From capture to immersive viewing of 3D HDR point clouds

C. Loscos [1] and P. Souchet [2] and T. Barrios [1] and G. Valenzise [3] and R. Cozot [4]

[1]University of Reims Champagne-Ardenne, France
[2]XD Productions, France
[3]Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France
[4]University of Littoral Côte d'Opale, IMAP Research Group / LISIC Laboratory, France

Figure 1: The ReVeRY project.

Abstract

The collaborators of the ReVeRY project address the design of a specific grid of cameras, a cost-efficient system that acquires at once several viewpoints, possibly under several exposures and the converting of multiview, multiexposed, video stream into a high quality 3D HDR point cloud. In the last two decades, industries and researchers proposed significant advances in media content acquisition systems in three main directions: increase of resolution and image quality with the new ultra-high-definition (UHD) standard; stereo capture for 3D content; and high-dynamic range (HDR) imaging. Compression, representation, and interoperability of these new media are active research fields in order to reduce data size and be perceptually accurate. The originality of the project is to address both HDR and depth through the entire pipeline. Creativity is enhanced by several tools, which answer challenges at the different stages of the pipeline: camera setup, data processing, capture visualisation, virtual camera controller, compression, perceptually guided immersive visualisation. It is the experience acquired by the researchers of the project that is exposed in this tutorial.

CCS Concepts

• ***Computing methodologies*** $\rightarrow$ ***Computational photography; Image processing; Virtual reality; Perception; 3D imaging;***

1. Introduction

In the last two decades, industries and researchers proposed significant advances in media content acquisition systems in three main directions: increase of resolution and image quality with the new ultra-high-definition (UHD) standard that uses 3840x2160 pixels resolution (also called 4K resolution); stereo capture for 3D content (depth information); and high-dynamic range (HDR) imaging raising the dynamic range of the image to at least 16-fstops. These recent advances addressed the full media production pipeline: acquisition, image data enhancement, and display, with the development of 3D and grid cameras, HDR imaging, UHD resolution, autostereoscopic displays, immersive VR headsets, HDR displays. These new technologies raise incontestable enthusiasm by both professionals and end users, but are currently limited by low creative content potential. For instance, todays offered 360° panoramic image for VR immersive visualization would not be convincing for a natural light outdoor landscape. The user would be perceptually limited in the range of intensity and restricted to rotating navigation. Among other objectives, the ReVeRY project wants to address solutions to enable user perception of high intensity ranges as well as free navigation inside the scene in an embedded distributed media adaptive to the diversity of nowadays displays. In other words, there should be no capability difference when virtually visualizing real or synthetic scenes. The ReVeRY project has conducted fundamental research to address the full pipeline from acquisition to display. Its aims are to answer to currently known limitations:

1. Rig capture still presents major chalenges, both in terms of equipment set up and data flow management,
2. Depth and HDR content is now predominant in many applications but higher resolution shouldn't be neglected,

DOI: 10.2312/egt.20221054

3. Compression, representation, and interoperability of these new media are active research fields in order to reduce data size and to be perceptually accurate.
4. Displaying such content on current restitution equipment needs adapted solutions.

This tutorial presents a complete pipeline to create 3D immersive content from a grid of production cameras. It summarizes the work produced for 4 years in a french funded multi-partner project, the ANR ReVeRY project. It is the experience acquired by the researchers of the project that is exposed in this tutorial. The pipeline is complete, from the camera set up to immersive viewing through data processing, content creation and perceptually-driven encoding.

2. Speakers

Tutorial organizer:

- **Céline Loscos**, LICIIS laboratory, University of Reims Champagne-Ardenne, *celine.loscos@univ-reims.fr*, https://cv.archives-ouvertes.fr/celine-loscos
Cèline Loscos has been a Professor of computer science at University of Reims Champagne-Ardenne since 2010. She obtained her PhD in computer science at Joseph Fourier University (Grenoble, France) in 1999. After a postdoctoral fellowship (2000-2001) at University College London, United Kingdom, she was appointed lecturer. In 2007, she joined the University of Girona, Spain. She conducts her research in the LICIIS laboratory. Her research topics focus on computational photography, 3D imaging, and virtual reality. She is the coordinator of the ANR ReVeRy project (2017-2022).

Other speakers in presenting order:

- **Philippe Souchet**, XD Productions, *philippe.souchet@xdprod.com*, https://www.xdprod.com/
Philippe Souchet has been Chief Technology Officer at XD Productions since 1999. He got an MSc in computer vision at Paris VII Jussieu in 1993. As a former game developper for Sony Psygnosis between 1994 and 1999, he participated in the first soccer simulations using motion capture for the video games series "Adidas Power Soccer". He leads Research Developpemnt efforts of XD Productions in markerless motion capture, 3D reconstruction and volumetric capture, along with their dissemination in the broadcast industry, XD also being a producer of TV Shows and Motion Pictures.
- **Giuseppe Valenzise**, Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, *giuseppe.valenzise@l2s.centralesupelec.fr*, https://l2s.centralesupelec.fr/u/valenzise-giuseppe/
Giuseppe Valenzise is a researcher at the Centre National de la Recherche Scientifique (CNRS) in the Laboratoire des Signaux et Systèmes, CentraleSupelec, University Paris-Saclay, France. He completed a Ph.D. in Information Technology at the Politecnico di Milano, Italy, 2011. From 2012 to 2016 he was with the Laboratoire Traitement et Communication de l'Information (LTCI) of Telecom Paristech. He got the French "Habilitation à diriger des recherches" from Université Paris-Sud in 2019. His research interests span different fields of image and video processing, including traditional and learning-based image and video compression, light fields and point cloud coding, image/video quality assessment, high dynamic range imaging and applications of machine learning to image and video analysis. He is co-author of more than 100 research publications and of several award-winning papers. He is the recipient of the EURASIP Early Career Award 2018. Dr. Valenzise serves as Associate Editor for IEEE Transactions on Image Processing as well as for Elsevier Signal Processing: Image communication. He was program co-chair of the EUVIP 2021 conference. He is a member of the MMSP and IVMSP technical committees of the IEEE Signal Processing Society, as well as a member of the Technical Area Committee on Visual Information Processing of EURASIP.
- **Théo Barrios**, LICIIS laboratory, University of Reims Champagne-Ardenne, *theo.barrios@univ-reims.fr*
Théo Barrios has been a PhD student at University of Reims Champagne-Ardenne since 2018. He obtained a Master Degree in Computer Science and Applied Mathematics at ENSEEIHT enigneering school. His Master project covered room mapping from LiDAR point clouds. His PhD research topic is on 3D reconstruction from color images from camera arrays.
- **Rémi Cozot**, University of Littoral Côte d'Opale, IMAP Research Group / LISIC Laboratory, *remi.cozot@univ-littoral.fr*, http://cozot.free.fr/
Rémi Cozot is a full professor at the University of Littoral Opal Cost located in Calais, France. Before that, he completed a PhD from the University of Rennes in 1996. He got an associate professor position at the University of Rennes in 1997, until 2019. His research focusses on image appearance modeling, visual perception, image aesthetic, and especially style/aesthetic aware HDR image processing. He has been involved in many french national projects and European projects in the field of HDR image processing and visual perception. He is the associated editor of the visual computer journal.

3. Tutorial details

3.1. Keywords

This tutorial frontiers 3D vision, data compression, and computer graphics.

3.2. Tutorial length

We proposed a half-day tutorial, with four presentations of 45-minute each.

3.3. A detailed outline of the tutorial

The tutorial is composed of four parts, each part presenting a step of the pipeline, going from acquisition to display. Each part is planned for 45 minutes.

1. **Camera grid setup and camera controller** - *speaker:speakers: P. Souchet*.

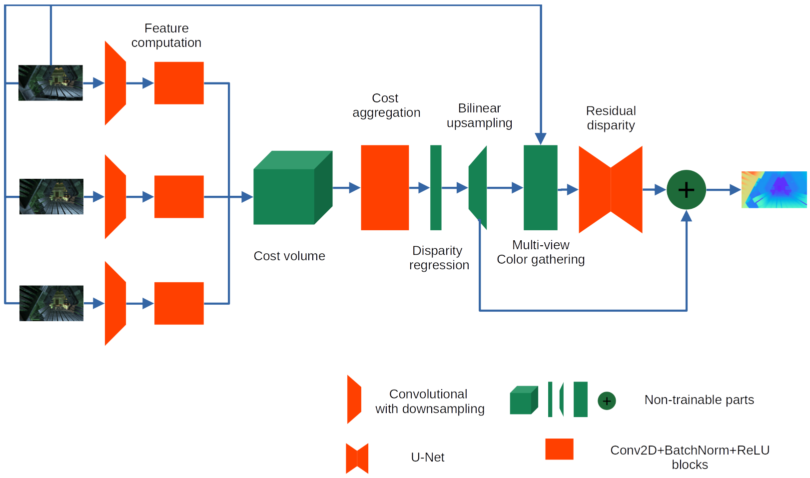


Figure 2: Pipeline used to reconstruct a 3D point cloud from camera grid pictures [BGPL22].

a. **Multi-view, multi-exposure camera grid**
The role of XD Productions, as industrial partner with a long experience of multiview capturing systems, was to specify, design and build the prototype of a grid of 4x4 UHD cameras, allowing real time 3D Reconstruction of HDR point clouds from synchronized multi-exposed video streams (see Figure 3).
The images can be processed in real time or recorded on disk for more complex algorithms, demanding a lot of processing power along with important storage and network bandwidth. Therefore, the system is composed of several acquisition units, linked to one multi GPU computing unit. The units communicate through 10GB ethernet connections, to allow the transfer of 16 4K-video streams in real time.

Figure 3: Camera and camera capture setup.

b. **Controlling software** The development of the software layer was designed to allow each partner to add its personal brick, best fitting its needs. Thus, a modular architecture was chosen, allowing easy testing of different algorithms and rendering techniques, and greater adaptability to coming states of the art.
The main modules of the REVERY software include:

- display of the 16 video streams (see Figure 4),
- remote control of the camera (for parameters such as gamma, zoom, focus, exposure, ...),
- camera calibration,
- rectification,
- 3D interactive rendering of resulting point clouds.

Figure 4: Display of 16 multi-exposed, video streams.

2. **3D HDR content reconstruction** - *speakers: C. Loscos and T. Barrios*.
In this part, we will expose advances in depth reconstruction from grid of cameras, HDR reconstruction for single and multiple view, and how it combines to produce a 3D HDR point cloud. Recent advances show that machine learning, like [KFR*18], helps robustly producing 3D point clouds. We show that it is possible to extend the concept to camera grid with large baselines [BGPL22] (see Figure 2. We specifically address camera grid configuration, and the challenges associated to large baselines. We review previous work on HDR imaging, especially those combing depth and HDR reconstruction [BLV*12] [BVL19] [OLMA13], and more recent machine learning-based approaches which need only one image as an input to generate an HDR image [EKD*17] [SRK20] and can be adapted to multiple views [MZCL22]. Examples of results are shown in Figure 5.
3. **3D point cloud coding and quality assessment** - *speaker: G. Valenzise*.
We present the state-of-the-art coding methods for point clouds,

Figure 5: HDR reconstruction results after machine learning from one view of [EKD*17], [SRK20], and [MZCL22] compared with the reference on the left hand side (LDR and HDR images).

and in particular the new MPEG G-PCC and V-PCC standards [CPZ*21], as well as recently proposed learning-based compression approaches [QVD, QVD20, NQVD21]. The latter have been shown to provide substantial coding gains compared to conventional methods, see Figure 6. We will then discuss briefly how to assess the quality of compressed point clouds, from simple distance metrics for geometric distortion [T*17] to more recent data-driven approaches [CQVD21, QCVD21].

(a) Original (b) (c)

Figure 6: Qualitative evaluation of geometry compression on “soldier”. (a) Original point cloud. (b) Learning-based method in [QVD20]. (c) G-PCC (Trisoup). The errors are displayed according to the color scale on the right. The learning-based method has a better point-to-point error than G-PCC (66.59dB vs. 65.87dB) for the same bitrate (0.19 bits per point).

4. **Immersive 3D HDR visualisation** - *speaker: R. Cozot*
In this part, we will expose solutions to display HDR 3D point clouds on display units of various characteristics. The objective of these solutions is twofold. The first objective is concerned with the rendering of HDR 3D contents on mainstream displays. The solutions we propose allow improving the quality of the rendering of contents (HDR 3D point clouds) on mainstream displays and HMDs (Head Mounted Displays). This improvement result from subjective evaluations we have conducted on the perception of color on HMDs. In this first part, we will detail, first, a solution to tone mapping 360^o HDR Images [GCB19] [GCLM20]. Then we will move to the challenge of tone mapping 3D dynamic scenes [GLC20]. The second objective is the stylization of 3D contents represented by point clouds. While there exist many stylization techniques applied to images (filters, blurring or vignetting effects, etc.), the stylization of 3D contents has aroused little interest. For this reason, we will present a stylization method consisting of transferring the color of a point cloud to another [GCLMB21]. This method is example-based and accounts for the geometry of the point clouds. Our results, illustrated in Figure 7, and evaluations have shown a significant improvement compared to existing color transfer methods.

3.4. Necessary background

We expect participants to know basics of computer vision and 3D imaging. It is addressed to researchers interesting in comprehending a set of issues which could be encountered when addressing the creation of immersive content from real capture.

3.5. Historical context

This tutorial was never given before. However, the tutorial organizer, C. Loscos, has given twice a tutorial on "3D Video: from Capture to Interactive Display", at Eurographics 2014 and 2015. This tutorial addresses similar problems, but exposes advanced, recent solutions. In addition, G. Valensize recently presented the tutorial "Learning-based Point Cloud Processing and Codings" at ICIIP 2021 (`https://www.2021.ieeeicip.org/Tutorials.asp`) from which content is going to be selected to compose the 3rd part of the tutorial.

4. Acknowledgements

The work presented in this tutorial is part of the ReVeRY project (`https://revery.univ-reims.fr`). The project was funded by the *Agence National pour la Recherche* (Projet-ANR-17-CE23-0020).

References

[BGPL22] Barrios T., Gerhards J., Prévost S., Loscos C.: A fast and flexible network for wide-baseline lightfield camera array disparity inference. In *Proceedings of the Eurographics conference, Poster session* (2022). 3

[BLV*12] Bonnard J., Loscos C., Valette G., Nourrit J.-M., Lucas L.: High-dynamic range video acquisition with a multiview camera. In *Optics, Photonics, and Digital Technologies for Multimedia Applications II* (2012), Schelkens P., Ebrahimi T., Cristòbal G., Truchetet F., Saarikko P., (Eds.), vol. 8436, International Society for Optics and Photonics, SPIE, pp. 81 – 91. 3

[BVL19] Bonnard J., Valette G., Loscos C.: Disparity-based HDR imaging. In *Digital Image & Signal Processing (DISP)* (Oxford, United Kingdom, 2019). URL: `https://hal.univ-reims.fr/hal-02099818`. 3

[CPZ*21] Cao C., Preda M., Zakharchenko V., Jang E. S., Zaharia T.: Compression of sparse and dense dynamic point clouds—methods and standards. *Proceedings of the IEEE 109*, 9 (2021), 1537–1558. `doi:10.1109/JPROC.2021.3085957`. 4

[CQVD21] Chetouani A., Quach M., Valenzise G., Dufaux F.: Convolutional Neural Network for 3D Point Cloud Quality Assessment with Reference. In *IEEE International Workshop on Multimedia Signal Processing (MMSP'2021)* (Tampere, Finland, Oct. 2021). URL: `https://hal.archives-ouvertes.fr/hal-03298952`. 4

[EKD*17] Eilertsen G., Kronander J., Denes G., Mantiuk R. K., Unger J.: HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics 36*, 6 (2017), 1–15. `doi:10.1145/3130800.3130816`. 3, 4

Figure 7: (Left) Tone mapping 360 HDR image according to viewport only. (Middle) Tone mapping 360 HDR image combining global image and viewport image. (Right) Tone mapping 360 HDR image according to global image only.

[GCB19] GOUDÉ I., COZOT R., BANTERLE F.: Hmd-tmo: A tone mapping operator for 360^o hdr images visualization for head mounted displays. In *Computer Graphics International Conference* (2019), Springer, pp. 216–227. 4

[GCLM20] GOUDÉ I., COZOT R., LE MEUR O.: A perceptually coherent tmo for visualization of 360^o hdr images on hmd. In *Transactions on Computational Science XXXVII*. Springer, 2020, pp. 109–128. 4

[GCLMB21] GOUDÉ I., COZOT R., LE MEUR O., BOUATOUCH K.: Example-based colour transfer for 3d point clouds. In *Computer Graphics Forum* (2021), vol. 40, Wiley Online Library, pp. 428–446. 4

[GLC20] GOUDÉ I., LACOCHE J., COZOT R.: Tone mapping high dynamic 3d scenes with global lightness coherency. *Computers & Graphics 91* (2020), 243–251. 4

[KFR*18] KHAMIS S., FANELLO S., RHEMANN C., KOWDLE A., VALENTIN J., IZADI S.: Stereonet: Guided hierarchical refinement for real-time edge-aware depth prediction. *arXiv:1807.08865 [cs]* (Jul 2018). arXiv: 1807.08865. URL: http://arxiv.org/abs/1807.08865. 3

[MZCL22] MOHAN A., ZHANG J., COZOT R., LOSCOS C.: Consistent multi- and single-view hdr-image reconstruction fromsingle exposures. In *Proceedings of the Eurographics workshop on Intelligent Cinematography and Editing* (2022). 3, 4

[NQVD21] NGUYEN D. T., QUACH M., VALENZISE G., DUHAMEL P.: Lossless Coding of Point Cloud Geometry using a Deep Generative Model. *IEEE Transactions on Circuits and Systems for Video Technology* (2021). 4

[OLMA13] OROZCO R. R., LOSCOS C., MARTIN I., ARTUSI A.: Patch-based registration for auto-stereoscopic HDR content creation. *HDRi2013 - First International Conference and SME Workshop on HDR imaging* (2013), 7. 3

[QCVD21] QUACH M., CHETOUANI A., VALENZISE G., DUFAUX F.: A deep perceptual metric for 3D point clouds. In *Image Quality and System Performance, IS&T International Symposium on Electronic Imaging (EI 2021)* (San Francisco, United States, Jan. 2021). URL: https://hal.archives-ouvertes.fr/hal-03108242. 4

[QVD] QUACH M., VALENZISE G., DUFAUX F.: Learning convolutional transforms for lossy point cloud geometry compression. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4320–4324. ISSN: 1522-4880. doi:10.1109/ICIP.2019.8803413. 4

[QVD20] QUACH M., VALENZISE G., DUFAUX F.: Improved deep point cloud geometry compression. In *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)* (2020), pp. 1–6. doi:10.1109/MMSP48831.2020.9287077. 4

[SRK20] SANTOS M. S., REN T. I., KALANTARI N. K.: Single image hdr reconstruction using a cnn with masked features and perceptual loss. *ACM Transactions on Graphics 39*, 4 (2020). doi:10.1145/3386569.3392403. 3, 4

[T*17] TIAN D., ET AL.: Geometric distortion metrics for point cloud compression. In *2017 IEEE Intl. Conf. on Image Process. (ICIP)* (Beijing, Sept. 2017), IEEE, pp. 3460–3464. URL: http://ieeexplore.ieee.org/document/8296925/, doi:10.1109/ICIP.2017.8296925. 4

EUROGRAPHICS 2022
43RD ANNUAL CONFERENCE OF
THE EUROPEAN ASSOCIATION FOR COMPUTER GRAPHICS
UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE
REIMS · FRANCE
APRIL 25-29 / 2022

FROM CAPTURE TO IMMERSIVE VIEWING OF 3D HDR POINT CLOUD

INTRODUCTION

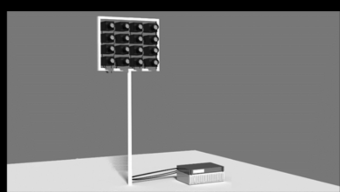
- Context
  - 3 main capture types :
    - Ultra-high-definition (UHD) : image definition and quality
    - Stereo capture for 3D : depth, multi-view
    - High-dynamic range (HDR) : higher luminance range

INTRODUCTION

- Context
  - 3 main capture types
  - Current limits:
    - No uniform representation
    - Limits creative industries to chose one format (UHD, HDR or 3D).

