LEARNING CONVOLUTIONAL TRANSFORMS FOR LOSSY POINT CLOUD GEOMETRY COMPRESSION

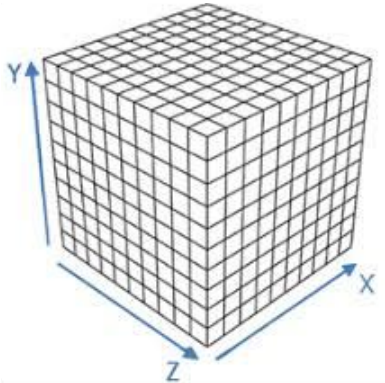
Maurice Quach, Giuseppe Valenzise and Frederic Dufaux
L2S, CNRS, CentraleSupelec, Université Paris-Saclay

AGENCE NATIONALE DE LA RECHERCHE
ANR
ID: REVERY ANR-17-CE23-0020

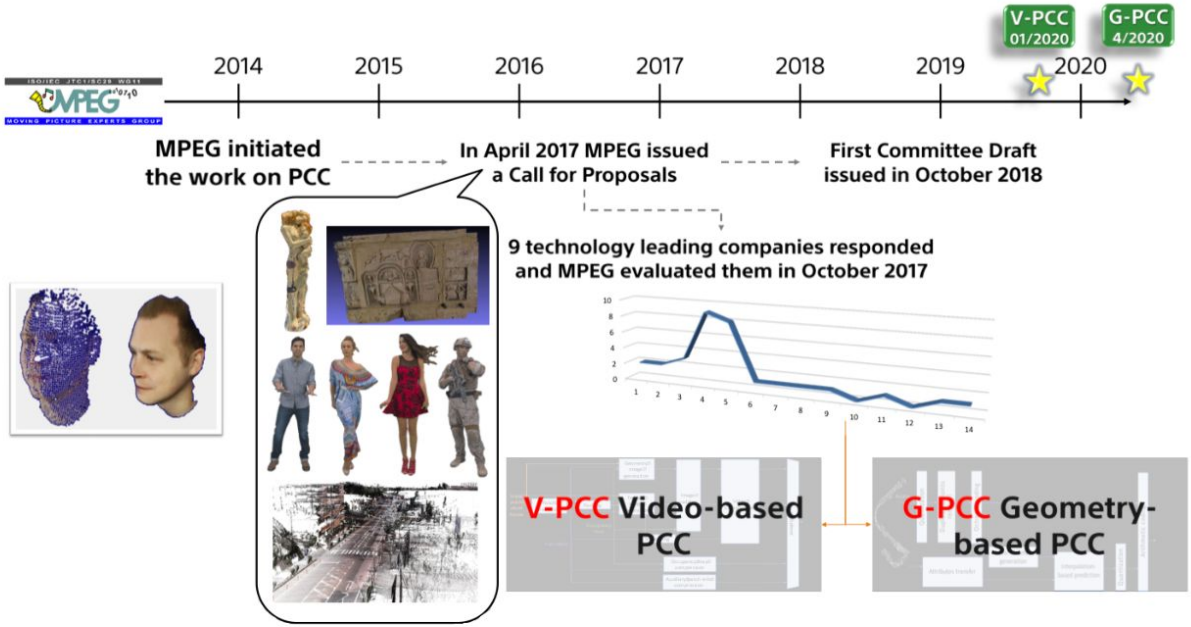
L2S
Laboratoire des Signaux & Systèmes
cnrs
CentraleSupélec
UNIVERSITE PARIS SUD
université PARIS-SACLAY

Point Cloud

- General definition: set of (x, y, z) elements
- Point attributes: colors, normals…
- Voxelized point cloud/Voxel grid at resolution r:



MPEG PCC

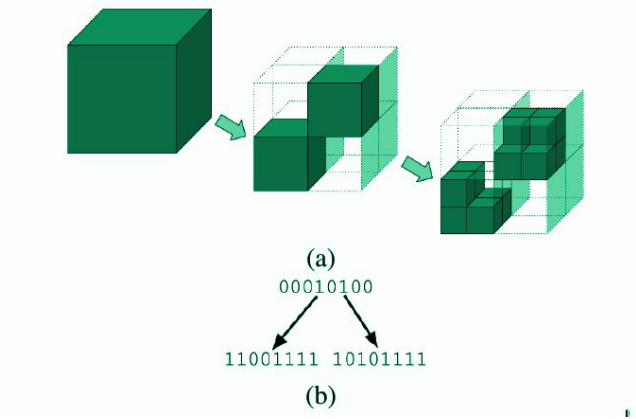


from Danillo Graziosi

Point Cloud Compression

Geometry

- Mainly based on octrees



Diogo C. Garcia et al., ICIP (2018)

