Point Cloud Coding: The *Status Quo*

João Ascenso, Fernando Pereira
Instituto Superior Técnico – Universidade de Lisboa
Instituto de Telecomunicações
Lisboa – PORTUGAL

IEEE International Conference on Multimedia and Expo (ICME)
London, UK
10th July 2020

Tutorial Outline

1. 3D Visual Representation and Coding (F.Pereira)
2. Plenoptic Function based Imaging (F.Pereira)

Questions

3. Point Cloud Coding: Basic Approaches (F.Pereira)

Questions

4. Point Cloud Coding: Standardization, part 1 (J.Ascenso)

Questions

Point Cloud Coding: Standardization, part 2 (J.Ascenso)

5. Point Cloud Quality Assessment (J.Ascenso)
6. Summary and Trends (J.Ascenso)

Questions

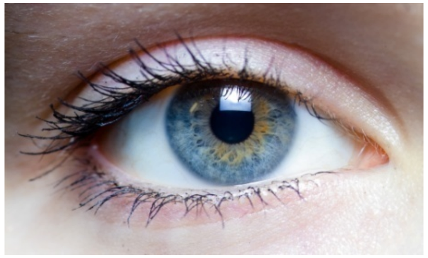
TÉCNICO LISBOA

1

3D Visual Representation and Coding


Light is Life ... Both are Complex !

The World and the Humans ...



Visual, Visual, Visual …

- **It is believed that up to 50% of the human brain is involved in some way in processing visual information**
  - Reflects the significance of vision for <u>function and survival</u>
  - Also explains its capacity to <u>entertain, and inform</u>

- **Visual experiences are important drivers:**
  - Nearly two-thirds of the global population will have Internet access by 2023
  - 5G speeds will be 13 times higher than the average mobile connection by 2023
  - By 2023, 66 percent of connected flat-panel TV sets will be 4K
  - By 2022, video traffic, which includes internet video and IP video on demand (VOD), will account for 82% of the total global IP-traffic

- **New, more immersive and effective visual experiences are continuously asked for !**

from Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper

Visual Representation and Coding: What and Why ?

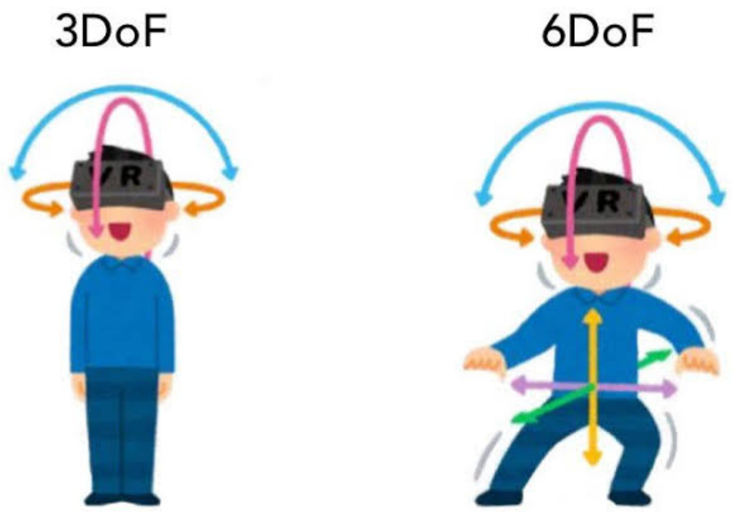
- Replicating the visual world in an efficient way
- Driven/conditioned by available sensors, transmission/storage channels, displays and devices
- .... and by the Human Visual System
- To offer in an efficient, effective, immersive, resilient, scalable, adaptive, simple, ... way
- The relevant set of funcionalities
- For each target application/service
- To provide the best USER EXPERIENCE !

