloud coding with light-field like coding scheme, inter-view prediction that yields very good results.

Summary

Recent papers,

- S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. César, P. A. Chou, R. A. Cohen, M. Krivokuca, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. J. Tabatabai, A. M. Tourapis, V. Zakharchenko: Emerging MPEG Standards for Point Cloud Compression. *IEEE J. Emerg. Sel. Topics Circuits Syst*. 9(1): 133-148 (2019)
- L. Li, Z. Li, S. Liu, H. Li, "Occupancy-map-based rate distortion optimization for video-based point cloud compression", *IEEE Int'l Conf on Image Processing* (ICIP), 2019.
- A. Akhtar, B. Kathariya, Z. Li, "Low Latency Scalable Point Cloud Communication", *IEEE Int'l Conf on Image Processing* (ICIP), 2019.
- A. Akhtar, J. Ma, R. Shafin, J. Bai, L. Li, Z. Li, and L. Liu, "Low Latency Scalable Point Cloud Communication in VANETs using V2I Communication", *IEEE Int'l Conf on Communication* (ICC), Shanghai, 2019.
- L. Li, Z. Li, V. Zakharchenko, and J. Chen, "Advanced 3D Motion Prediction for Video Based Point Cloud Attributes Compression", *IEEE Data Compression Conf* (DCC), Snowbird, USA, 2019.
- B. Kathariya, L. Li, Z. Li, and J. Alvarez, "Scalable Point Cloud Geometry Coding with Binary Tree Embedded Quadtree", *IEEE Int'l Conf. on Multimedia & Expo* (ICME) ,San Diego, USA, 2018.
- B. Kathariya, V. Zakharchenko, Z. Li, and J. Chen, "Level-of-detail generation using binary-tree for lifting scheme in LiDAR point cloud attributes coding" , *IEEE Data Compression Conf* (DCC), Snowbird, USA, 2019.
- B. Kathariya, L. Li, Z. Li, and J. Alvarez, "Lossless Dynamic Point Cloud Geometry Compression with Inter Compensation and Traveling Salesman Prediction", *IEEE Data Compression Conference* (DCC) , Snow Bird, 2018.
- Y. Shao, Q. Zhang, G. Li and Z. Li, "Hybrid Point Cloud Attribute Compression Using Slice-based Layered Structure and Block-based Intra Prediction" , *ACM Multimedia* , Seoul, Korea, 2018.

- Q & A