Attributes

- Graph Transform
- Region-Adaptive Hierarchical Transform (RAHT) ~ Haar wavelets
- Gaussian Process Transforms
- Mapping to a 2D plane
- …

Model Architecture

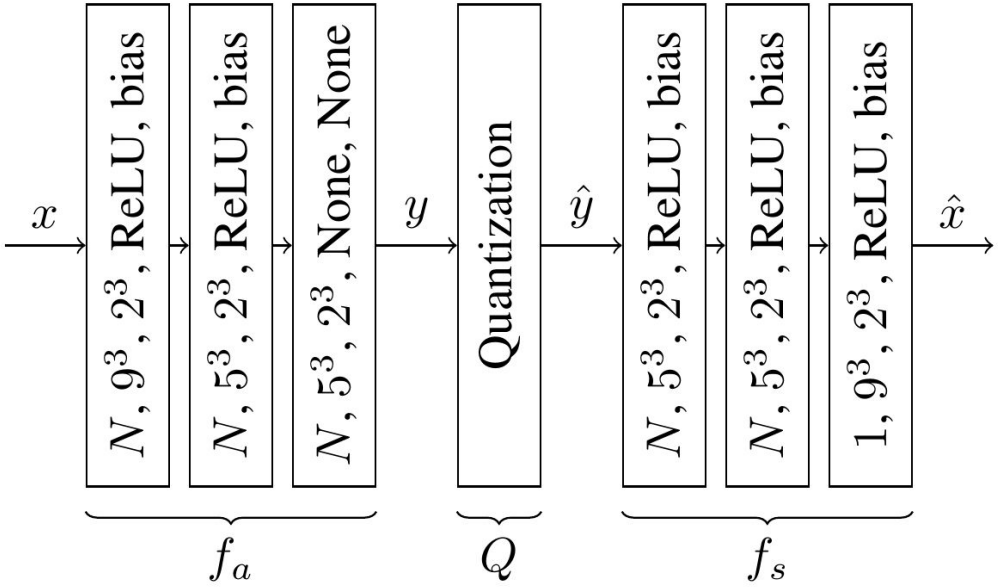
Transforms:

- *Analysis* f_a
- *Synthesis* f_s

Quantization:

- Training: *uniform noise*
- Evaluation: *quantization*

Lightweight model



number of feature maps, filter size, strides, activation and bias

Loss Function

- **Distortion**:
  Focal loss to compensate for class imbalance
- **Rate**:
  Entropy of the quantized latent space / number of occupied voxels (bpov)
- Decoding can be interpreted as binary classification on the voxel grid

$$FL(p_z^t) = -\alpha_z(1 - p_z^t)^\gamma log(p_z^t)$$

$$L = \lambda D + R$$

(inverted formulation because of scale)

Methodology

- Training on **ModelNet40**: mesh dataset with 12,311 models and 40 classes
  Random sampling at resolution *64*

- Evaluation on **Microsoft Voxelized Upper Bodies**: point cloud dataset with 5 sequences, 30 fps (7 to 10 seconds), total of 1202 frames at resolution ***512***

Evaluation

- Baseline (Rufael Mekuria et al., TCSVT 2017): Octree coding
- Quality metric:
  Point to plane symmetric PSNR
  (Dong Tian et al., ICIP 2017)
- Bitrates (Deflate algorithm and range coding) in bits per occupied voxel (bpov)