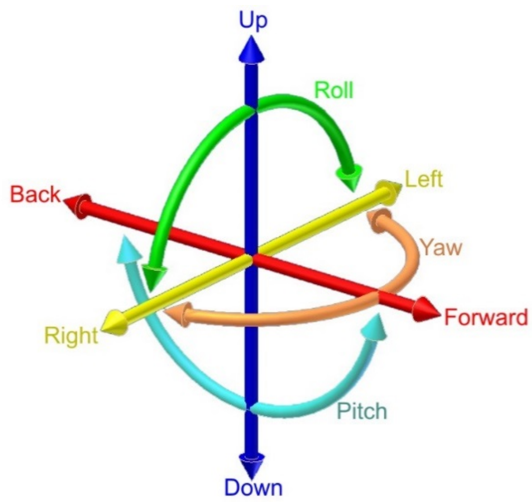
Immersion as Much as ... Real

Creating Virtual Realities ...

How to Assess Immersion: The Visual Degrees of Freedom

- **Degrees of Freedom (DoF) refer to the movement of a rigid body inside space, this means the *"different basic ways in which an object can move"*.**
- **There are 6-DoF in total, essentially 3 translations and 3 rotations:**
  - Translations: a body is free to translate in 3 degrees of freedom, forward/back, up/down, left/right.
  - Rotations: a body can also rotate with 3 degrees of freedom, pitch, yaw, and roll.