INTRODUCTION

- Global objective
  - Replace the traditional video stream by a rich UHD, HDR lightfield represented as a 3D point cloud in a dedicated format

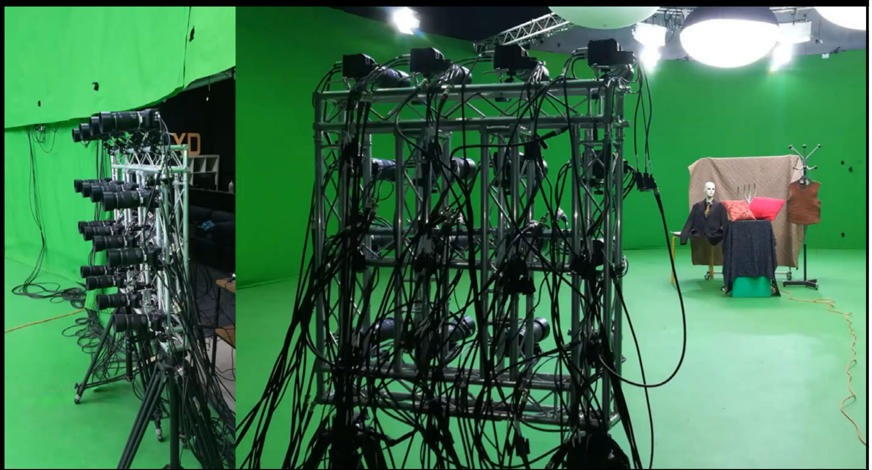


OVERVIEW OF THE PROJECT PIPELINE

- Multiview/multi exposure acquisition
- HDR/Point cloud reconstruction
- Data representation and encoding of HDR point clouds
- Visualisation on various display devices
- Quality of experience

PART I: CAMERA GRID PROTOTYPE

- 4x4 grid of cameras
- 4K video streams
- Genlock sync
- Multi-exposure patterns
- Cluster of PCs + software :
    - Remote control
    - Recording
    - Real time visualization
    - Interactive tools for directors



Part I : Multi-exposed shootings

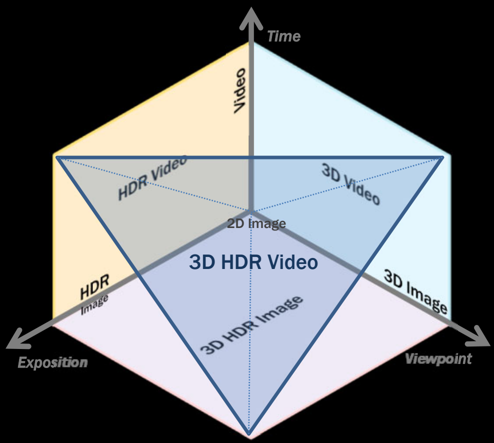
- Various challenging scenes
- For reconstruction and HDR
- Multiple objects sparsed in depth
- Overexposed & shadows
- Repetitive patterns
- Transparent and shiny objects

3D HDR reconstruction

Jennifer Bonnard[1], Gilles Valette[1], Raissel Ramirez[1,2], Ignacio Martin[2], Alessandro Artusi[2], Céline Loscos[1]

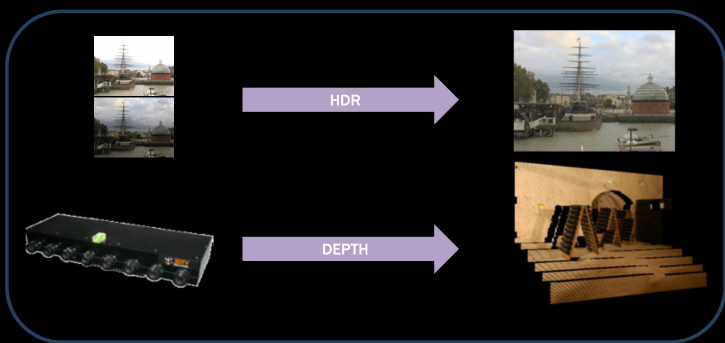
[1]Université de Reims Champagne-Ardenne, France
[2]University of Girona, Spain

METHOD CLASSIFICATION – 3 AXIS



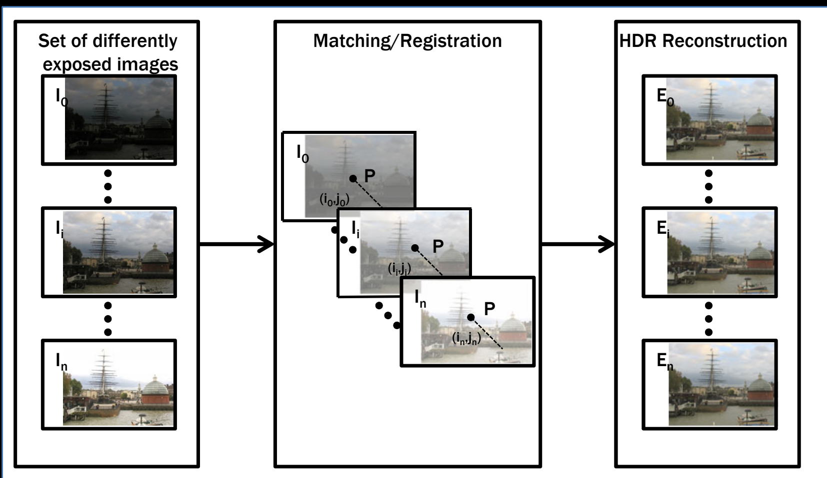
3D HDR VIDEO

- **3D point cloud colors in HDR format**



3D HDR

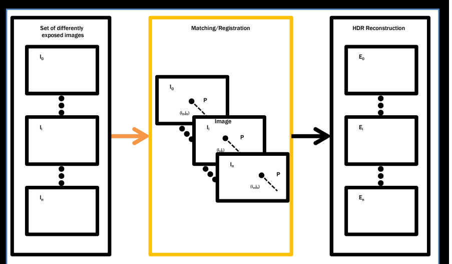
General pipeline for HDR imaging



Matching/Registration

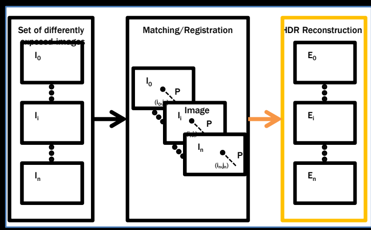
- From one view point
  - Static scenes
    - Image alignement
  - Dynamic scenes
    - Motion estimate, 2 solutions:
      - Removing the dynamic object
      - Aligning moving parts

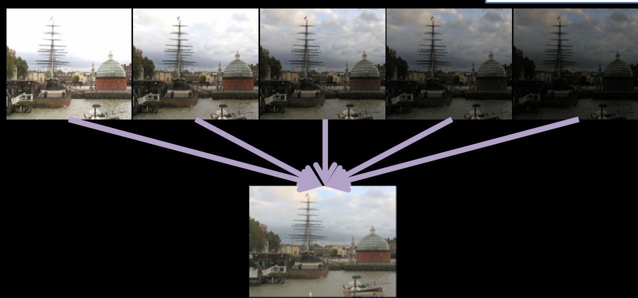
- Multiscopic images
  - Pixel registration
    - Belief propagation [LIN 09]
    - 3D estimation [LU 11]
    - Disparity [BONNARD 12]
    - Patch based [RAMIREZ 15]



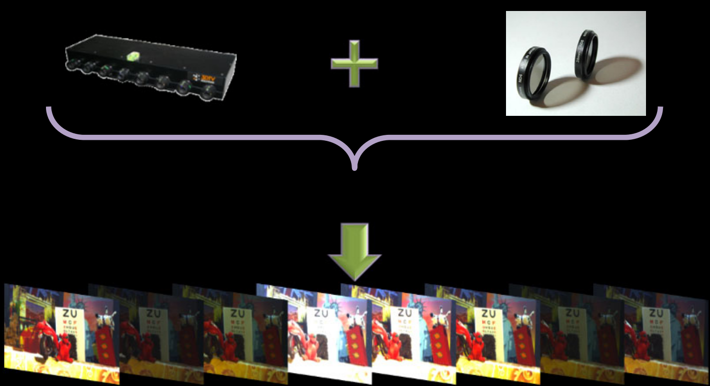
HDR Reconstruction

- HDR Reconstruction
  - Weighted average [Debevec 97]
  - Exposure fusion [Mertens07]



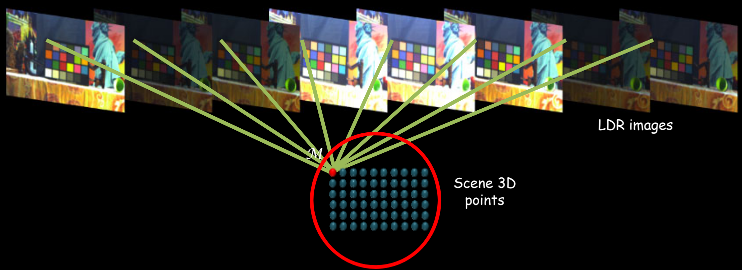


Multiview acquisition [Bonnard 12]

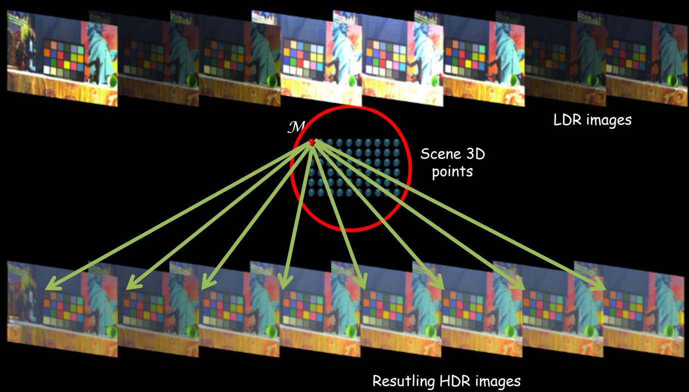


PIXEL REGISTRATION [BONNARD 12]

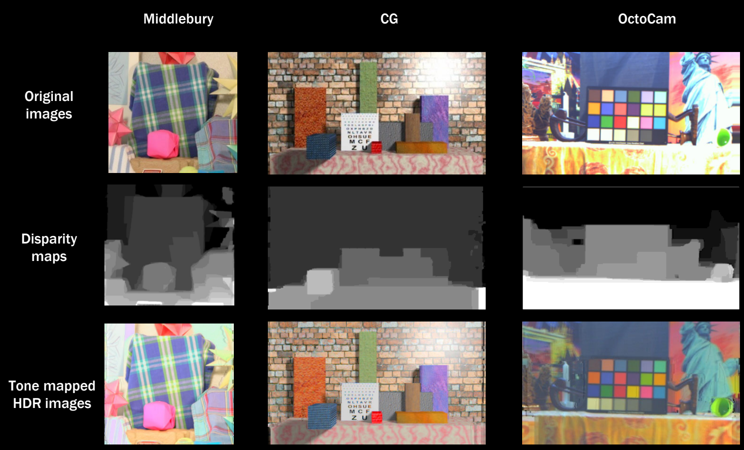
- Using disparities



HDR RECOVERY [BONNARD 12]

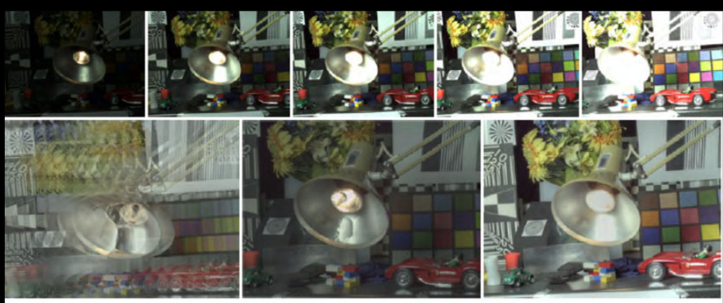


RESULTS [BONNARD 12]



PATCH-BASED 3D HDR IMAGES [RAMIREZ 15]

- Random patch match guided by the epipolar line



(a) Non aligned (b) Bätz *et al.*[BRG+14] (c) Our Result

Conclusion

- It is possible to generate an HDR point cloud from multi-view, multi-exposed images
- Difficulties raised:
  - Processing of all images
  - Alignment
- [Bonnard 12] based on the disparity estimate
  - The use of disparity for image alignment
  - Complexity of the resolution for under- or over-exposed areas
  - The quality of the disparity resolution directly impacts the HDR reconstruction
- Finally, is it a good idea to address depth reconstruction and HDR at the same time?
  - Nowadays, sensors have increased their capture capacity
  - Decision of the ReVeRY project: proposal of two separate learning-based approaches for multi-view systems
    - One for HDR reconstruction
    - Another for depth resolution

Consistent Multi- and Single-View HDR-Image Reconstruction from Single Exposures

Aditya Mohan[2], Jing Zhang[1], Rémi Cozot[1] and Céline Loscos[2]

[1] Université du Littoral Côte d'Opale
[2] Université de Reims Champagne-Ardenne

A. Mohan, J. Zhang, R. Cozot, C. Loscos: ***Consistent Multi- and Single-View HDR-Image Reconstruction from Single Exposures***. Eurographics Workshop on Intelligenent cinematography and Editing. April, 2022.

Introduction

- HDR image = Image with higher luminance range

LDR HDR

	Images LDR	Images HDR
Codage	8 bits	16-32 bits
Données	Integer	Floating point
Informations	Colors	Luminance

Introduction

LDR multi-view images

↓

HDR multi-view images

Input:

16 images of same exposure organized in a 4X4 camera grid

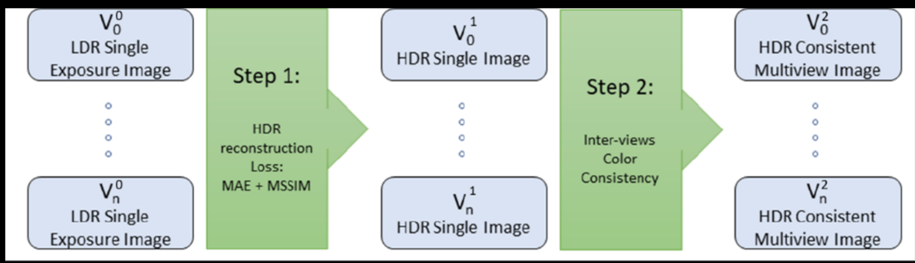
Output:

16 HDR images

Proposition of a LDR-to-HDR algorithm

Target: results as close to the ground truth as possible

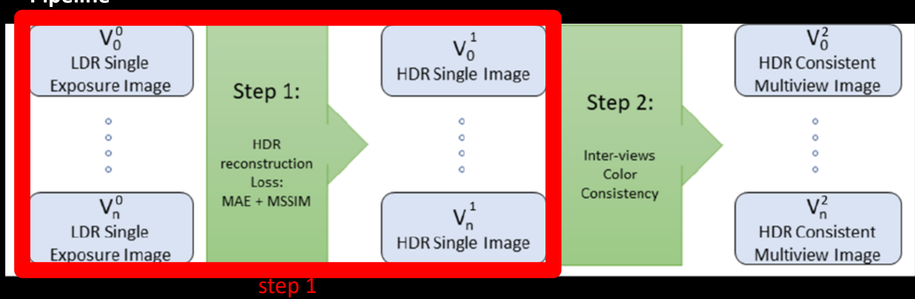
OVERALL PIPELINE



[5] EILERTSEN G., KRONANDER J., DENES G., MANTIUK R. K.,UNGER J.: ***HDR image reconstruction from a single exposure using deep CNNs***. 1–15.

METHOD – STEP 1

Pipeline



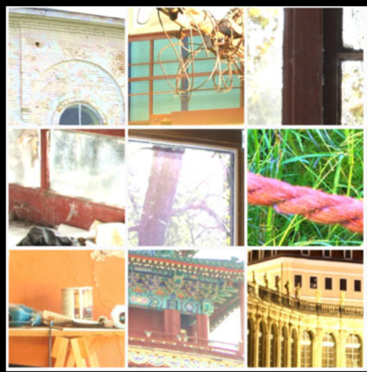
[5] EILERTSEN G., KRONANDER J., DENES G., MANTIUK R. K.,UNGER J.: ***HDR image reconstruction from a single exposure using deep CNNs***. 1–15.

Method – Step 1: Training Database

3 948 images HDR

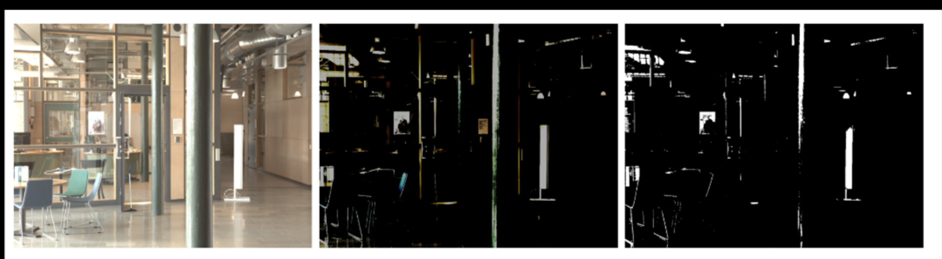
Caméra virtuelle

281 895 paires d'images HDR / LDR



Method – Step 1: Preprocessing

Masks on over-exposed areas



Original | After substracting rg chromaticity | Alpha masks

Method – Step 1: Loss functions

Training loss functions : MAE + MSSIM

Combining

Mean Absolute Error (MAE) :

$$\mathrm{MAE} = \frac{\sum_{c,i} |\widehat{y_{c,i}} - y_{c,i}|}{3 \cdot n}$$

To preserve colors and luminance

Multi-scale Structural Similarity Index :

$$MS\text{-}SSIM(x,y) = [l_M(x,y)]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j}$$

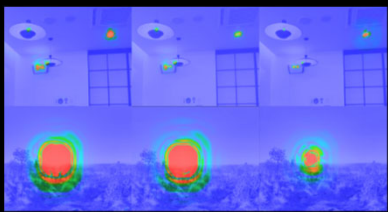
To preserve contrasts in high frequence areas

Step 1 - Results

LDR input HDR groundtruth HDRCNN MaskCNN Ours

Step 1 - Results

HDR-VDP – visual result



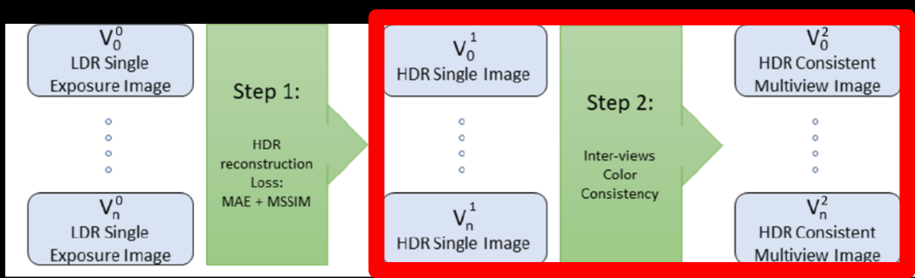
HDRCNN MaskCNN **Ours**

Metric	HDRCNN	MaskCNN	Our
Dynamic Range Error	1.051	1.026	0.921

Metric	HDRCNN	MaskCNN	Our
HDR-VDP	33.75	34.02	34.43
PSNR	57.59	57.54	55.64
Harmonic-IQA	0.314	0.315	0.310
SSIM	0.29	0.29	0.29

Average scores over 40 images of different scenes

Method – Step 2



Step 2

[5] EILERTSEN G., KRONANDER J., DENES G., MANTIUK R. K.,UNGER J.: ***HDR image reconstruction from a single exposure using deep CNNs***. 1–15.

Method - STEP 2

Multiview coherence

Images are downsampled, gathered as a group of pictures and passed again in the network as a

The output is a corrected coherent HDR value for the group of picture

Images are upsampled

Results - STEP 2

Multiview HDR Images (step 2)

Metric	Independent Views	Grid Views
SAD	3831483.42	1388318.70
NCC	0.014	0.22

Multiview consistency evaluation(step 2)

CONCLUSION

Neural network solution to extend LDR to HDR values

Improve the state of the art

- Extend luminance to closer values to ground truth HDR
- Multiview coherence consideration

REFERENCES

[1] BANTERLE F., LEDDA P., DEBATTISTA K., CHALMERS A.: ***Inverse tone mapping***. Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia - GRAPHITE 06 (2006). doi:10.1145/1174429.1174489. 1

[2] DEBEVEC P.: ***A median cut algorithm for light probe sampling***.ACM SIGGRAPH 2008 classes on - SIGGRAPH 08 (2008).

[3] DEBEVEC P. E., MALIK J.: ***Recovering high dynamic rangeradiance maps from photographs***. ACM SIGGRAPH 2008 classes on -SIGGRAPH 08 (2008). doi:10.1145/1401132.1401174. 1

[4] DOUTRE C., NASIOPOULOS P.: ***Color correction preprocessingfor multiview video coding***. IEEE Transactions on Circuits and Systems for Video Technology 19, 9 (2009), 1400–1406. doi:10.1109/tcsvt.2009.2022780. 3

[5] EILERTSEN G., KRONANDER J., DENES G., MANTIUK R. K.,UNGER J.: ***HDR image reconstruction from a single exposure using deep CNNs***. 1–15. doi:10.1145/3130800.3130816. 1, 2, 3

[6] FECKER U., BARKOWSKY M., KAUP A.: ***Histogram-based prefiltering for luminance and chrominance compensation of multiview video***. IEEE Transactions on Circuits and Systems for Video Technology18, 9 (2008),1258–1267.doi:10.1109/tcsvt.2008.926997. 2

[7] LU F., 252 JI X., DAI Q., ER G.: ***Multi-view stereo reconstruction with high dynamic range texture***. Computer Vision – ACCV 2010Lecture Notes in Computer Science (2011), 412–425. doi:10.1007/978-3-642-19309-5_32. 1

[8] MANTIUK R., KIM K. J., REMPEL A. G., HEIDRICH W.:***Hdr-vdp-2***. ACM SIGGRAPH 2011 papers on - SIGGRAPH 11 (2011).