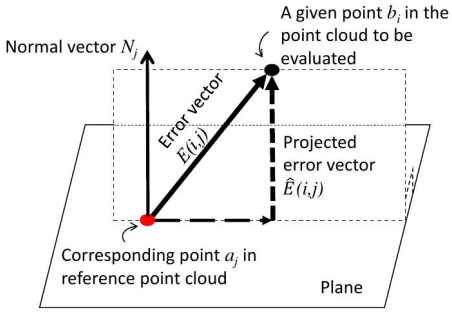
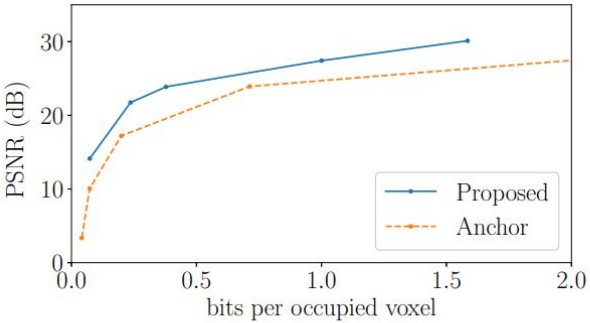


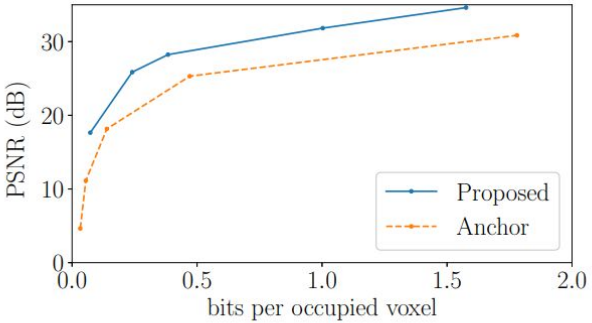
Fig. 1: Point-to-point distance vs. point-to-plane distance.

$$e_{symm}(A,B) = \min(e(A,B), e(B,A))$$

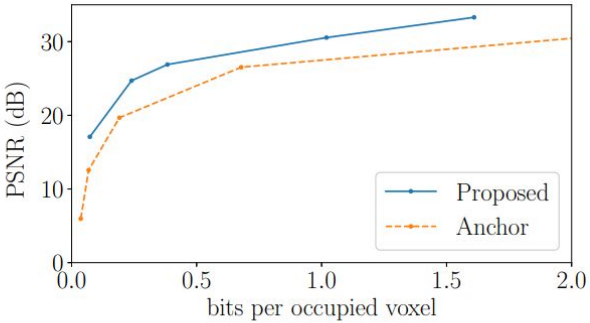
Quantitative results



(a) Andrew sequence (−47.8% BDBR)



(b) David sequence (−55.7% BDBR)



(c) Phil sequence (−49.0% BDBR)

Qualitative results

- Original point cloud
- 370,798 points

Qualitative results

- Compressed point cloud using our method (λ = 10e-6)
- 1,302,027 points
- p2plane PSNR: 29.22 dB
- 0.071 bpov
- Some holes are filled

Qualitative results

- Octree compressed point cloud (Depth 6)
- 5,963 points
- p2plane PSNR: 23.98 dB
- 0.058 bpov

Conclusion

- **51.5%** Bjontegaard-delta bitrates (BDBR) savings in average
- **No exponential diminution** in the number of points at low bitrates
- **Efficient** learning based point cloud compression
- **Quality metrics** can be further improved to better reflect perception

Code and supplementary material

Point Cloud	msft/phil9/ply/frame0000.ply
Number of points	370798

Point Cloud	msft_dec_000001/phil9/ply/frame0000.ply.bin.pl
Number of points	1302027
Bits per occupied voxels	0.07072314305902405
A->B p2point PSNR (dB)	41.7388
B->A p2point PSNR (dB)	27.9759

Point Cloud	cwi-pcl-codec-samples/msft_6/phil9/ply/frame0
Number of points	5963
Bits per occupied voxels	0.05825274138479713
A->B p2point PSNR (dB)	19.2723
B->A p2point PSNR (dB)	22.5908

- Source code: https://github.com/mauriceqch/pcc_geo_cnn
- Samples: https://www.mauricequach.com/pcc_geo_cnn_samples/

Questions ?

Additional Slides

Complexity

- Specifications:
  - CPU: Intel(R) Xeon(R) Bronze 3106 CPU @ 1.70GHz CPU
  - GPU: Nvidia GeForce GTX 1080 Ti
- Our method: (resolution 512)
  - About 0.5s for compression (GPU)
  - About 1m20s for decompression (using CPU as GPU implementation seems to have issues on high resolutions)
- Octree:
  - Under 5s

Complexity

Performance bug for CONV3D_TRANSPOSE on GPU for resolution 512 so CPU is faster