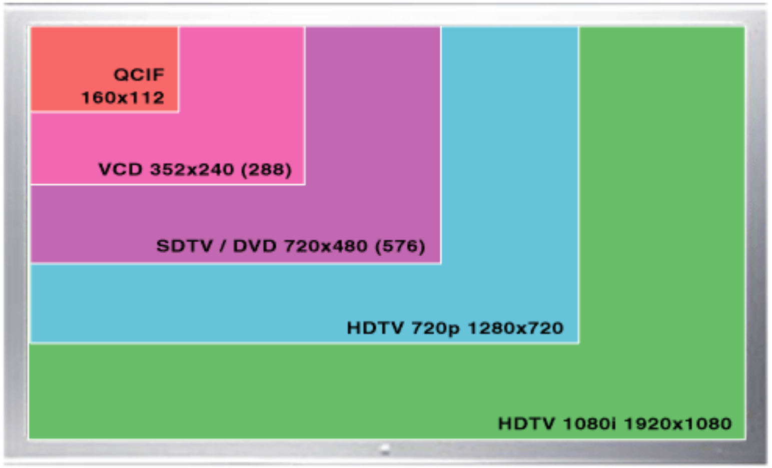


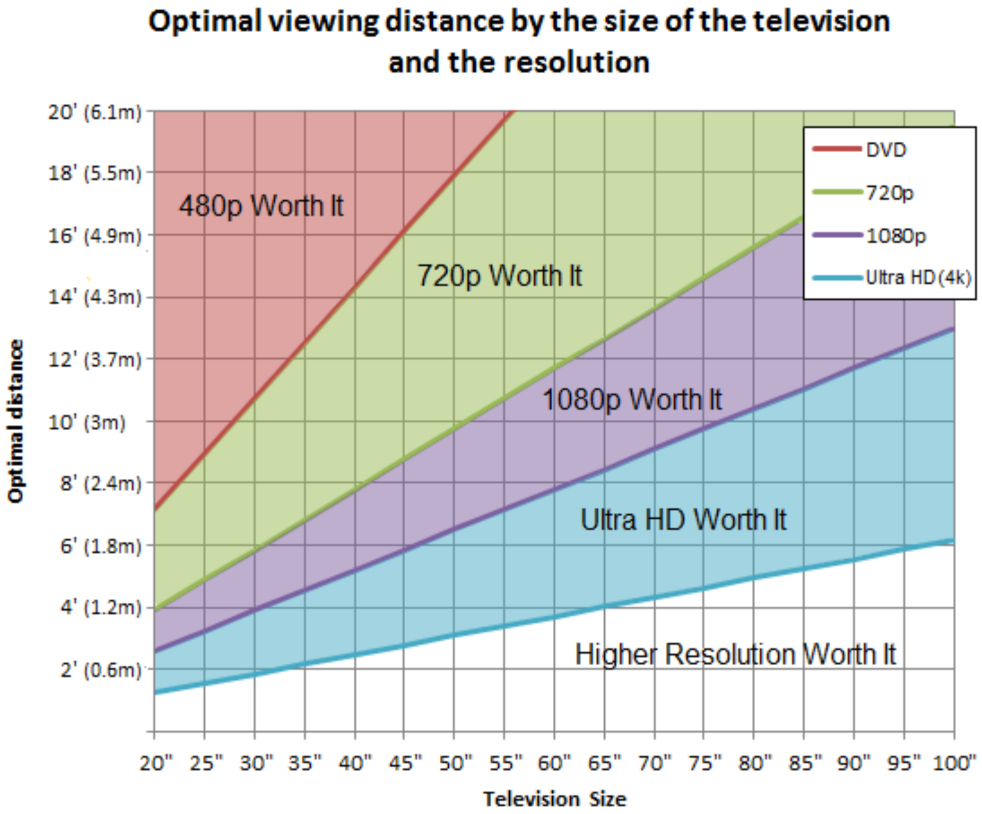
For a Long Time: 0-DoF Experiences

The Frame-based Video Model: a 2D Window to the World

0-DoF

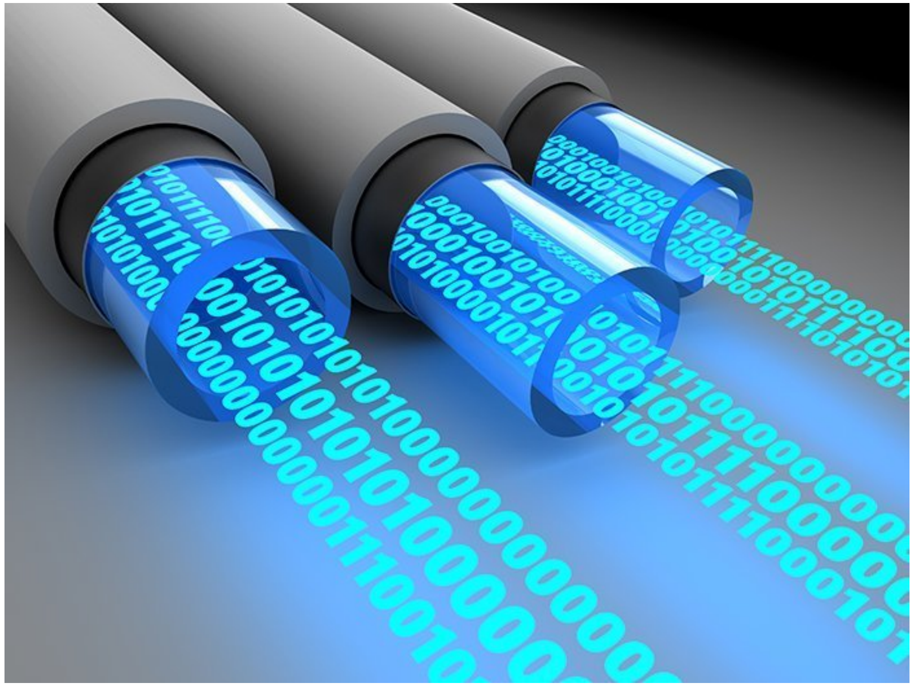
Spatial Resolution: Racing for Immersion