[9] MORIWAKI K., YOSHIHASHI R., KAWAKAMI R.: ***Hybrid Loss for Learning Single-Image-based HDR Reconstruction***. 21. 2

[10] ROUSSELOT M., DUCLOUX X., MEUR O. L., COZOT R.:***Quality metric aggregation for hdr/wcg images***. 2019 IEEE Inter263 national Conference on Image Processing (ICIP) (2019). doi:10.1109/icip.2019.8803635. 3

[11] SANTOS M. S., REN T. I., KALANTARI N. K.: ***Single imageHDR reconstruction using a CNN with masked features and perceptual loss***. ACM Transactions on Graphics 39, 4 (2020). doi:10.1145/3386569.3392403. 1, 3

[12] WANG Z., SIMONCELLI E., BOVIK A.: ***Multiscale structural similarity for image quality assessment.*** In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, IEEE, pp. 1398–1402. doi:10.1109/ACSSC.2003.1292216. 2

[13] ZHAO H., GALLO O., FROSIO I., KAUTZ J.: ***Loss functions for image restoration with neural networks.*** IEEE Transactions on Com putational Imaging 3, 1 (2017), 47–57. doi:10.1109/tci.2016.2644865. 2

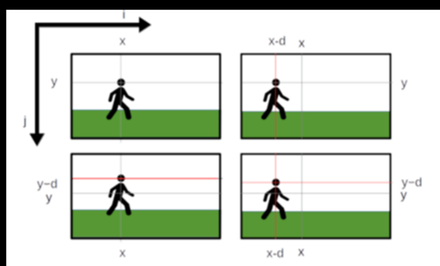
DISPARITY INFERENCE FOR WIDE-BASELINE LIGHTFIELD CAMERA ARRAY

Théo Barrios, Julien Gerhards, Stéphanie Prévost, Céline Loscos

Université de Reims Champagne-Ardenne

DEFINITION OF THE PROBLEM TO SOLVE

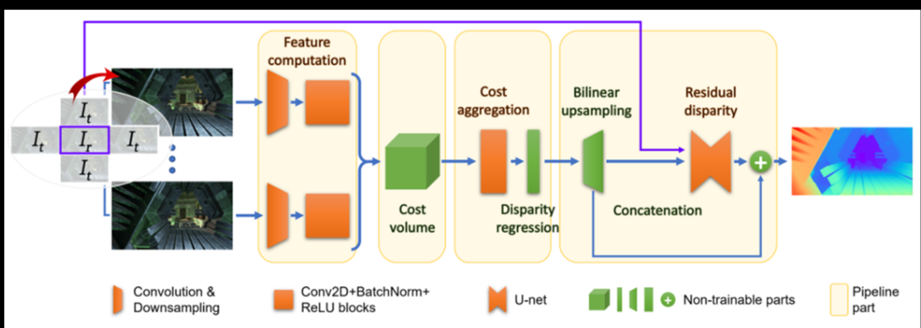
- Estimating depth from images on a 4x4 grid
- Process all images in the grid
- Propose floating-point disparities for more precision
- Process the highest resolution possible (UHD, 4k)
- Offer rapid treatment (1-n fps)
- Adapt to high camera spacing (Disparity values> 100)
- Vertical and horizontal disparities
- Additional difficulties: images located at edges and corners



Approach based on neural network

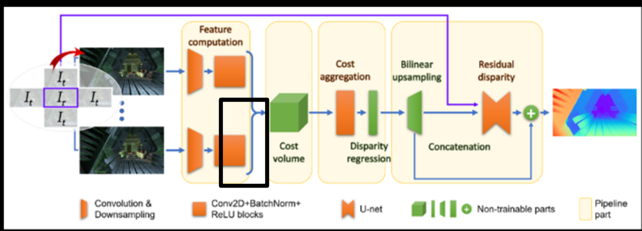
- For each image I_r of the 4X4 grid
- input : I_r + 2-4 images (I_t) in a cross around I_r
- Output : Disparity map (1)
- Solution : Deep-learning

Network description



Two-step cost volume

- First step : One cost-volume for each I_t
- Averaging cost on the horizontal / vertical part of the cross →
- two costs concatenated for cost aggregation.
- Can be used with any width and height camera array at any position with a given set of weights



Results

- Speed
  - 1,5s per view in 4k
  - 3fps in fullHD
  - 6fps in 960x540
- Quality
  - Good within the required FOV
  - Requires denoising for optimal result

Conclusion

3D reconstruction for large-baseline camera grids

- One disparity map per grid image
- Interactive time
- Can handle high resolutions (4K) and various array width and heights.
- Precise results, requires a denoising pass for application
- Different array width and heights require fine-tuning for better performance.

Coding techniques for point clouds

Giuseppe Valenzise, Centrale Supelec, CNRS

OUTLINE

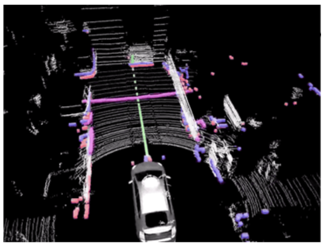
Outline

- Coding techniques for point clouds
- Quality assessment and benchmark of the different approaches
- Trends and summary

CODING TECHNIQUES FOR POINT CLOUDS

- Introduction and basic coding tools
- MPEG PCC standardization
- Learning-based techniques

MOTIVATION

- Typical point cloud video size:
  - 1 million points per frame
  - 30 frames/second
  -  color -> ~ 3.6 Gbps

- Example: Velodyne HDL-64 LiDAR sensor
  - Over 100k points per sweep
  - 3 billion points per hour

WHY CODING POINT CLOUDS IS DIFFICULT?

- Non-regular sampling
  - Geometry is expensive to code

- Spatially varying density
  - “Holes” in some regions

- Sparsity
  - Lack of spatial correlation

Some general coding approaches

- 2D projections
- Voxelization
- Octrees
- Graphs
- Using sensing structure

Some references to recent surveys on Point Cloud Compression

- C. Cao, M. Preda, V. Zakharchenko, E. S. Jang, and T. Zaharia, "Compression of Sparse and Dense Dynamic Point Clouds–Methods and Standards," *Proceedings of the IEEE*, pp. 1–22, 2021
- F. Pereira, A. Dricot, J. Ascenso, and C. Brites, "Point cloud coding: A privileged view driven by a classification taxonomy," *Signal Processing: Image Communication*, vol. 85, p. 115862, Jul. 2020
- Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T. and Tabatabai, A., 2020. An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC). *APSIPA Transactions on Signal and Information Processing*, 9.

3D to 2D projection

- Reduce the problem to multiple 2D image coding instances
- Effective when the point cloud is dense enough to get smooth projections
- Used in MPEG V

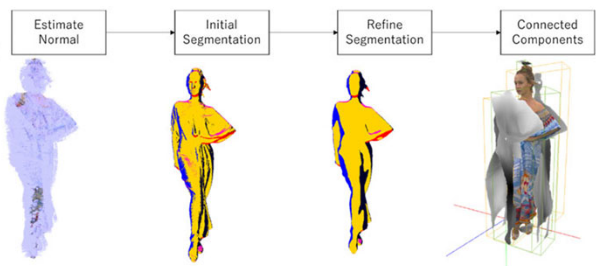


Image credit: Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T. and Tabatabai, A., 2020. An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC). APSIPA Transactions on Signal and Information Processing, 9.

VOXELIZATION

- Quantize the 3D coordinates of points to a given bit depth
  - Define the point clouds on a regular 3D lattice
  - Geometry represented as binary **occupancy** maps
  - Attributes resampled over the voxel grid
- Introduce distortion wrt original point cloud
- *Highly inefficient to deal with sparsity!*
  - Most of the 3D space is empty
  - It might require a very high bit depth to represent precisely sparse point clo...

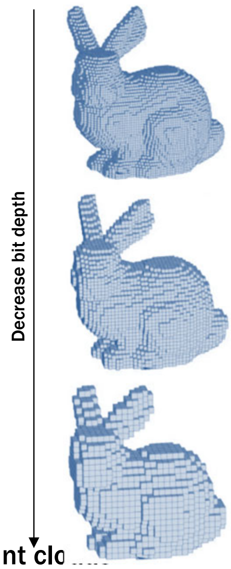


Image credit: Yong Zhou, Han Lu, Gongxian Wang, Junfeng Wang, Weidong Li, Voxelization modelling based finite element simulation and process parameter optimization for Fused Filament Fabrication, Materials & Design, Volume 187, 2020.

DEALING WITH SPARSITY: TREE-BASED PARTITIONING

- Divide the space hierarchically
  - E.g., KD-tree or typically octree
  - Remove empty space
- **Octree**
  - Recursive s
  - Each octree

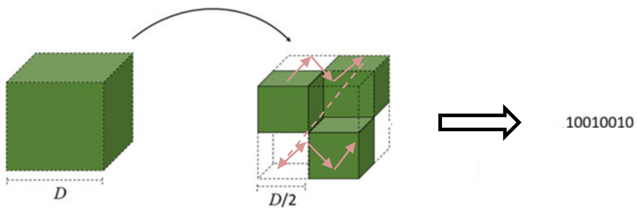
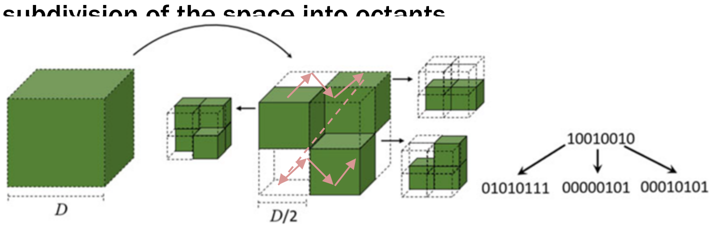


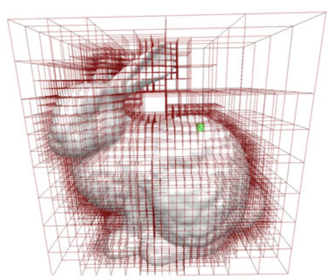
Image credit: Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T. and Tabatabai, A., 2020. An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC). APSIPA Transactions on Signal and Information Processing, 9.

Dealing with sparsity: tree-based partitioning

- Divide the space hierarchically
  - E.g., KD-tree or typically octree
  - Remove empty space
- Octree
  - Recursive subdivision of the space into octants
  - Each octr



Octree-based coding

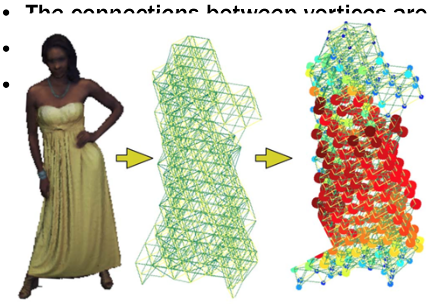


- Widely used since the early PC coding methods
  - E.g., the Point Cloud Library (PCL)[1]
  - Benchmark codec in the MPEG G-PCC CfP (2017)[2]
- Essential coding tool in MPEG G-PCC
- Basic functionalities:
  - Arithmetic coding of voxel occupancies using previous nodes as context
  - Detail encoding (e.g., residuals) or surface approximations
  - Attributes averaged over the points in the leaf nodes
  - Temporal prediction possible by matching nodes

[1]J. Kammerl, N. Blodow, R. B. Rusu, S. Gedikli, M. Beetz and E. Steinbach, "Real-time compression of point cloud streams," IEEE International Conference on Robotics and Automation, 2012, pp. 778-785
[2]R. Mekuria, K. Blom and P. Cesar, "Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 4, pp. 828-842, April 2017
Image credit: Castro, R., Lewiner, T., Lopes, H., Tavares, G. and Bordignon, A., Sep. 2008. Statistical optimization of octree searches. In Computer Graphics Forum (Vol. 27, No. 6, pp. 1557-1566). Oxford, UK: Blackwell Publishing Ltd.

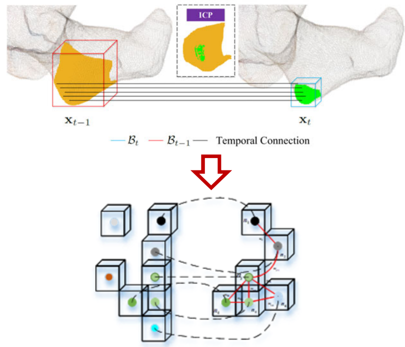
GRAPH METHODS

- Graph construction
  - The connections between vertices are obtained by k-NN
  - … ce between points
  - … over a graph



Graph construction

Motion Estimation

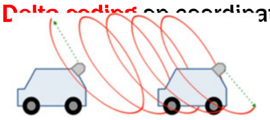


Optimal Inter-Prediction and Transform
- **Generalized Graph Fourier Transform (GGFT)**

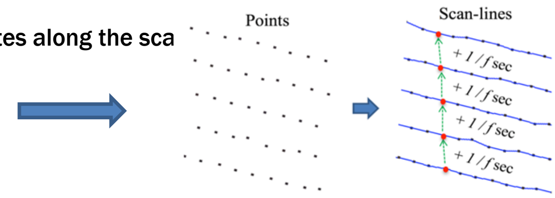
D. Thanou, P. A. Chou and P. Frossard, "Graph-Based Compression of Dynamic 3D Point Cloud Sequences," in IEEE Transactions on Image Processing, vol. 25, no. 4, pp. 1765-1778, April 2016
Y. Xu et al., "Predictive Generalized Graph Fourier Transform for Attribute Compression of Dynamic Point Clouds," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 5, pp. 1968-1982, May 2021

LIDAR DATA: LEVERAGE THE ACQUISITION MODEL

- Structured:
  - If metadata available (GPS data, timestamps, sensor information, etc.)
  - The point cloud becomes an *ordered* set
- Sequential scan lines
  - Delta coding on coordinates along the sca…



Trajectory of the laser beam

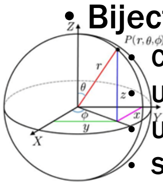


Kohira, K. and Masuda, H.: Point-cloud Compression For Vehicle-based Mobile Mapping Systems Using Portable Network Graphics, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2/W4, 99–106, 2017.
Xianxiong Liu, Yue Wang, Qingwu Hu and Dengbo Yu, "A scan-line-based data compression approach for point clouds: Lossless and effective," 2016 4th International Workshop on Earth Observation and Remote Sensing Applications, 2016

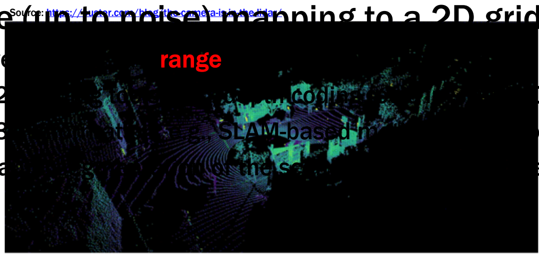
LiDAR data: leverage the acquisition model

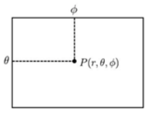
- Bijective (up to noise) mapping to a 2D grid:

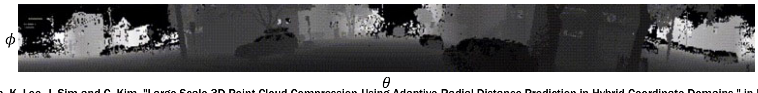
Source: https://ouster.com/blog/the-camera-is-in-the-lidar/

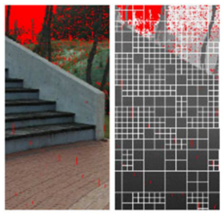


  - Conve … + radial distance)
  - Use 2 … C)
  - Use 3 … n
  - Sema … etc.)









Quadtree decomposition of a range image[1]

[1]J. Ahn, K. Lee, J. Sim and C. Kim, "Large-Scale 3D Point Cloud Compression Using Adaptive Radial Distance Prediction in Hybrid Coordinate Domains," in IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 3, pp. 422-434, April 2015

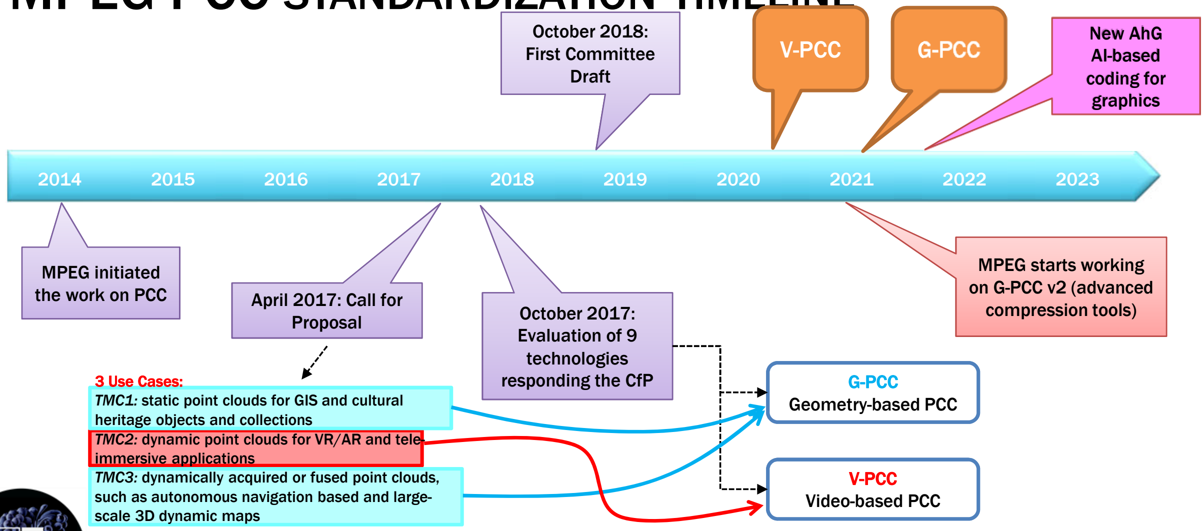
C. Tu, E. Takeuchi, C. Miyajima and K. Takeda, "Continuous point cloud data compression using SLAM based prediction," 2017 IEEE Intelligent Vehicles Symposium (IV), 2017, pp. 1744-1751

X. Sun, H. Ma, Y. Sun and M. Liu, "A Novel Point Cloud Compression Algorithm Based on Clustering," in IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 2132-2139, April 2019

Coding techniques for point clouds

- Introduction and basic coding tools
- MPEG PCC standardization
- Learning-based techniques

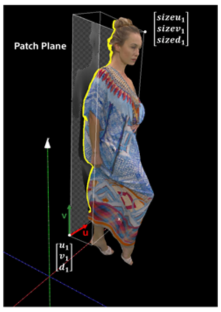
MPEG PCC STANDARDIZATION TIMELINE



MPEG V-PCC

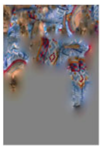
- Video-based Point Cloud Compression

Projection-based coding principle



Occupancy map

Geometry

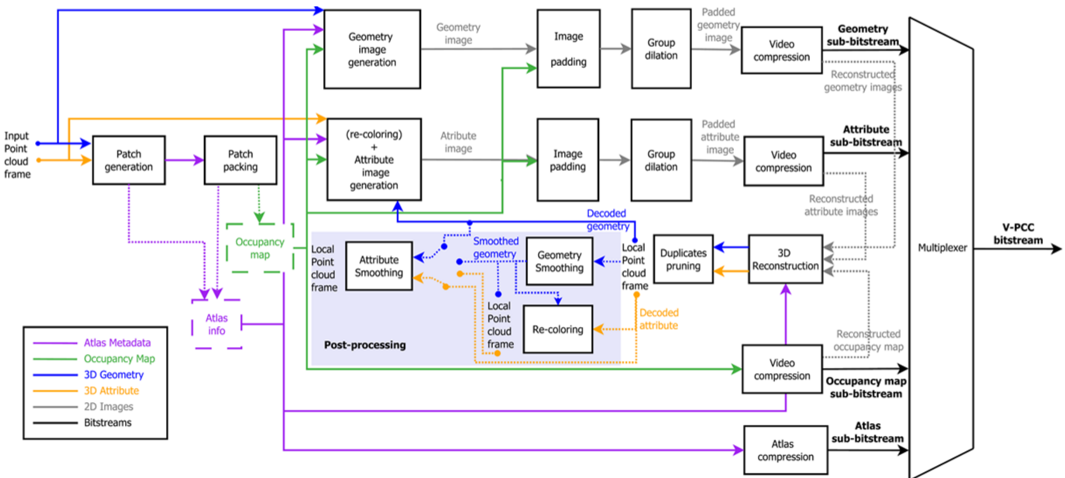


Attribute images

Source code:
https://github.com/MPEGGroup/mpeg-pcc-tmc2

Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T. and Tabatabai, A., 2020. An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC). APSIPA Transactions on Signal and Information Processing, 9.

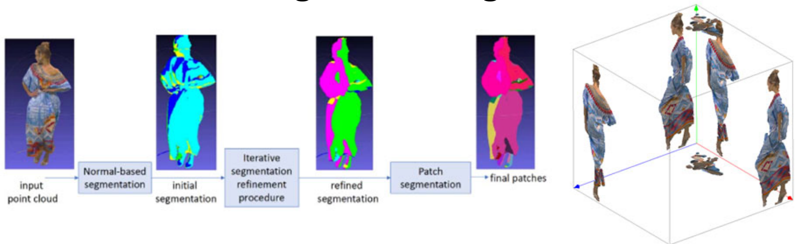
V-PCC codec architecture (TMC2)



MPEG 3DG, V-PCC codec description, ISO/IEC JTC1/SC29/WG11 N18892, 2019.

V-PCC CODEC ARCHITECTURE (TMC2)

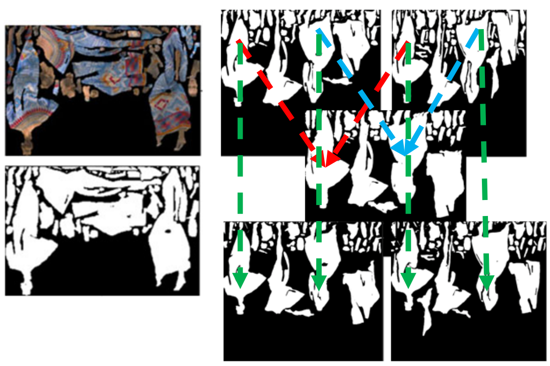
- Estimate projection direction for each patch
  - Based on normals
- The 3D patch is projected orthogonally to one of the six faces of the axis-aligned bounding box



MPEG 3DG, V-PCC codec description, ISO/IEC JTC1/SC29/WG11 N18892, 2019.

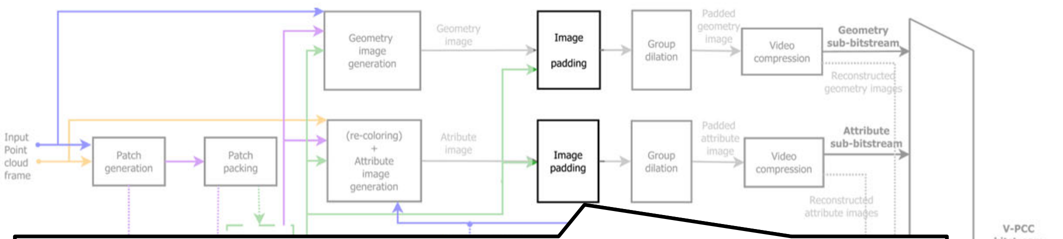
V-PCC CODEC ARCHITECTURE (TMC2)

- Placement of projected 2D patches in a 2D image
- Iterative and heuristic process
- Enforce temporal consistency

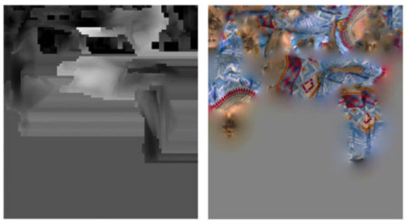


MPEG 3DG, V-PCC codec description, ISO/IEC JTC1/SC29/WG11 N18892, 2019.

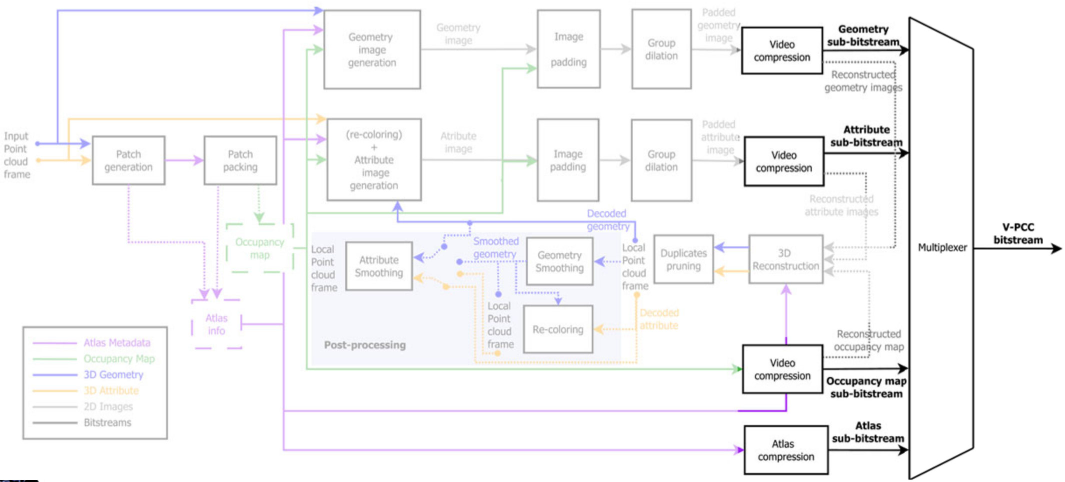
V-PCC CODEC ARCHITECTURE (TMC2)



- Fill the empty space between patches using a padding function
- Generate piecewise smooth image easier to code



V-PCC CODEC ARCHITECTURE (TMC2)

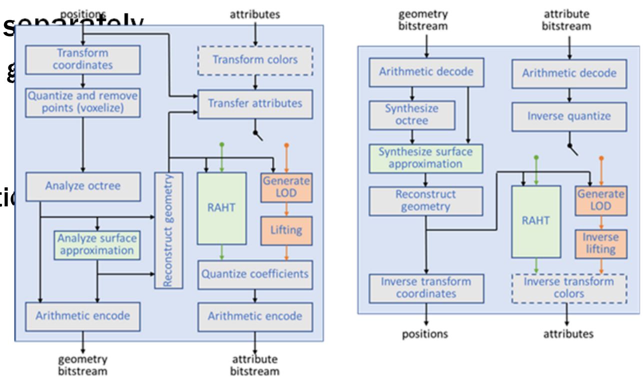


MPEG G-PCC

- Geometry-based Point Cloud Compression

OCTREE-BASED APPROACH

- Geometry and attributes are encoded separately
  - Attribute coding depends on decoded g
- Workflow:
  - Coordinate transformation + voxelizati
  - Geometry coding (octree/pred)
  - Transform (attributes)
  - Arithmetic coding



MPEG 3DG, V-PCC codec description, ISO/IEC JTC 1/SC 29/WG 7 N 0099, April 2021

GEOMETRY CODING IN G-PCC

- Two basic approaches:

1. … ral mo

- *Limitations of a vanilla octree coding*:
  - Isolated points are expensive to code
  - Number of points exponentially decreasing at low bitrates
  - Does not use the local geometric structures
  - Does not use structure/side information when available

GEOMETRY CODING IN G-PCC

- Two basic approaches:

1. Octree coding augmented with several modes

   (Inferred) Direct coding mode (for isolated points)

Geometry coding in G-PCC

- Two basic approaches:

1. Octree coding augmented with several modes
   (Inferred) Direct coding mode
   Triangle soup (trisoup)

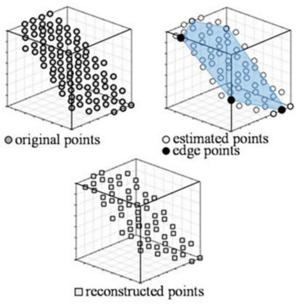
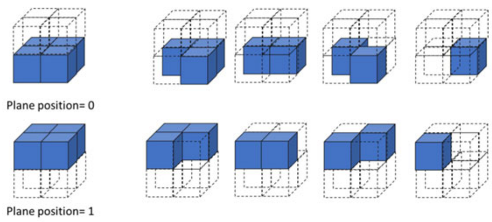


Image credit: A. Dricot and J. Ascenso, "Hybrid Octree-Plane Point Cloud Geometry Coding," 2019 27th European Signal Processing Conference (EUSIPCO), 2019, pp. 1-5.

Geometry coding in G-PCC

- Two basic approaches:

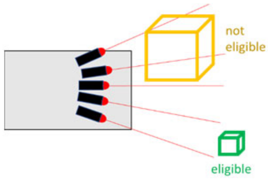
1. Octree coding augmented with several modes
   (Inferred) Direct coding mode
   Triangle soup (trisoup)
   Planar modes



GEOMETRY CODING IN G-PCC

- Two basic approaches:

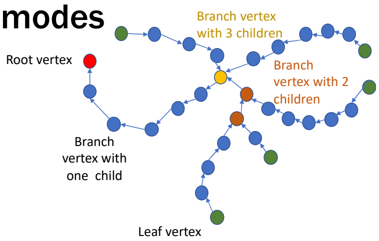
1. Octree coding augmented with several modes
   - (Inferred) Direct coding mode
   - Trisoup
   - Planar modes
   - **Angular modes**



GEOMETRY CODING IN G-PCC

- Two basic approaches:

1. Octree coding augmented with several modes
   - (Inferred) Direct coding mode
   - Trisoup
   - Planar modes
   - Angular modes

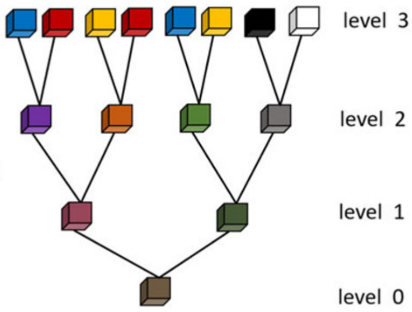


Predictive geometry coding (low-latency)

ATTRIBUTE CODING IN G-PCC

- Two tools:

1. Region-Adaptive Hierarchical Transform (RA
   Haar-inspired transform on octree structure



Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T. and Tabatabai, A., 2020. An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC). APSIPA Transactions on Signal and Information Processing, 9.

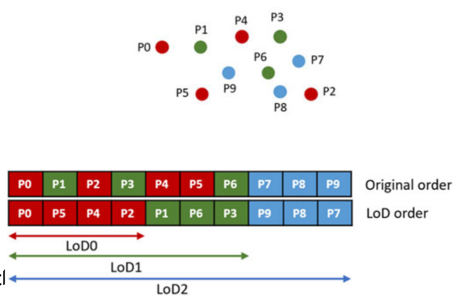
ANR IRISA

ATTRIBUTE CODING IN G-PCC

- Two tools:

1. Region-Adaptive Hierarchical Transform (RAHT)
   Haar-inspired transform on octree structure

2. Predicting/Lifting Transform
   Distance-based prediction
   Employs a Level of Detail (LoD) representation that distributes tl



Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T. and Tabatabai, A., 2020. An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC). APSIPA Transactions on Signal and Information Processing, 9.

ANR IRISA

Take-away on MPEG standardization

- Two standards

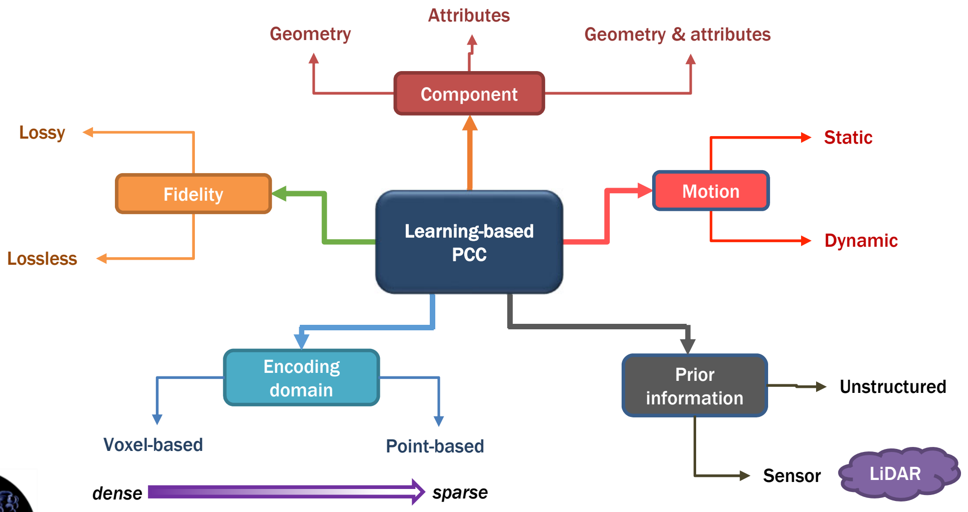
- V-PCC:
  - 2D projection-based
  - Dense PC
  - Dynamic content
  - AR/VR applications

- G-PCC:
  - Mostly octree-based + many optimizations
  - Static content
  - Low-to-high density
  - Wide range of applications: AR/VR, cultural heritage, LiDAR (fused and scans), etc.

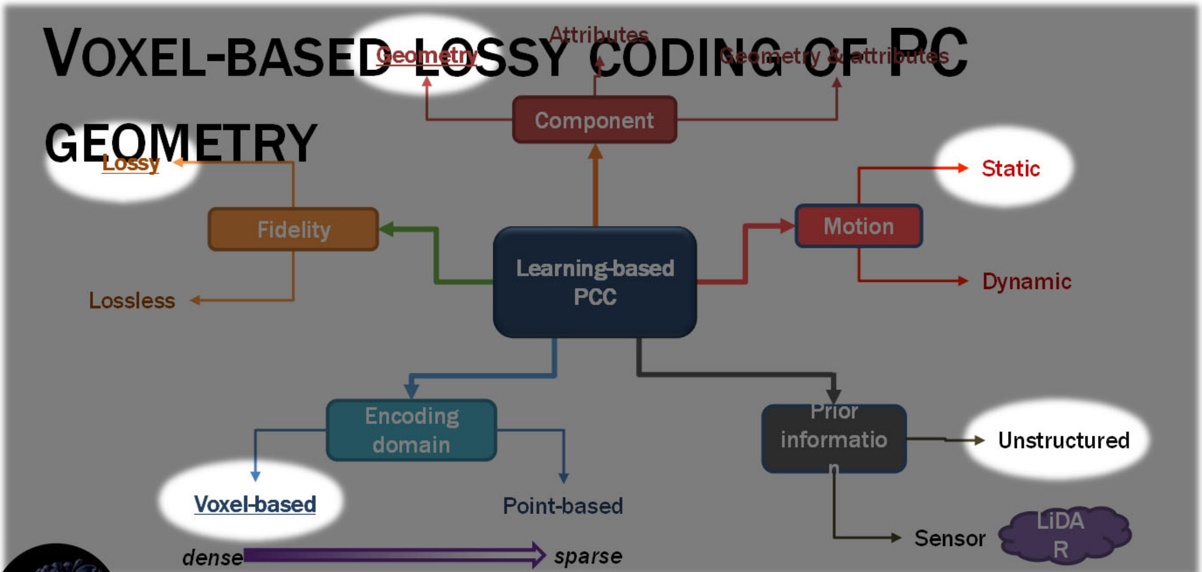
Coding techniques for point clouds

- Introduction and basic coding tools
- MPEG PCC standardization
- **Learning-based techniques**

Taxonomy



Voxel-based lossy coding of PC geometry

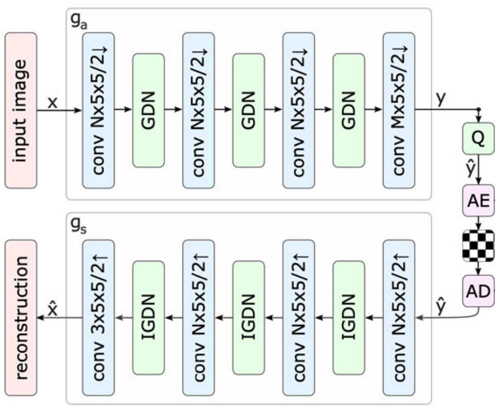


Voxel-based lossy coding of PC geometry

- Adapted to dense point clouds
- Similar to learning-based 2D image compression
  - Auto-encoder based approach
  - Entropy bottleneck and quantization

Background: learning-based image compression

- Var

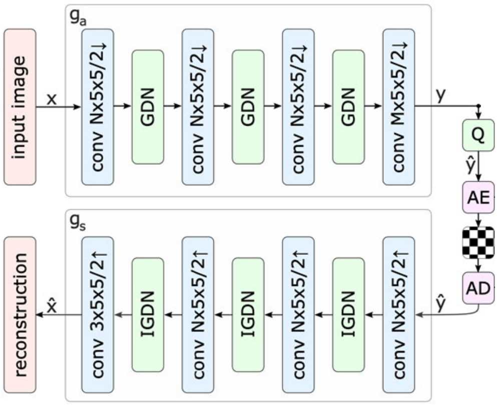


- Optimized end-to-end
- Quantization
  - Non differentiable
  - Backward pass (in training):
    $$\hat{y}_i = y_i + \mathcal{U}\left[-\frac{1}{2}, \frac{1}{2}\right]$$
  - Inference:
    $$\hat{y}_i = \text{round}(y_i)$$
- Entropy coding
  - Differential entropy for training
- What is learned:
  - Analysis transform
  - Synthesis transform
  - Probability distribution of latent variables

Ballé, J., Laparra, V., Simoncelli, E., "End-to-end optimized image compression". In *International Conference on Learning Representations*. 2016

BACKGROUND: LEARNING-BASED IMAGE COMPRESSION

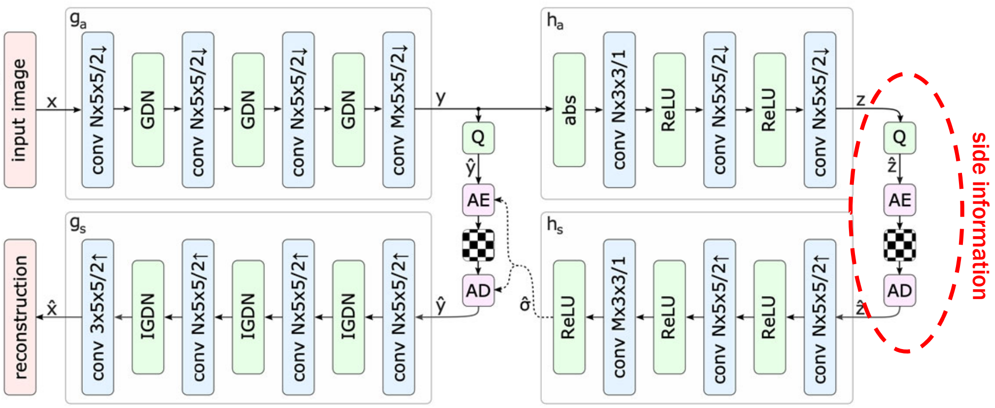
- The
  $\prod p_y$
- In pı



Ballé, J., Laparra, V., Simoncelli, E., "End-to-end optimized image compression". In *International Conference on Learning Representations*. 2016

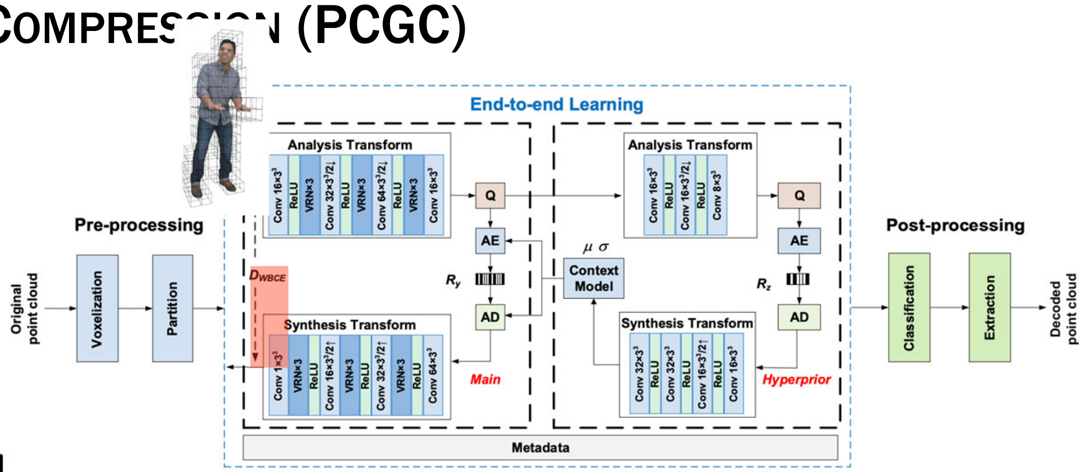
BACKGROUND: LEARNING-BASED IMAGE COMPRESSION

- Varia
- Hype



Ballé, J., Minnen, D., Singh, S., Hwang, S.J. and Johnston, N., "Variational image compression with a scale hyperprior". In *International Conference on Learning Representations*. Feb. 2018

Learning-based Point Cloud Geometry Compression (PCGC)



J. Wang, H. Zhu, H. Liu and Z. Ma, "Lossy Point Cloud Geometry Compression via End-to-End Learning," in IEEE Transactions on Circuits and Systems for Video Technology, 2021

Classification

- Cast the reconstruction problem as a classification one
- Loss function
  - e the probability that a voxel is occupied

$$p_t = \begin{cases} p & \text{if the voxel was indeed occupied} \\ 1-p & \text{otherwise} \end{cases}$$

Probability of assigning the correct class

- Binary Cross Entropy (BCE):

$$BCE(p_t) = -\log p_t$$

Classification: Weighted Binary Cross-Entropy

$$p_t = \begin{cases} p & \text{if the voxel was indeed occupied} \\ 1-p & \text{otherwise} \end{cases}$$

Probability of assigning the correct class

- Binary Cross Entropy (BCE):

$$BCE(p_t) = -\log p_t$$

- Take into account class imbalance in the minimization (most voxels are e
- **Weighted Binary Cross-Entropy (WBCE)**

$$WBCE(p_t) = -\alpha_t \log p_t$$

with $\alpha_t = \alpha$ if the voxel ground-truth occupation is 1, and $\alpha_t = 1 - \alpha$ otherwise.

Typically $\alpha \propto \dfrac{\text{total num. voxels}}{\text{num. occupied voxels}}$

✓ Balances the importance of occupied/not occupied voxels

✗ Does not differentiate between easy/hard examples
Most of the empty voxels are easily classified and do not bring much information to learning

J. … Zhu, H. … nd Z. Ma, "Lossy Point Cloud Geometry Compression via End-to-End Learning," in IEEE Transactions on Circuits an… Syst… r Video Technology, 2021

Classification: Focal Loss

$$p_t = \begin{cases} p & \text{if the voxel was indeed occupied} \\ 1-p & \text{otherwise} \end{cases}$$

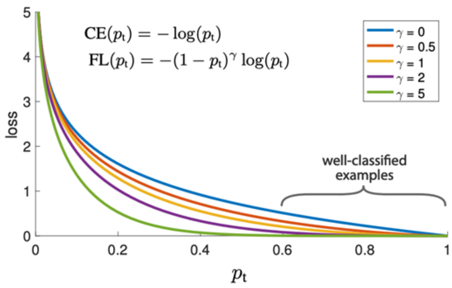
Probability of assigning the correct class

- Binary Cross Entropy (BCE):

$$BCE(p_t) = -\log p_t$$

- Take into account class imbalance in the minimization
- **Focal Loss (FL)**

$$FL(p_t) = -\alpha_t\,(1-p_t)^{\gamma} \log p_t$$



Voxels easy to classify have less weight in the loss

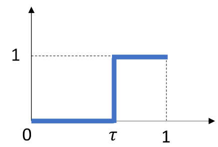
with $\alpha_t = \alpha$ if the voxel ground-truth occupation is 1, and $\alpha_t = 1 - \alpha$ oth

γ is a focusing parameter

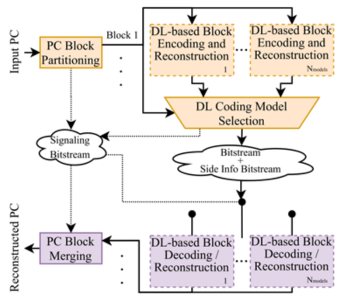
Maurice Quach, Giuseppe Valenzise, Frédéric Dufaux. Learning Convolutional Transforms for Lossy Point Cloud Geometry Compression. *IEEE International Conference on Image Processing (ICIP'2019)*, Sep 2019, Taipei, Taiwan.
T. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2999-3007, doi: 10.1109/ICCV.2017.324.

Dealing with variable spatial density

DEALING WITH VARIABLE SPATIAL DENSITY

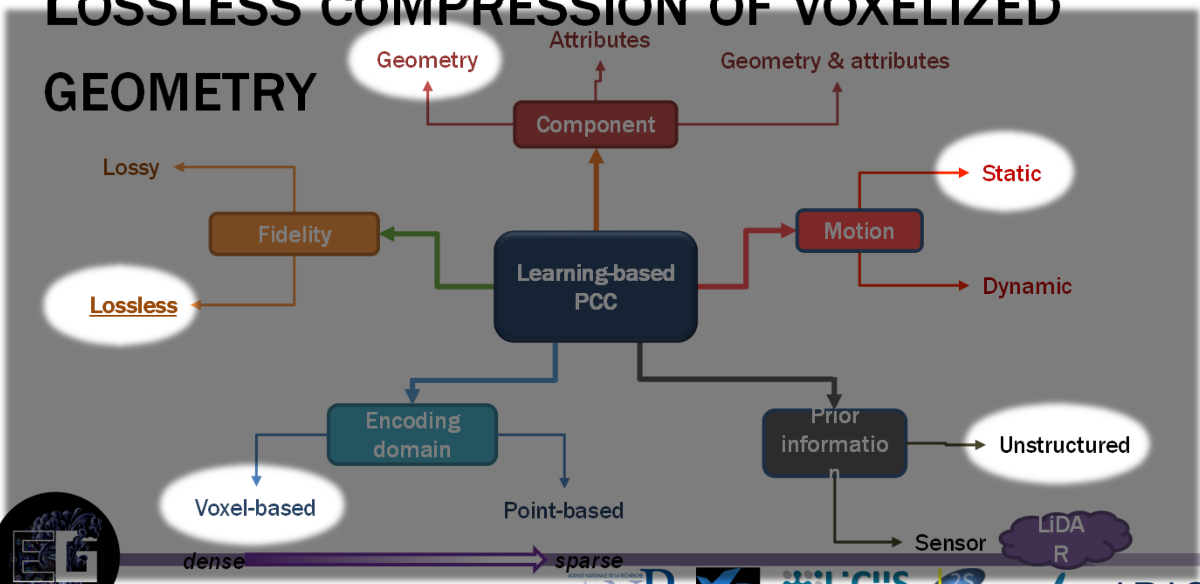


- Adaptive thresholding
  - Use a threshold to binarize the estimated occupancy probability
  - Fixed threshold (e.g., 0.5)[1] is suboptimal
  - Optimized over the whole PC and transmitted[2]
  - Optimized block by block and transmitted[3]
- Adaptive model (ADL-PCC[4])
  - Train a different network for each α in the focal loss
  - Select the best model using RDO
  - Signal the model index in the bitstream



[1] Maurice Quach, Giuseppe Valenzise, Frédéric Dufaux. Learning Convolutional Transforms for Lossy Point Cloud Geometry Compression. *IEEE International Conference on Image Processing (ICIP 2019)*, Sep 2019, Taipei, Taiwan.
[2] J. Wang, H. Zhu, H. Liu and Z. Ma, "Lossy Point Cloud Geometry Compression via End-to-End Learning," in IEEE Transactions on Circuits and Systems for Video Technology, 2021
[3] Maurice Quach, Giuseppe Valenzise, Frédéric Dufaux. Improved Deep Point Cloud Geometry Compression. *IEEE International Workshop on Multimedia Signal Processing (MMSP'2020)*, Sep 2020, Tampere, Finland.
[4] A. F. R. Guarda, N. M. M. Rodrigues and F. Pereira, "Adaptive Deep Learning-Based Point Cloud Geometry Coding," in IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 2, pp. 415-430, Feb. 2021

LOSSLESS COMPRESSION OF VOXELIZED GEOMETRY

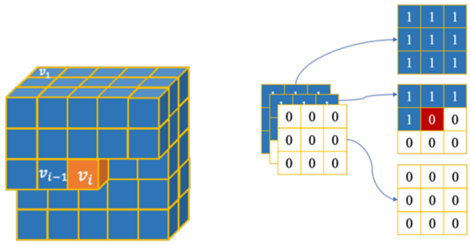


VoxelDNN

- Goal: estimate the occupancy probabilities to use in a context-based arithmetic codec
- Autoregressive model:
  - Inspired by PixelCNN[1]
  - Factorize the joint probability of voxel occupancy

$$p(v) = \prod_{i=1}^{d^3} \underbrace{p(v_i | v_{i-1}, v_{i-2}, \ldots, v_1)}_{\text{Modeled with a DNN}}$$

  - Masked convolution to enforce causality for correct decoding



(a) 3D voxel context (b) 3D type A mask

Dat Thanh Nguyen, Maurice Quach, Giuseppe Valenzise, Pierre Duhamel. Lossless Coding of Point Cloud Geometry using a Deep Generative Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021

VoxelDNN

- Goal: estimate the occupancy probabilities to use in a context-based arithmetic codec
- Autoregressive model:
  - Inspired by PixelCNN[1]
  - Factorize the joint probability of voxel occupancy

$$p(v) = \prod_{i=1}^{d^3} \underbrace{p(v_i | v_{i-1}, v_{i-2}, \ldots, v_1)}_{\text{Modeled with a DNN}}$$

  - Masked convolution to enforce causality for correct decoding

- Trained with *cross-entropy loss*

$$H(p, \hat{p}) = \mathbb{E}_{v \sim p(v)} \left[\sum_{i=1}^{d^3} -\log \hat{p}(v_i) \right]$$

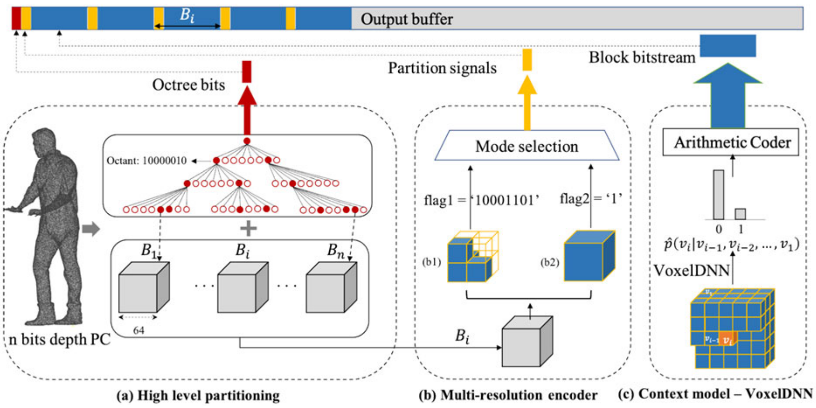
$$= H(p) + \boxed{D_{KL}(p \| \hat{p})}$$

  - Minimizes the distance between the estimated occupancy probability and the ground truth
  - Different from the lossy case!

Dat Thanh Nguyen, Maurice Quach, Giuseppe Valenzise, Pierre Duhamel. Lossless Coding of Point Cloud Geometry using a Deep Generative Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021

VOXELDNN: GENERAL ARCHITECTURE

- Hybrid
- VoxelD



prediction

Dat Thanh Nguyen, Maurice Quach, Giuseppe Valenzise, Pierre Duhamel. Lossless Coding of Point Cloud Geometry using a Deep Generative Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021

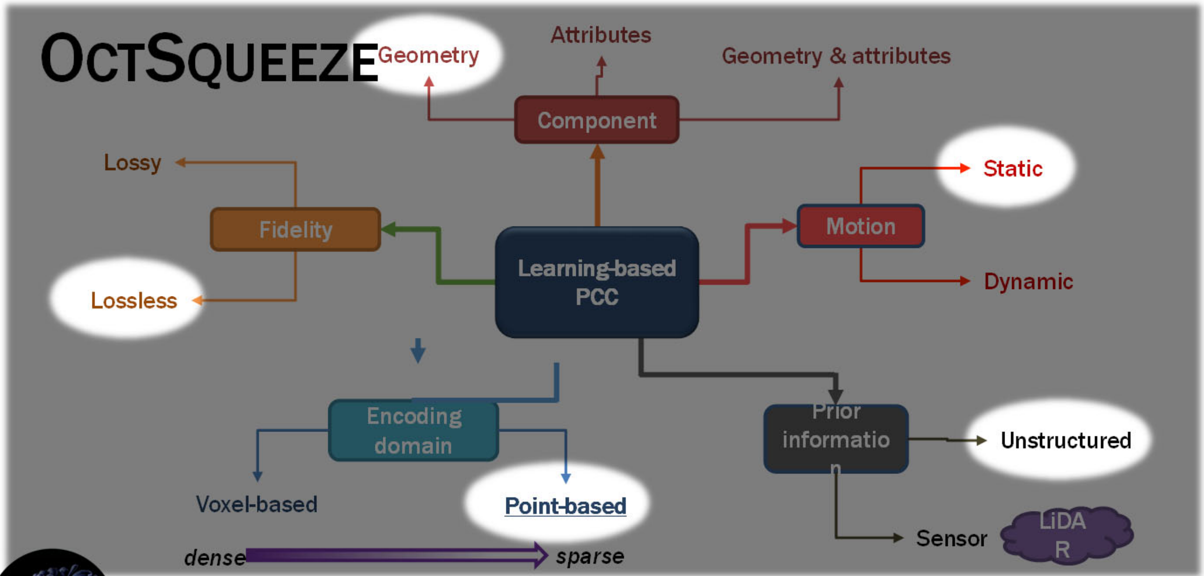
VOXELDNN

- **Pro's:**
  - Significant bitrate reductions
  - Flexible to be extended with larger contexts
- **Con's:**
  - Poorer performance on sparser PCs
  - Sequential voxel-by-voxel decoding
    - High computational complexity
    - Approximations (MSVoxelDNN)

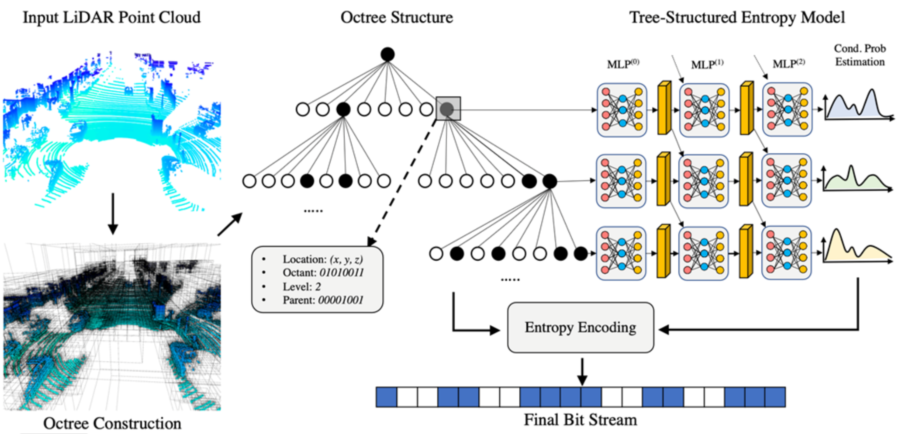
		G-PCC	Baseline + DA + CE	
Dataset	Point Cloud	bpov	bpov	Gain over G-PCC
MVUB	Phil	1.1599	0.8252	-28.86%
	Ricardo	1.0673	0.7572	-29.05%
	Average	**1.1136**	**0.7912**	**-28.95%**
8i	Redandblack	1.0893	0.7003	-35.71%
	Loot	0.9524	0.6084	-36.12%
	Thaidancer	0.9990	0.6627	-33.66%
	Boxer	0.9492	0.5906	-37.78%
	Average	**0.9975**	**0.6405**	**-35.79%**
CAT1	Frog	1.8990	1.7071	-10.11%
	Arco Valentino	4.8531	4,9900	+2.82%
	Shiva	3.6716	3.5135	-4.31%
	Average	**3.4746**	**3.4035**	**-3.86%**
USP	BumbaMeuBoi	5.4068	5.066	-6.29%
	RomanOiLight	1.8604	1.6231	-12.76%
	Average	**3.6336**	**3.4855**	**-9.52%**

Dat Thanh Nguyen, Maurice Quach, Giuseppe Valenzise, Pierre Duhamel. Lossless Coding of Point Cloud Geometry using a Deep Generative Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021
Dat Thanh Nguyen, Maurice Quach, Giuseppe Valenzise, Pierre Duhamel. Multiscale deep context modeling for lossless point cloud geometry compression. *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Jul 2021, Shenzhen (virtual), China.

OctSqueeze



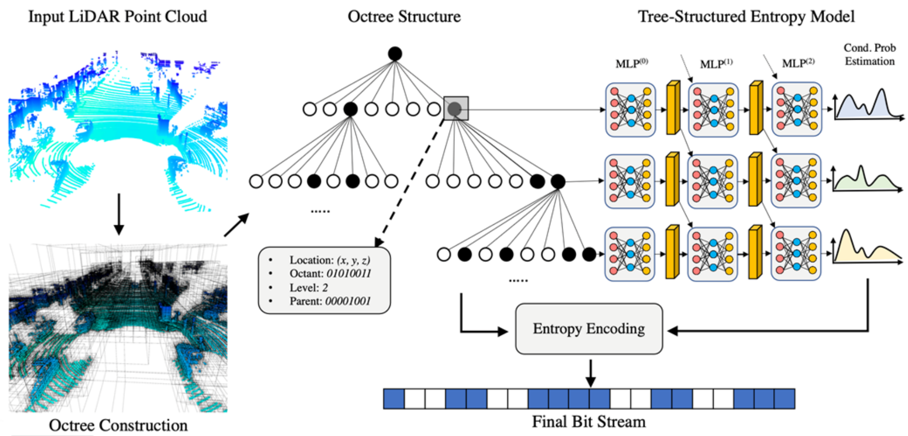
OctSqueeze



- x_i: occupancy of a node (1 byte)
- $\mathbf{x} = [x_1, x_2, \dots, x_n]$: octree structure
- $p(\mathbf{x})$: real occupancy distribution
- $q(\mathbf{x})$: estimated distribution
- **Goal: minimize CE** $H(p, q) = \mathbb{E}_{\mathbf{x}\sim p}[-\log_2 q(\mathbf{x})]$

L. Huang, S. Wang, K. Wong, J. Liu and R. Urtasun, "OctSqueeze: Octree-Structured Entropy Model for LiDAR Compression," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 1310-1320

OctSqueeze



- Factorize $q(\mathbf{x})$ as

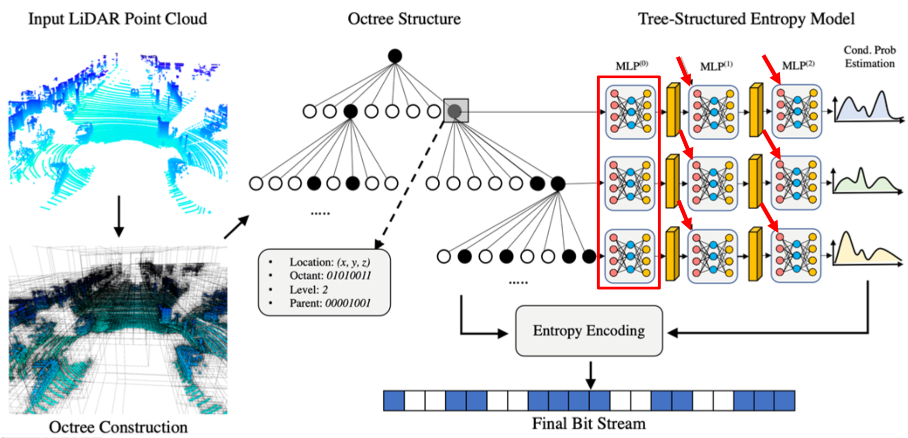
$$q(\mathbf{x}) = \prod_i q_i(x_i \mid \mathbf{x}_{\mathrm{an}(i)}, \mathbf{c}_i; \mathbf{w})$$

where:

- $\mathbf{x}_{an}$ is a set of ancestor nodes
- $\mathbf{c}_i$ is additional context (spatial location, level in the octree, etc.)

L. Huang, S. Wang, K. Wong, J. Liu and R. Urtasun, "OctSqueeze: Octree-Structured Entropy Model for LiDAR Compression," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 1310-1320

OctSqueeze



- Estimation of conditional probabilities $q(x_i)$

1. Embedding of the current node using MLP

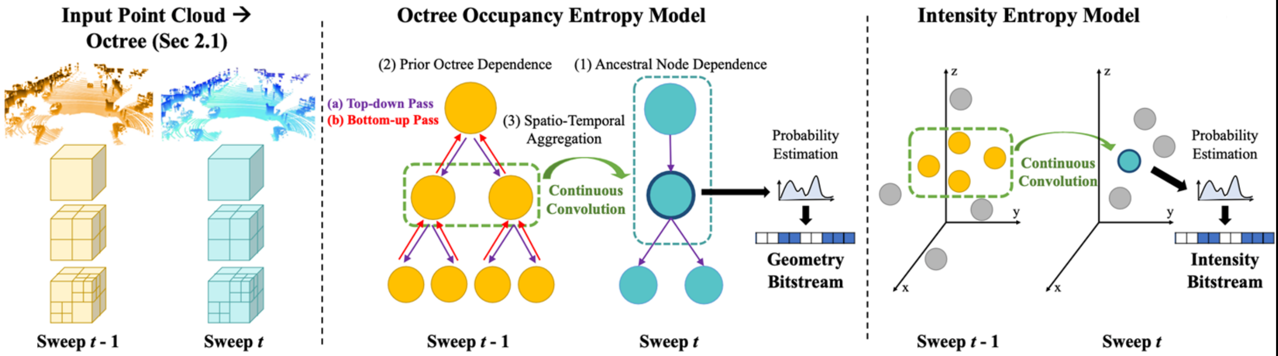
$$\mathbf{h}_i^{(0)} = \mathrm{MLP}^{(0)}(\mathbf{c}_i)$$

2. Adding the information of previos encoded nodes through parent

$$\mathbf{h}_i^{(k)} = \mathrm{MLP}^{(k)}([\mathbf{h}_i^{(k-1)}, \mathbf{h}_{\mathrm{pa}(i)}^{(k-1)}])$$

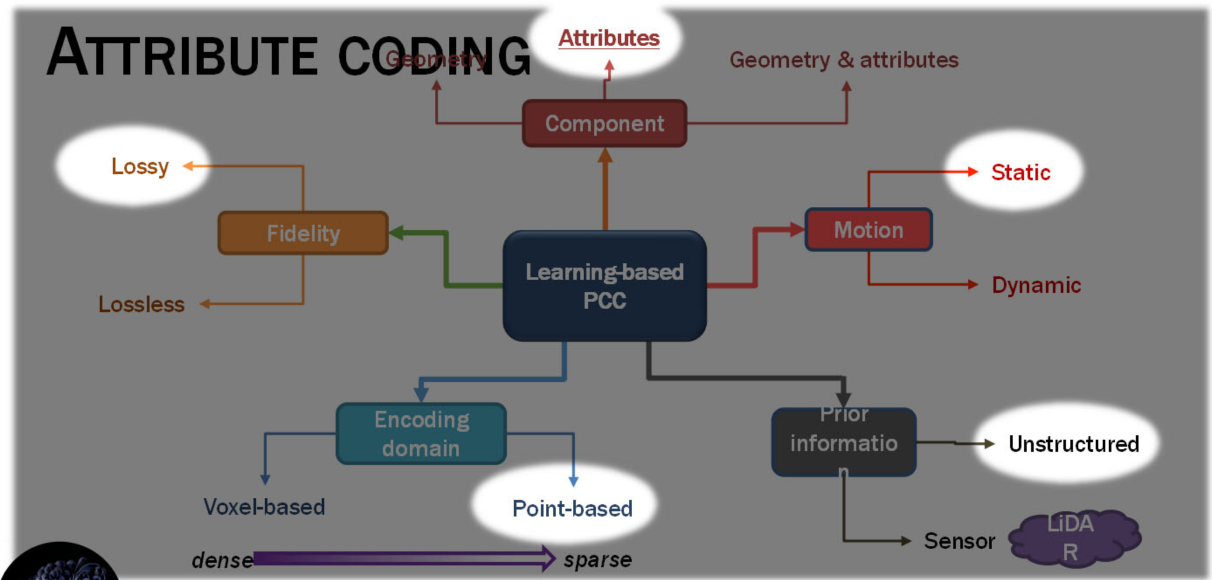
L. Huang, S. Wang, K. Wong, J. Liu and R. Urtasun, "OctSqueeze: Octree-Structured Entropy Model for LiDAR Compression," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 1310-1320

Extension to dynamic point clouds and attributes



Biswas, S., Liu, J., Wong, K., Wang, S. and Urtasun, R., 2020. MuSCLE: Multi Sweep Compression of LiDAR using Deep Entropy Models. *Advances in Neural Information Processing Systems, 33.*

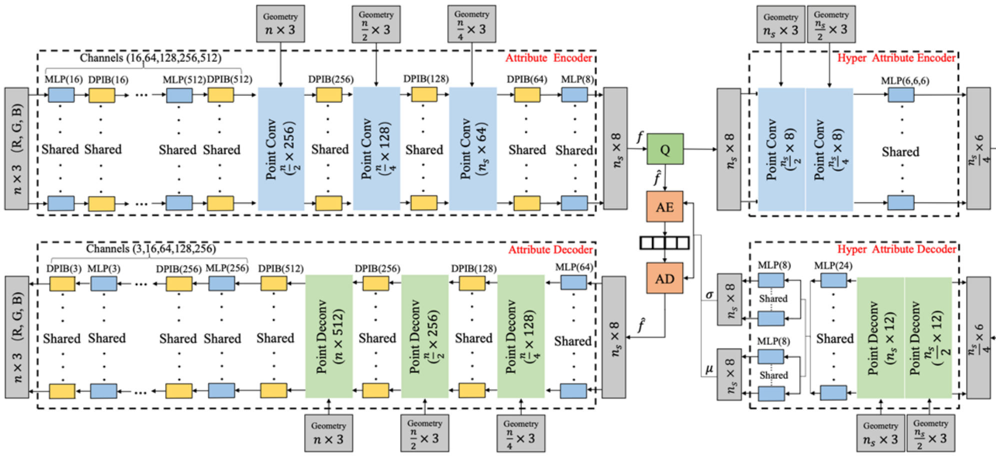
Attribute coding



13/04/2022

DEEP-PCAC

•



X. Sheng, L. Li, D. Liu, Z. Xiong, Z. Li and F. Wu, "Deep-PCAC: An End-to-End Deep Lossy Compression Framework for Point Cloud Attributes," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2021.3086711.

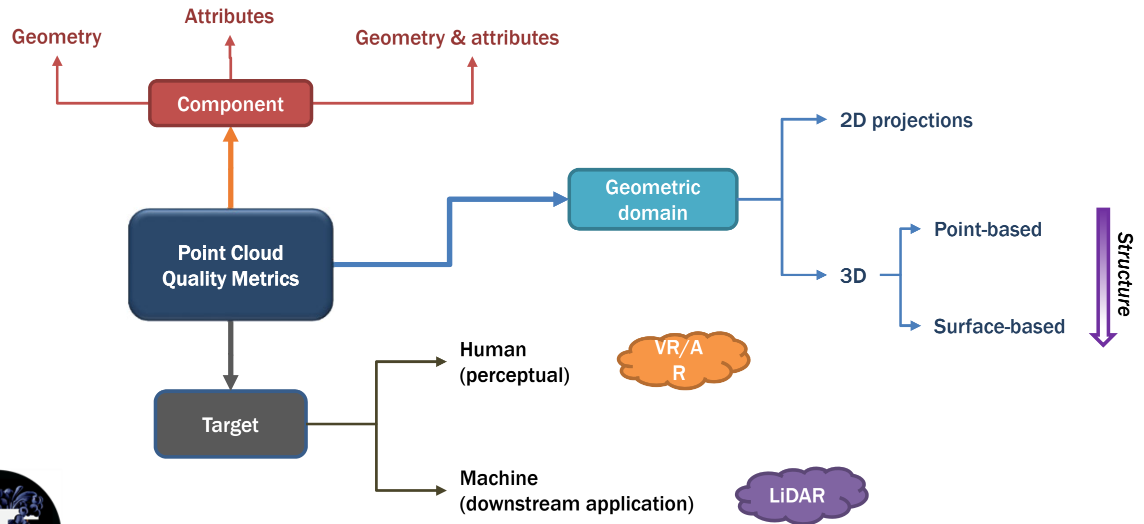
TAKE-AWAY ON LEARNING-BASED PCC

- Significantly inspired by recent advances in 2D learning-based compression
  - VAE, generative models (auto-regressive)
- Mainly two kinds of encoding backbones employed
  - Voxel-based convolution (sparse convolution possible)
  - Point-based (PointNet/PointNet++) convolution
- Geometry (occupancy) coding is cast as a classification problem
  - Adapting to varying spatial density is fundamental
- ...le done on dynamic PCs
- ...AR scans: special case

QUALITY ASSESSMENT AND BENCHMARK

- Objective quality metrics
- Performance of PC codecs

APPROACHES TO ASSESS POINT CLOUD QUALITY



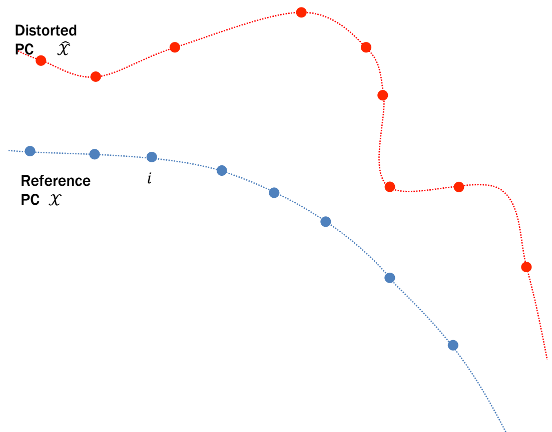
2D PROJECTION-BASED APPROACH

- Same principle as V-PCC
  - Joint geometry & texture
  - Use conventional 2D quality metrics on views
  - Fuse the score:
- Well-correlated wi

(a) $K = 6$ (b) $K = 12$ (c) $K = 42$

Torlig, E.M., Alexiou, E., Fonseca, T.A., de Queiroz, R.L. and Ebrahimi, T., 2018, September. A novel methodology for quality assessment of voxelized point clouds. In *Applications of Digital Image Processing XLI* (Vol. 10752, p. 107520I). International Society for Optics and Photonics.
E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1-6

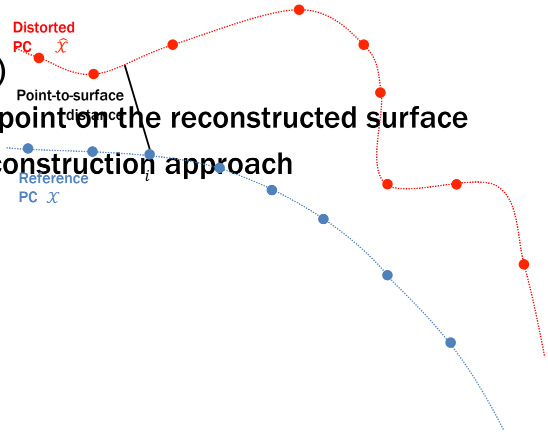
3D DISTANCE METRICS



D. Tian, H. Ochimizu, C. Feng, R. Cohen and A. Vetro, "Geometric distortion metrics for point cloud compression," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3460-3464

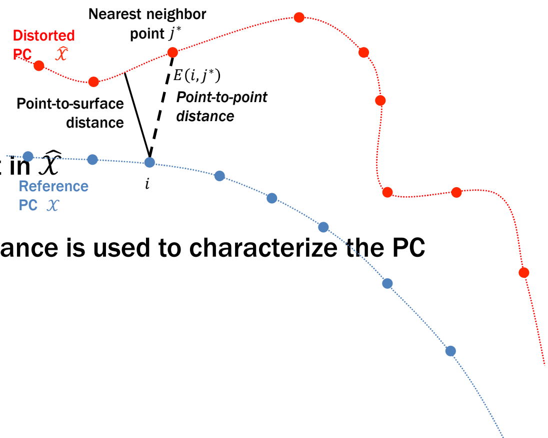
3D DISTANCE METRICS

- Point-to-surface (a.k.a. cloud-to-mesh)
  - For each point in $\mathcal{X}$, find the nearest point on the reconstructed surface
  - Heavily dependent on the surface reconstruction approach
  - Difficult to use for PC compression



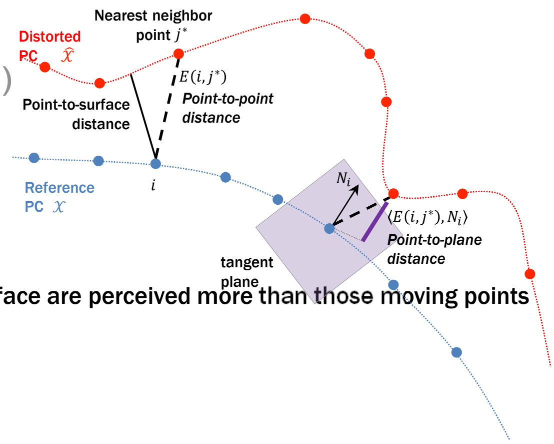
3D DISTANCE METRICS

- Point-to-surface (a.k.a. cloud-to-mesh)
- Point-to-point (a.k.a. cloud-to-cloud)
  - For each point in $\mathcal{X}$, find the nearest point in $\hat{\mathcal{X}}$
  - Euclidean distance in the 3D space
  - The average or maximum (Hausdorff) distance is used to characterize the PC distortion
  - Fails to account for surface structures



3D DISTANCE METRICS

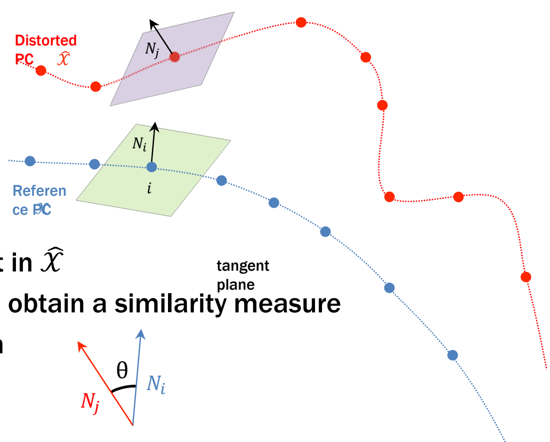
- Point-to-surface (a.k.a. cloud-to-mesh)
- Point-to-point (a.k.a. cloud-to-cloud)
- **Point-to-plane**
  - For each point in $\mathcal{X}$, find the nearest point in $\hat{\mathcal{X}}$
  - Project the error vector on the local normal
  - Errors that push points *away* from the local surface are perceived more than those moving points *along* the surface
  - Approximate the local surface with a plane



D. Tian, H. Ochimizu, C. Feng, R. Cohen and A. Vetro, "Geometric distortion metrics for point cloud compression," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3460-3464

3D DISTANCE METRICS

- Point-to-surface (a.k.a. cloud-to-mesh)
- Point-to-point (a.k.a. cloud-to-cloud)
- Point-to-plane
- **Plane-to-plane**
  - For each point in $\mathcal{X}$, find the nearest point in $\hat{\mathcal{X}}$
  - Compare the normals at the two points to obtain a similarity measure
  - Better correlations on octree compression



E. Alexiou and T. Ebrahimi, "Point Cloud Quality Assessment Metric Based on Angular Similarity," 2018 IEEE International Conference on Multimedia and Expo (ICME), 2018, pp. 1-6

3D DISTANCE METRICS

- **Point-to-Point → D1 metric**
- **Point-to-Plane → D2 metric**

- Error pooling
  - Mean squared error

$$MSE_{\mathcal{X},\hat{\mathcal{X}}}^{D_1} = \frac{1}{n_{\mathcal{X}}} \sum_{\forall i \in \mathcal{X}} \|E(i,j)\|_2^2$$

$$MSE_{\mathcal{X},\hat{\mathcal{X}}}^{D_2} = \frac{1}{n_{\mathcal{X}}} \sum_{\forall i \in \mathcal{X}} \|E(i,j) \cdot N_i\|_2^2$$

  - **Asymmetric!**

$$MSE_{\mathcal{X},\hat{\mathcal{X}}} \neq MSE_{\hat{\mathcal{X}},\mathcal{X}}$$

Typical symmetrization: $MSE_{\text{sym}}(\mathcal{X},\hat{\mathcal{X}}) = \max(MSE_{\mathcal{X},\hat{\mathcal{X}}}, MSE_{\hat{\mathcal{X}},\mathcal{X}})$

PSNR FOR GEOMETRY METRICS

- Traditionally, the Peak Signal-to-Noise Ratio is used in 2D image/video

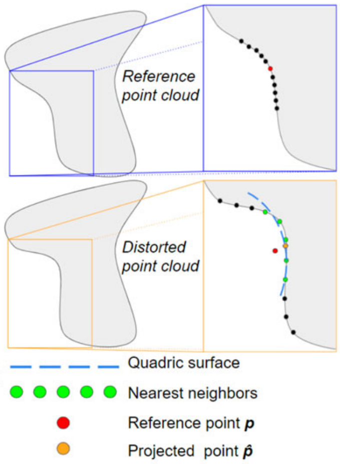
$$PSNR_{\text{sym}}(\mathcal{X},\hat{\mathcal{X}}) = 10 \log_{10} \frac{p^2}{MSE_{\text{sym}}(\mathcal{X},\hat{\mathcal{X}})}$$

- Normalization w.r.t. peak value p
  - The peak value should represent the energy of a pure noise signal
  - Easy to define for intensity, not for geometry…
  - Signal dependent
- Several solutions
  - For a voxelized PC with *b* bit-depth precision, $p = 2^b - 1$
  - Diagonal distance of bounding box
  - *Intrinsic resolution* (max or avg nearest neighbor distance)

D. Tian, H. Ochimizu, C. Feng, R. Cohen and A. Vetro, "Geometric distortion metrics for point cloud compression," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3460-3464
A. Javaheri, C. Brites, F. Pereira and J. Ascenso, "Improving PSNR-Based Quality Metrics Performance For Point Cloud Geometry," IEEE International Conference on Image Processing (ICIP), 2020, pp. 3438-3442

3D METRICS FOR GEOMETRY AND ATTRIBUTES

- Example: **PCQM**
  - quadric approximation for point matching (point-to-surface)
- Geometry-based features
  - Curvature comparison, contrast and structure
- Color-based features
  - Lightness comparison, contrast and structure
  - Chroma, hue comparison
- Linear combination of the features to obtain the global qua



G. Meynet, Y. Nehmé, J. Digne and G. Lavoué, "PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds," Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1-6

PERFORMANCE OF PC QUALITY METRICS

Alexiou et al.[1]
- 54 stimuli
- 20 subjects
- compression artifacts

	Inanimate objects				Human bodies			
	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
$\text{po2point}_{\text{MSE}}$	0.740	0.769	0.812	0.889	0.732	0.789	0.621	0.778
$\text{po2point}_{\text{Hausdorff}}$	0.735	0.758	0.819	0.889	0.732	0.781	0.621	0.778
$\text{po2plane}_{\text{MSE}}$	0.692	0.684	0.872	0.889	0.717	0.762	0.636	**0.741**
$\text{po2plane}_{\text{Hausdorff}}$	0.732	0.701	0.824	0.889	0.734	0.788	0.620	0.778
$\text{pl2plane}_{\text{RMS}}$	0.668	0.723	0.900	0.778	**0.782**	**0.813**	**0.568**	**0.741**
$\text{pl2plane}_{\text{MSE}}$	0.664	0.723	0.903	0.815	**0.782**	**0.813**	**0.568**	**0.741**
Color - PSNR_{YUV}	0.791	0.751	0.739	0.778	0.668	0.618	0.678	**0.741**
PSNR	0.739	0.672	0.814	0.704	0.740	0.771	0.613	0.815
SSIM	0.823	0.817	0.686	0.741	0.619	0.600	0.716	0.889
MS-SSIM	**0.884**	**0.855**	**0.566**	**0.630**	0.727	0.757	0.626	0.852
VIFP	0.693	0.645	0.871	0.778	0.662	0.566	0.683	0.778

SJTU dataset[2]
- 420 stimuli
- 64 subjects
- compression, noise, subsampling

3D metrics

Model	PLCC	SROCC	RMSE
MSE-p2point	0.0466	0.7009	2.4081
MSE-p2plane	0.0462	0.6881	2.4081
Hausdorff-p2point	0.6548	0.6189	1.8221
Hausdorff-p2plane	0.6325	0.6233	1.8673
PSNR-MSE-p2point	0.6699	0.7181	1.7898
PSNR-MSE-p2plane	0.6270	0.6669	1.8779
PSNR-Hausdorff-p2point	0.5988	0.5831	1.9307
PSNR-Hausdorff-p2plane	0.6129	0.5983	1.9048
PSNR-YUV	0.6311	0.6207	1.8701
PCQM	**0.8603**	**0.8465**	**1.2291**
Proposed	0.6076	0.6020	1.8635

[1]E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1-6
[2]Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang and J. Sun, "Predicting the Perceptual Quality of Point Cloud: A 3D-to-2D Projection-Based Exploration," in IEEE Transactions on Multimedia, Oct. 2020

PERFORMANCE OF PC QUALITY METRICS

Alexiou et al.[1]
- 54 stimuli
- 20 subjects
- compression artifacts

	Inanimate objects				Human bodies			
	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
$\text{po2point}_{\text{MSE}}$	0.740	0.769	0.812	0.889	0.732	0.789	0.621	0.778
$\text{po2point}_{\text{Hausdorff}}$	0.735	0.758	0.819	0.889	0.732	0.781	0.621	0.778
$\text{po2plane}_{\text{MSE}}$	0.692	0.684	0.872	0.889	0.717	0.762	0.636	**0.741**
$\text{po2plane}_{\text{Hausdorff}}$	0.732	0.701	0.824	0.889	0.734	0.788	0.620	0.778
$\text{pl2plane}_{\text{RMS}}$	0.668	0.723	0.900	0.778	**0.782**	**0.813**	**0.568**	**0.741**
$\text{pl2plane}_{\text{MSE}}$	0.664	0.723	0.903	0.815	**0.782**	**0.813**	**0.568**	**0.741**
Color - PSNR_{YUV}	0.791	0.751	0.739	0.778	0.668	0.618	0.678	**0.741**
PSNR	0.739	0.672	0.814	0.704	0.740	0.771	0.613	0.815
SSIM	0.823	0.817	0.686	0.741	0.619	0.600	0.716	0.889
MS-SSIM	**0.884**	**0.855**	**0.566**	**0.630**	0.727	0.757	0.626	0.852
VIFP	0.693	0.645	0.871	0.778	0.662	0.566	0.683	0.778

SJTU dataset[2]
- 420 stimuli
- 64 subjects
- compression, noise, subsampling

2D metrics

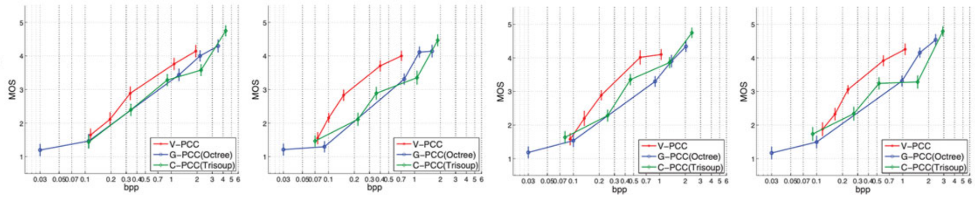
Model	MOS		
	PLCC	SROCC	RMSE
PSNR	0.2481	0.2512	2.3354
PSNR-HVS-M	0.2382	0.2615	2.3413
SSIM	0.3654	0.2789	2.2440
MS-SSIM	0.3659	0.2592	2.2437
IW-SSIM	0.4339	0.3285	2.1720
FSIM	0.3196	0.3019	2.2843
VIF	0.5243	0.5647	2.0653
NIQE	0.3262	-0.1149	2.2788
IL-NIQE	0.2703	-0.0478	2.3210
OG-IQA	0.1163	0.0214	2.3943
Proposed	**0.6076**	**0.6020**	**1.8635**

[1]E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), 2019, pp. 1-6
[2]Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang and J. Sun, "Predicting the Perceptual Quality of Point Cloud: A 3D-to-2D Projection-Based Exploration," in IEEE Transactions on Multimedia, Oct. 2020

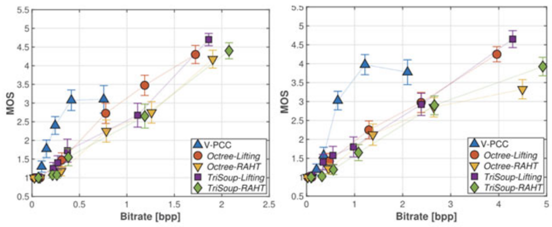
QUALITY ASSESSMENT AND BENCHMARK

- Objective quality metrics
- Performance of PC codecs

PERFORMANCE ASSESSMENT: G-PCC VS. V-PCC



(a) *Long Dress* (b) *Loot* (c) *Soldier* (d) *Red and Black*



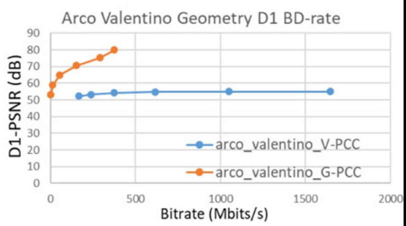
(e) *loot* (f) *longdress*

S. Perry et al., "Quality Evaluation Of Static Point Clouds Encoded Using MPEG Codecs," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 3428-3432, doi: 10.1109/ICIP40778.2020.9191308.

Alexiou, E., Viola, I., Borges, T., Fonseca, T., De Queiroz, R., & Ebrahimi, T. (2019). A comprehensive study of the rate-distortion performance in MPEG point cloud compression. *APSIPA Transactions on Signal and Information Processing, 8*, E27. doi:10.1017/ATSIP.2019.20

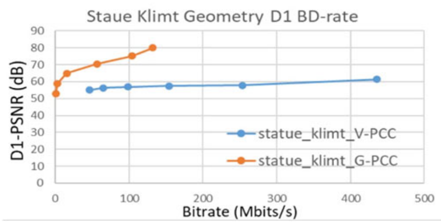
PERFORMANCE ASSESSMENT: G-PCC VS. V-PCC

Façade *Arco Valentino* *House without a roof*



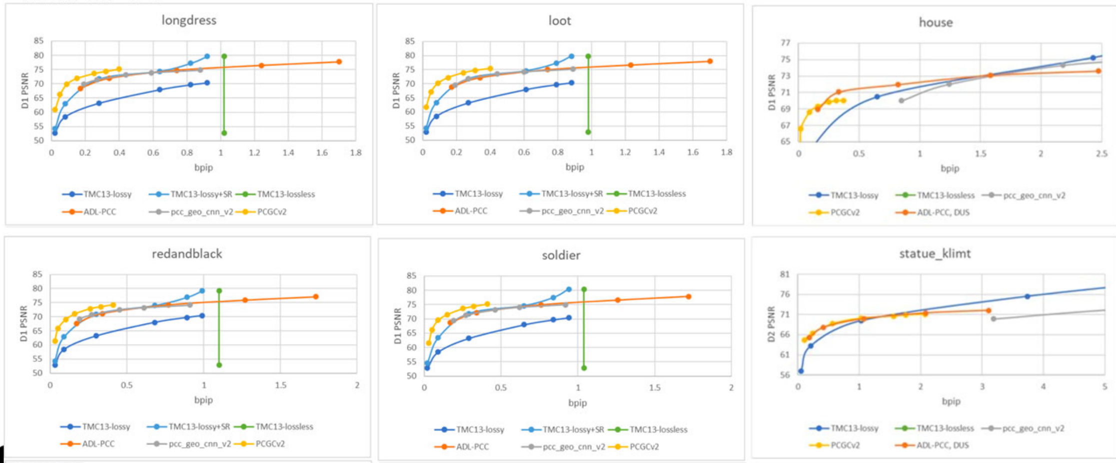
- Objects (sparse to dense, 11-20 bits per coordinate)

Egyptian mask *Shiva* *Klimt statue* *Frog*



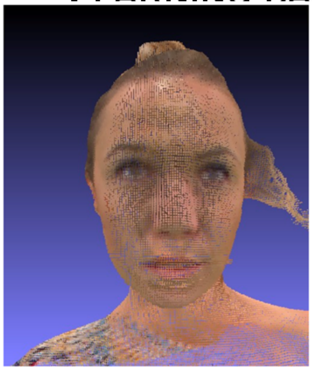
C. Cao, M. Preda, V. Zakharchenko, E. S. Jang and T. Zaharia, "Compression of Sparse and Dense Dynamic Point Clouds—Methods and Standards," in Proceedings of the IEEE, vol. 109, no. 9, pp. 1537-1558, Sept. 2021

GEOMETRY CODING: LEARNING-BASED CODECS VS. MPEG

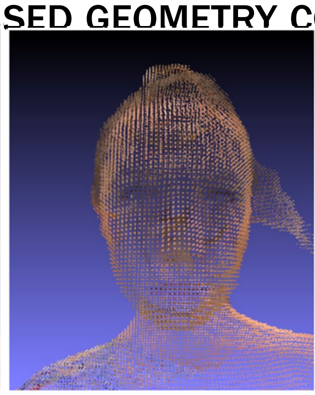


MPEG 3DGC, Performance analysis of currently AI-based available solutions for PCC, ISO/IEC JTC 1/SC 29/WG 7 N233, October 2021

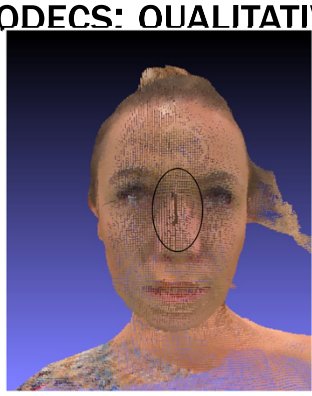
LEARNING-BASED GEOMETRY CODECS: QUALITATIVE RESULTS



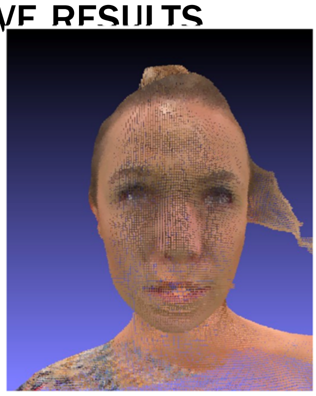
uncompressed



GPCC (TMC13v14)
0.28 bpip
D1-PSNR: 63.21 dB
D2-PSNR: 68.46 dB



Quach et al. (2020)
0.27 bpip
D1-PSNR: 71.41 dB
D2-PSNR: 74.99 dB

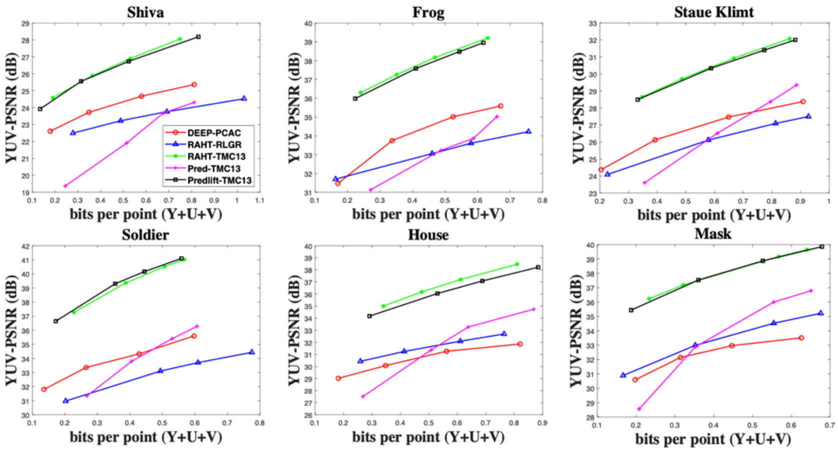


Wang et al. (2021)
0.25 bpip
D1-PSNR: 73.60 dB
D2-PSNR: 77.41 dB

MPEG 3DGC, Performance analysis of currently AI-based available solutions for PCC, ISO/IEC JTC 1/SC 29/WG 7 N233, October 2021
Maurice Quach, Giuseppe Valenzise, Frédéric Dufaux. Improved Deep Point Cloud Geometry Compression. *IEEE International Workshop on Multimedia Signal Processing (MMSP'2020)*, Sep 2020, Tampere, Finland.
J. Wang, D. Ding, Z. Li and Z. Ma, "Multiscale Point Cloud Geometry Compression," *Data Compression Conference (DCC)*, 2021

ATTRIBUTE CODING: LEARNING-BASED METHODS VS. MPEG

- Still lagg



X. Sheng, L. Li, D. Liu, Z. Xiong, Z. Li and F. Wu, "Deep-PCAC: An End-to-End Deep Lossy Compression Framework for Point Cloud Attributes," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2021.3086711.

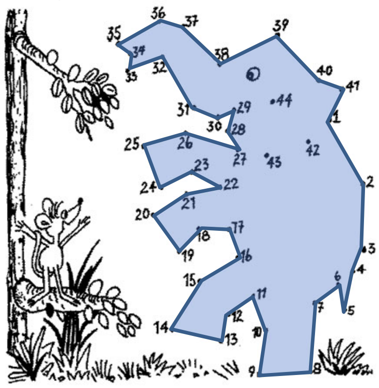
TAKE-AWAY ON PC QUALITY ASSESSMENT AND PCC BENCHMARK

- Quality metrics
  - 2D metrics appropriate for dense PC and distortions that do not significantly change density
  - Point-to-point easier to embed in end-to-end learning-based codecs
  - *No clear consensus on which is the good metric to use!*
- Benchmark of PC coding approaches
  - V-PCC outperforms G-PCC (only) on dense point clouds
  - Voxel-based VAE coding methods achieve state-of-the-art performance in coding *geometry* of dense PC
  - MPEG codecs achieve state-of-the-art performance on *attribute* compression
  - *A thorough subjective evaluation of learning-based codecs still missing!*

Trends and summary

Open Challenges in Point Cloud Coding and Quality Assessment

- Capture the underlying geometric structure
  - Variable spatial density
  - Extremely sparse sampling
  - Prior information: joint semantic interpretation and coding?
  - Modeling the acquisition
  - Perceptual loss?
- Joint geometry and attribute coding
  - Interdependence
- Perceptual quality assessment
  - Methodologies
  - Large dataset construction



Q&A

giuseppe.valenzise@centralesupelec.fr

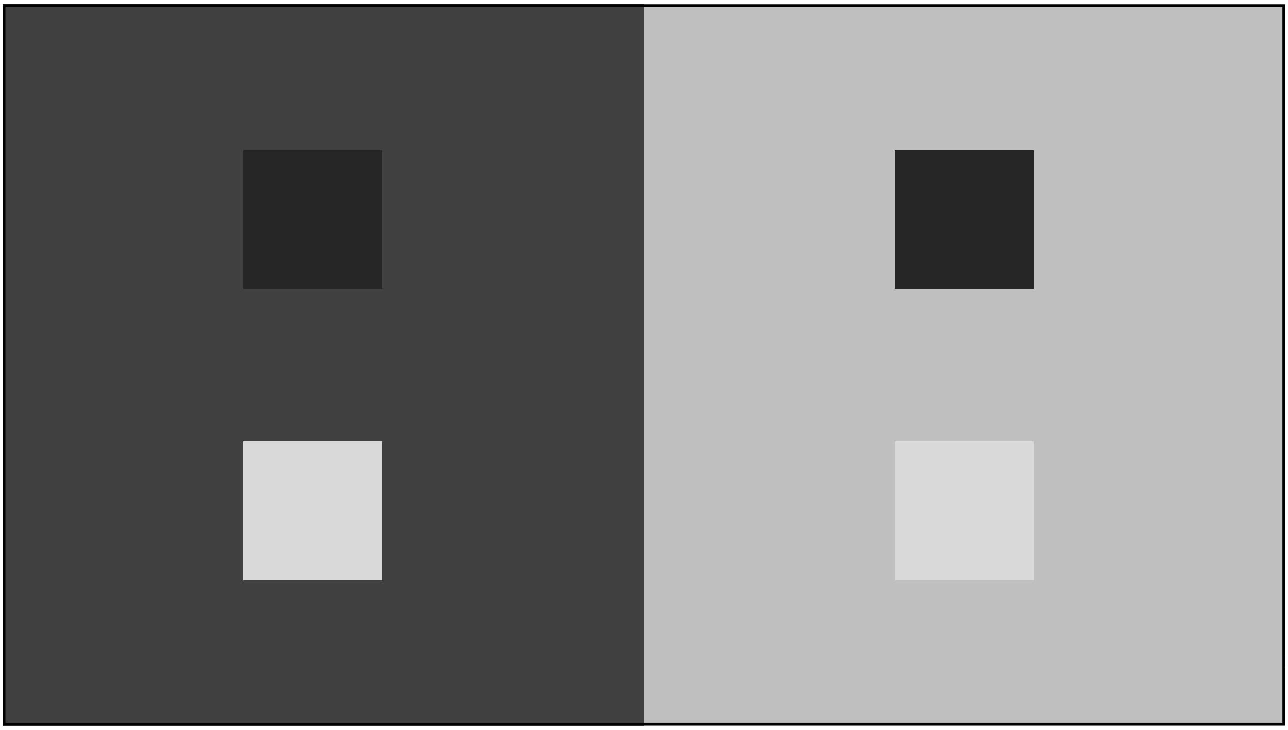
Immersive 3D HDR visualisation

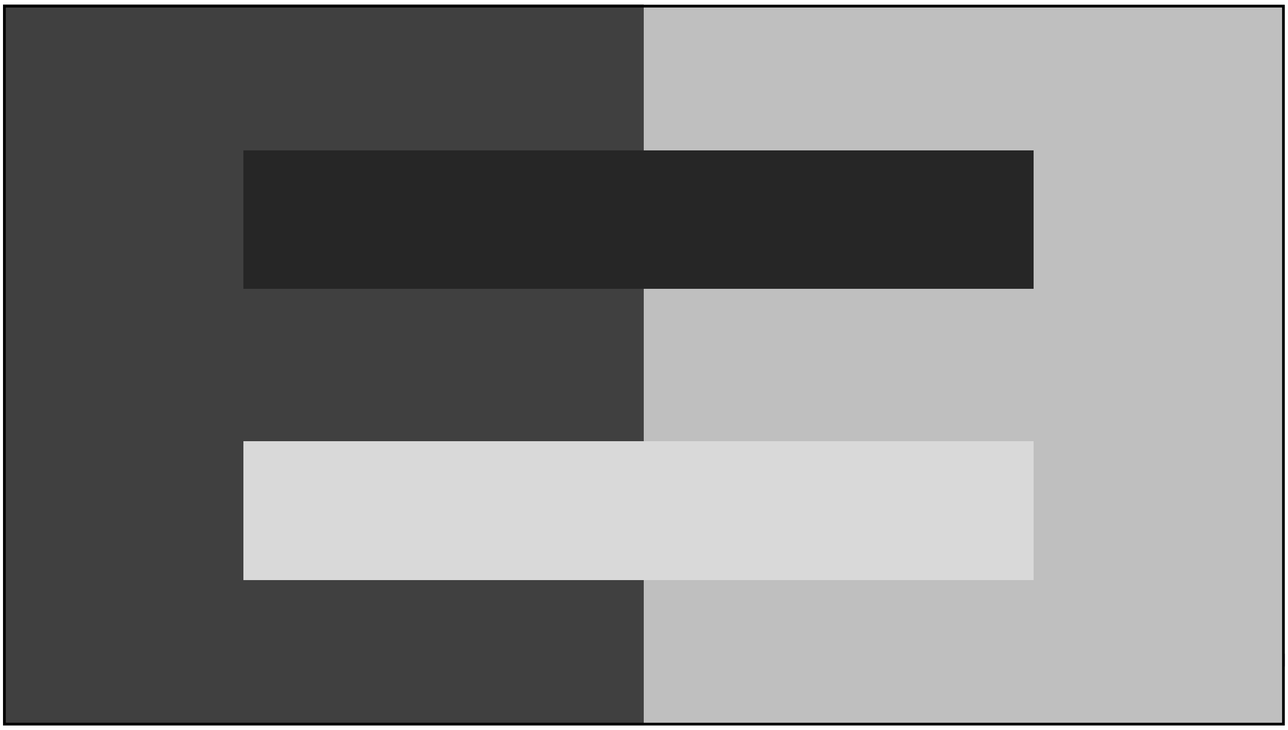
Ific Goudé, Rémi Cozot (speaker)

- **Evaluation of lightness and color perception on HMDs**
- A TMO for visualization of HDR panoramas on HMDs
- A TMO for HDR 3D scenes

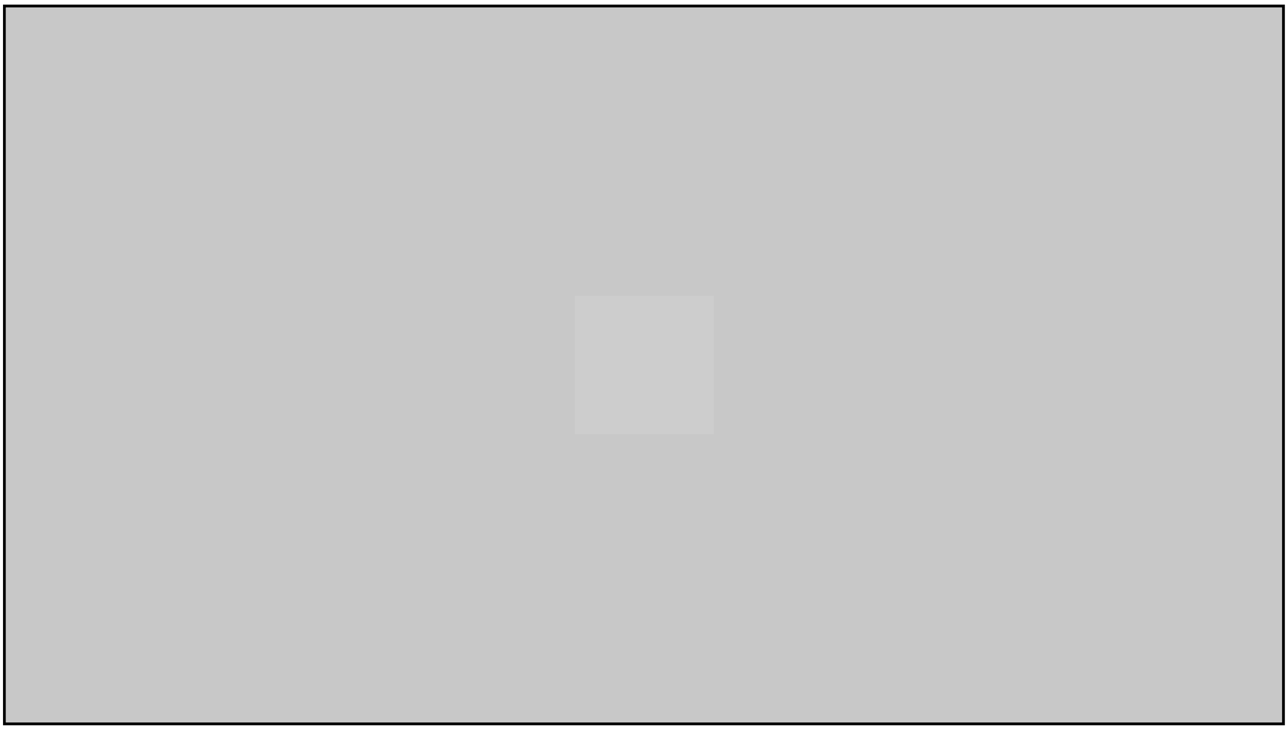
Human eye perception on Head Mounted Display

13/04/2022

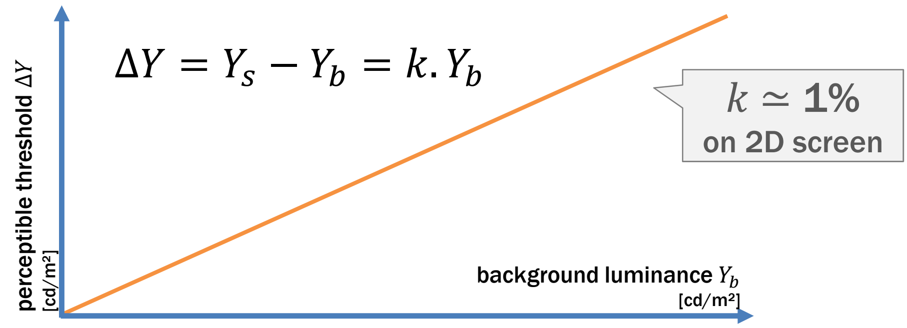




13/04/2022



The minimal perceptible threshold is proportional to the background luminance



The derivate of the response relative to the luminance: **lightness**

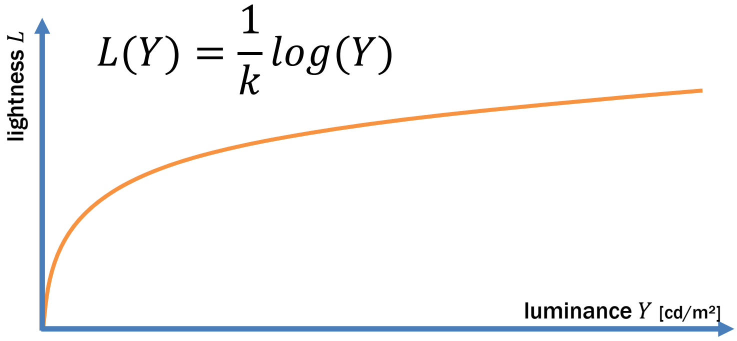
$$\frac{dL}{dY}(Y_b) = \frac{1}{\Delta Y(Y_b)}$$

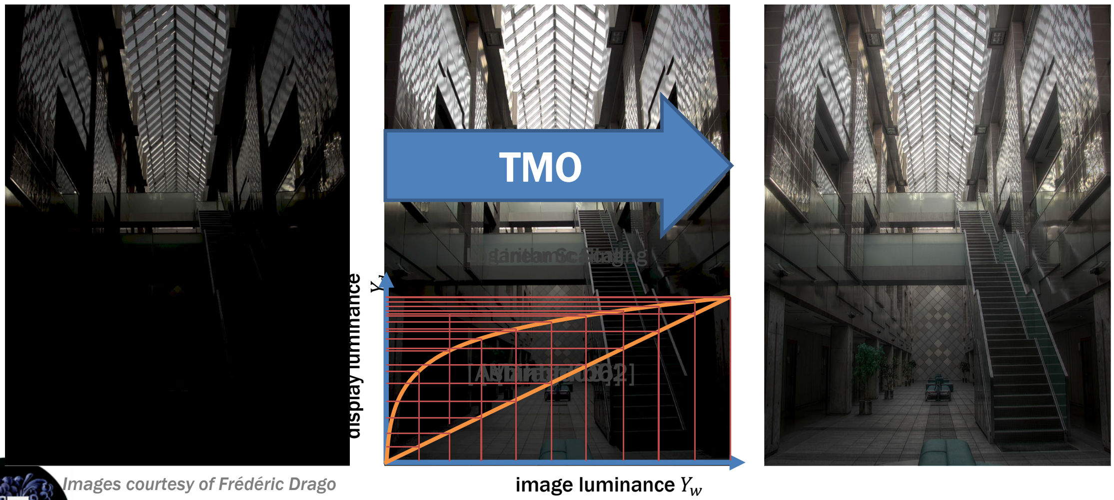
$$L(Y_b) = \int_0^{Y_b} \frac{1}{\Delta Y(Y_b)} dY$$

Weber: $\Delta Y = kY_b$

$$L(Y_b) = \int \frac{1}{kY_b} dY_b = \frac{1}{k} \times log(Y_b) + a$$

Our sensitive response is logarithmic





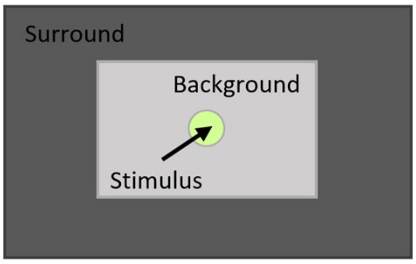
Images courtesy of Frédéric Drago

Do we have the same perception on HMD?

- Linearly proportional to the background?
- Logarithmic response?
- Same constant factor?
  - k ≃ 1% for screen visualization

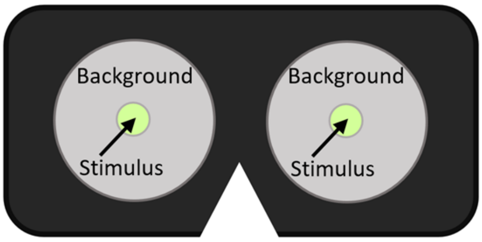
Lightness perception on HMD

CIECAM02

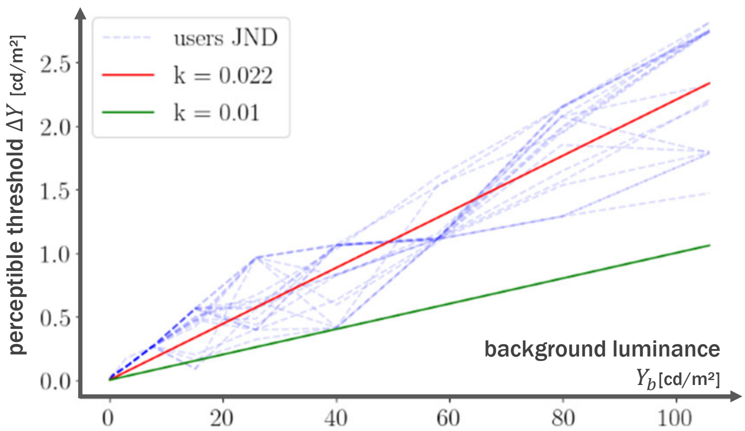


- Stimulus: 2°/ 4°
- Background: 20°
- Surround: Field of view

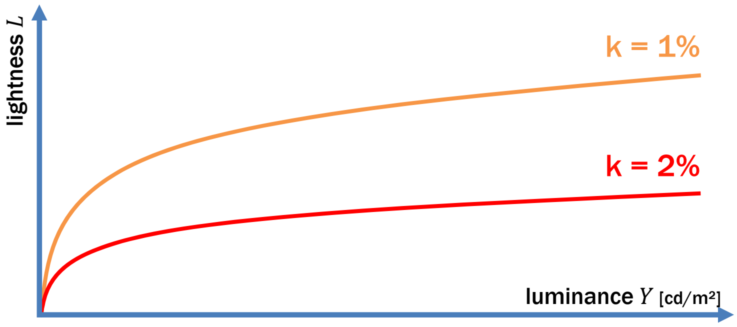
HMD



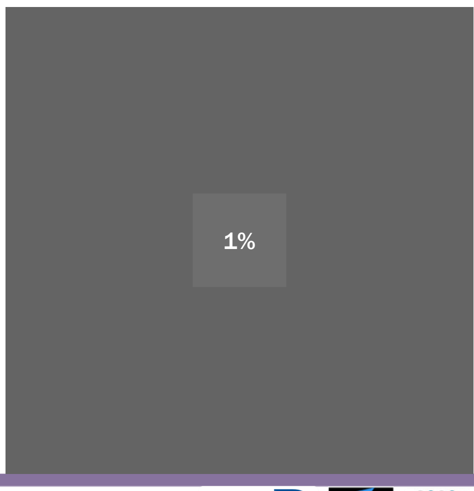
- Stimulus: 2°/ 4°
- Background: 100° (HMD FoV)
- ~~Surround: None~~



The sensitive response is still logarithmic

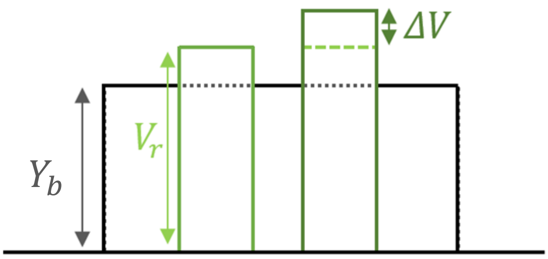


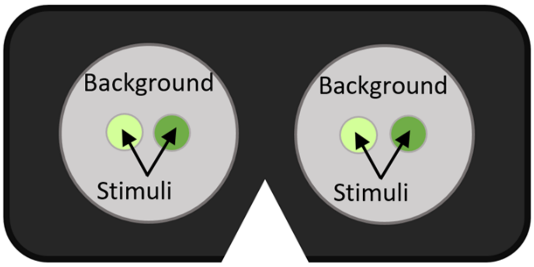
Stimulus is perceptible on 2D screen



Not anymore on HMD

- **Comparing two stimuli for a solid background**





Fechner (1860)

– $L = \frac{1}{k}\log(Y) + a$

CIELAB (1976)

– $L^* = 116\, f\left(\frac{Y}{Y_n}\right) - 16$ $\qquad$ $with\ f(t) = \begin{cases} t^{1/3}\ si\ t < \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29}\ sinon \end{cases}$

CIECAM02 (2002)

– $J = 100\left(\frac{A}{A_W}\right)^{c.z}$ $\qquad$ $with\ z = 1.48 + \sqrt{\frac{Y_b}{Y_W}},\ and\ c = \begin{cases} 0.525 & for\ Dark\ env \\ 0.590 & for\ Dim\ \ env \\ 0.690 & for\ Avg\ \ env \end{cases}$

Visualization conditions

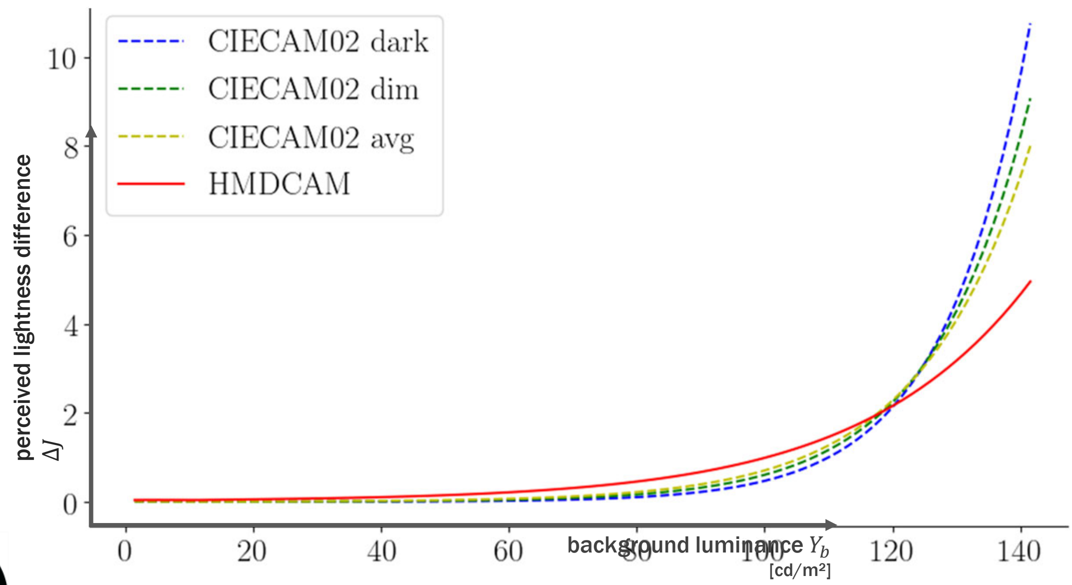
CIECAM02

$$-J = 100\left(\frac{A}{A_w}\right)^{c.z}$$

HMDCAM

$$-J = 100\left(\frac{A}{A_w}\right)^{c_L.z}$$

- $J = 100\left(\frac{A}{A_w}\right)^{c_L.z}$ $\quad with\ z = 1.48 + \sqrt{\frac{Y_b}{Y_w}}, \quad and\ c_L = \frac{c.r.\Delta Y_{a|Y_a=50}}{\Delta Y_a}$
- $\Delta Y_a = 1.88\ Y_a{}^{0.23} - 7.24\ Y_a{}^{0.11} + 8.26$
- $Y_a = F.Y_b + 0.2\ (1 - \mathrm{F})Y_{d_{max}}$
- $F = \begin{cases} 0.7379 + 0.392\ (1 - \exp(0.0221\ Y_b)), & if\ Y_b < 50\ cd/m^2 \\ 1 \quad , & otherwise \end{cases}$
- $r = \frac{0.01}{\mathrm{k}}$
- $\mathrm{k} = 0.022$



Error of the estimated perception for luminance and color [Goude20]

Background luminance *[cd/m²]*	15	50	90	125
CIECAM02 *(avg)* error [%]	13.1	18.9	17.3	9.7
HMDCAM error [%]	3.8	7.1	8.2	5.2

Color	Red	Green	Blue	Yellow
CIECAM02 *(avg)* error [%]	1.3	0.6	3.2	5.3
HMDCAM error [%]	0.7	0.5	1.7	2.8

- Evaluation of lightness and color perception on HMDs
- **A TMO for visualization of HDR panoramas on HMDs**
- A TMO for HDR 3D scenes

Tone Mapping High Dynamic Range 3D point cloud

HDR viewport

Viewport TMO [Reinhard02]

- $\bar{L}_w(t) = \frac{1}{N} \exp\left(\sum_{x,y} \log(Y_w(x,y,t) + \delta)\right)$
- $L(x,y,t) = \frac{a}{\bar{L}_w(t)} Y_w(x,y,t)$

$\bar{L}_w$ is the key-value
a is a user-defined parameter

- $\boldsymbol{Y_d(x,y,t) = \frac{L(x,y,t) \times \left(1 + \frac{L(x,y,t)}{Y_{w_{max}}(t)^2}\right)}{1 + L(x,y,t)}}$

Viewport TMO

Viewport TMO + eye adaptation [Yu15]

- $\bar{L}'_w(t) \quad = \quad \tau.\bar{L}_w(t) \quad + \quad (1-\tau).\bar{L}'_w(t-1)$
- $Y'_{w_{max}}(t) = \tau.Y_{w_{max}}(t) + (1-\tau).Y'_{w_{max}}(t-1)$
- $L(x,y,t) = \frac{a}{\bar{L}'_w(t)} L_w(x,y,t)$
- $\boldsymbol{Y_d(x,y,t) = \frac{L(x,y,t)\times\left(1+\frac{L(x,y,t)}{Y'_{w_{max}}(t)^2}\right)}{1+L(x,y,t)}}$

Viewport TMO + eye adaptation

Viewport TMO + eye adaptation

– Spatial coherency is not preserved

Global TMO

Global TMO [Ward97]

- **The Cumulative Distribution Function of the log-luminance histogram**

  $P\left(\log\left(Y_w(x,y)\right)\right)$

- **Histogram ceiling based on** our HMDCAM

- Scaled in the dynamic range of the display

  $$\boldsymbol{Y_d(x,y) = e^{ln\left(Y_{d_{min}}\right) + \left(ln\left(Y_{d_{max}}\right) - ln\left(Y_{d_{min}}\right)\right) \times P(Y_w(x,y))}}$$

Global TMO

HMD-TMO [Goude19]

- Linear combination of both **Viewport** and **Global** TMOs
  in the logarithmic domain
- $Y_d(x, y, t) = \alpha.\log\big(G(x, y)\big) + (1 - \alpha).\log(V(x, y, t))$

α is a user-defined weight

- $\boldsymbol{Y_d(x, y, t) = G(x, y)^{\alpha} \times V(x, y, t)^{1-\alpha}}$

HMD-TMO

HMD-TMO

– Spatial coherency is preserved

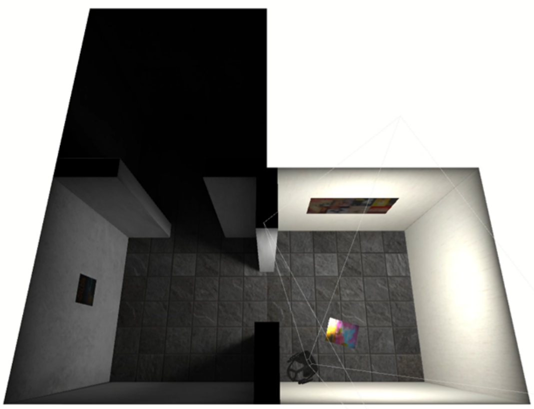
– Viewport contrasts are enhanced

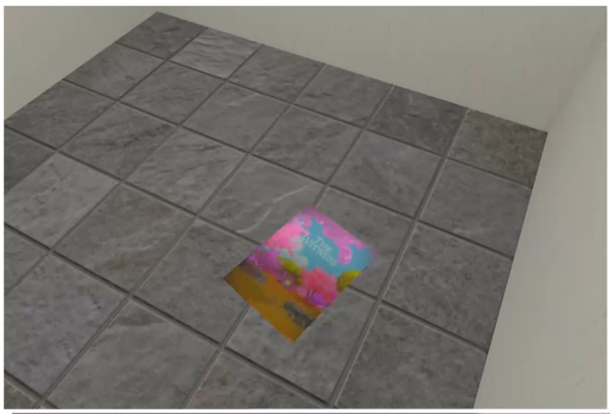
– Objective metric shows a better TMO quality

	[Reinhard02]	[Ward97]	[Yu15]	HMD-TMO [Goude19]
TMQI quality [TMQI]	0.798	0.854	0.865	**0.887**

- Evaluation of lightness and color perception on HMDs
- A TMO for visualization of HDR panoramas on HMDs
- **A TMO for HDR 3D scenes**

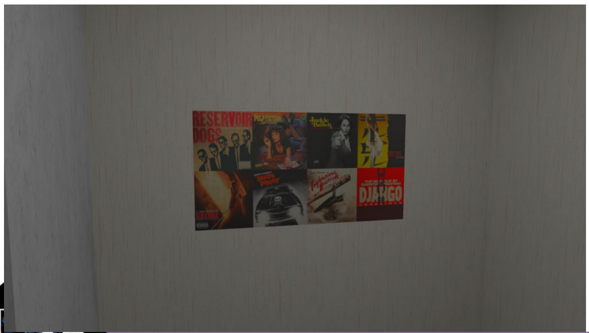
Viewport TMO + eye adaptation

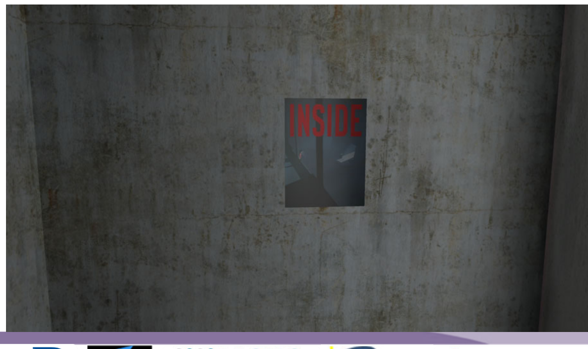




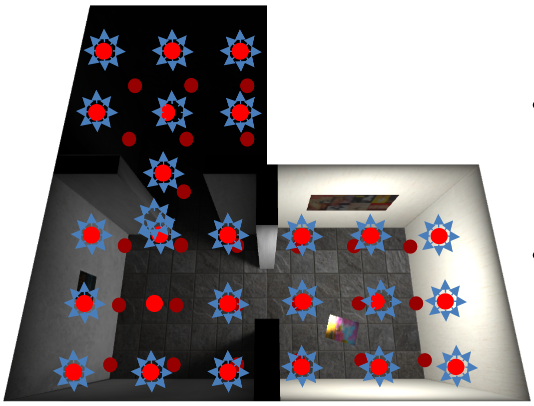
Viewport TMO + eye adaptation

– Spatial coherency is not preserved



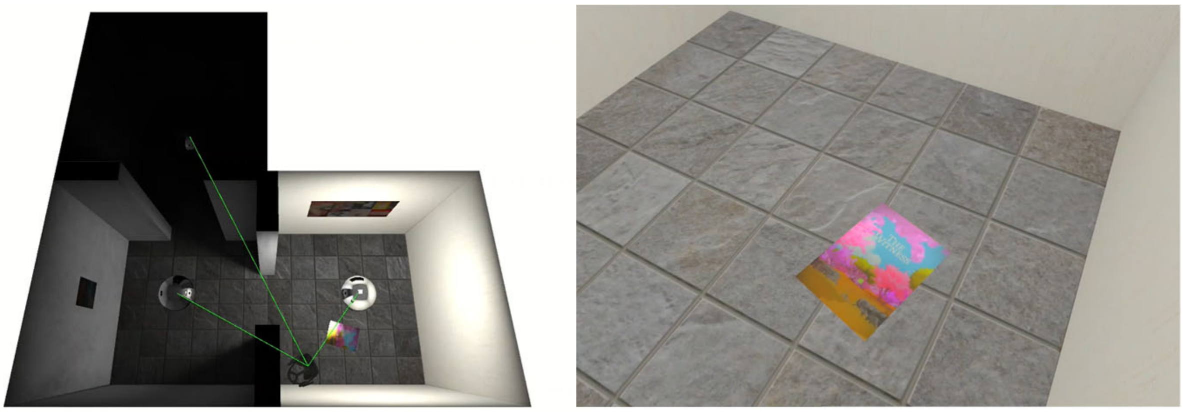


Global TMO



- What is the dynamic range of a 3D scene?
  - Light field
- Plenoptic function $L(x, y, z, \theta, \varphi, t)$ [Adelson91]
  - Camera position (x, y, z)
  - Camera orientation (θ, φ)
- To compute in real-time → impossible!

Global TMO



3D-TMO

3D-TMO

- Spatial coherency is preserved
- Viewport contrasts are enhanced
- Subjective studies show that our TMO is preferred

Fidelity	[Drago03]	Eye-adaptation (Unity)	3D-TMO [Goude20bis]
Scene 1	6.39 $\pm$ 1.82	6.56 $\pm$ 1.89	**7.50 $\pm$ 1.15**
Scene 2	6.17 $\pm$ 1.69	7.00 $\pm$ 1.64	**7.39 $\pm$ 1.58**

Global appreciation	[Drago03]	Eye-adaptation (Unity)	3D-TMO [Goude20bis]
Scene 1	6.83 $\pm$ 2.20	6.94 $\pm$ 1.66	**7.44 $\pm$ 1.38**
Scene 2	6.28 $\pm$ 1.81	7.22 $\pm$ 1.86	**7.61 $\pm$ 1.65**

Future work

For the ReVeRY project

- Combine 3D and HDR reconstruction
- Add multiexposure
- Add time coherence for video

Other directions

- More general camera configurations