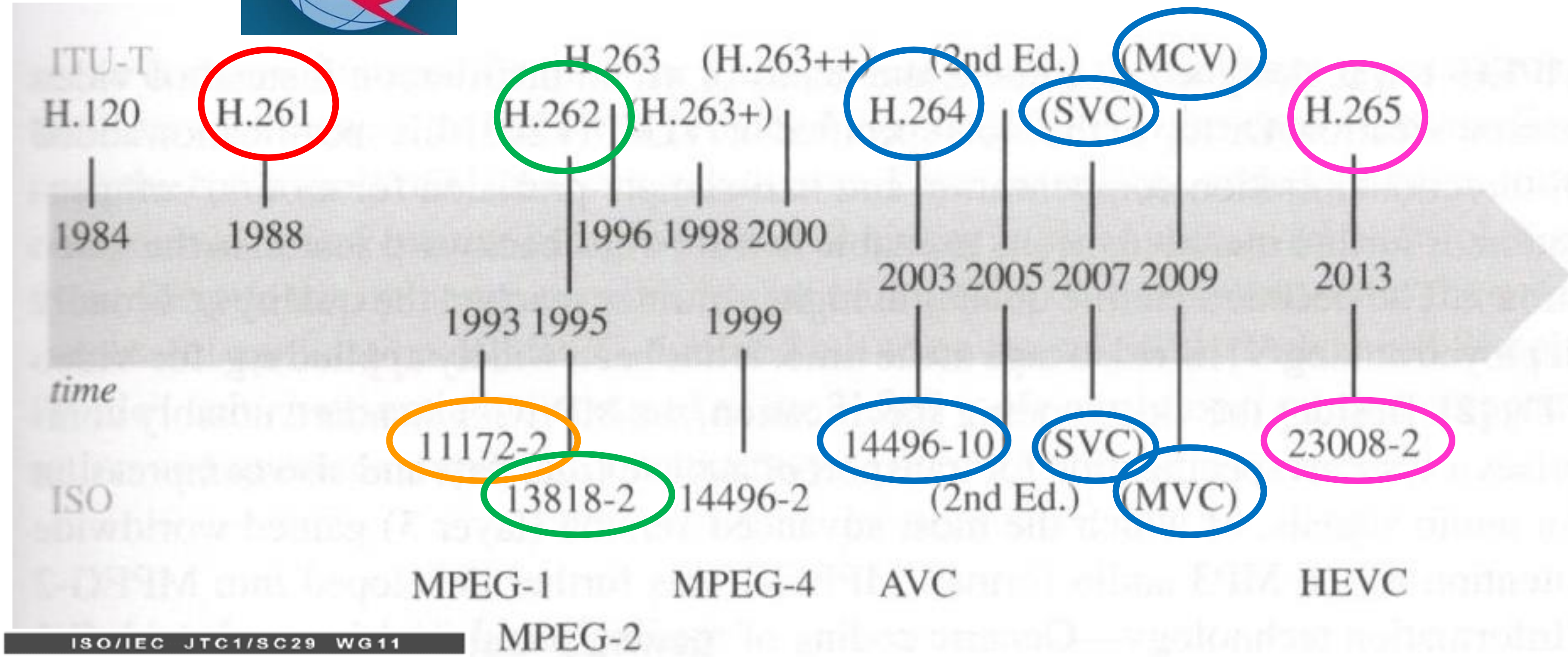


- Higher resolutions are relevant for shorter viewing distances ...
- Shorter viewing distances and large screens increase the sense of immersion ...

Bandwidth and Storage Resources are Always Scarce



Video Coding Standards Over Time …



from M. Wien, "High Efficiency Video Coding", Springer, 2015

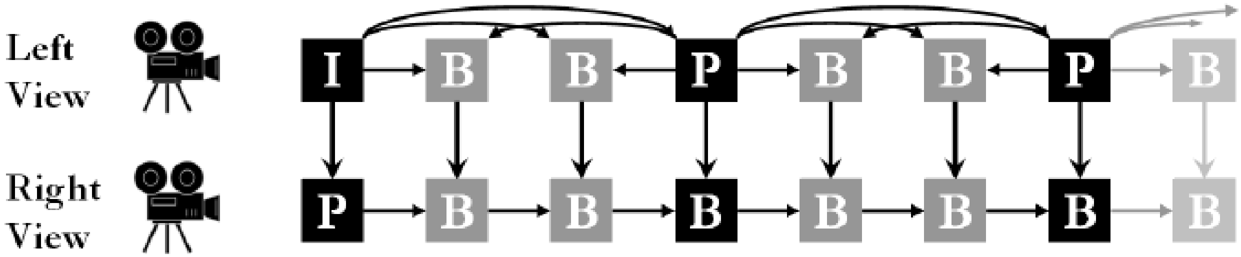

It’s a 3D World !

The 6-DoF Dream for the Future

3D Video: Stereo at the Beginning ...


EXIT

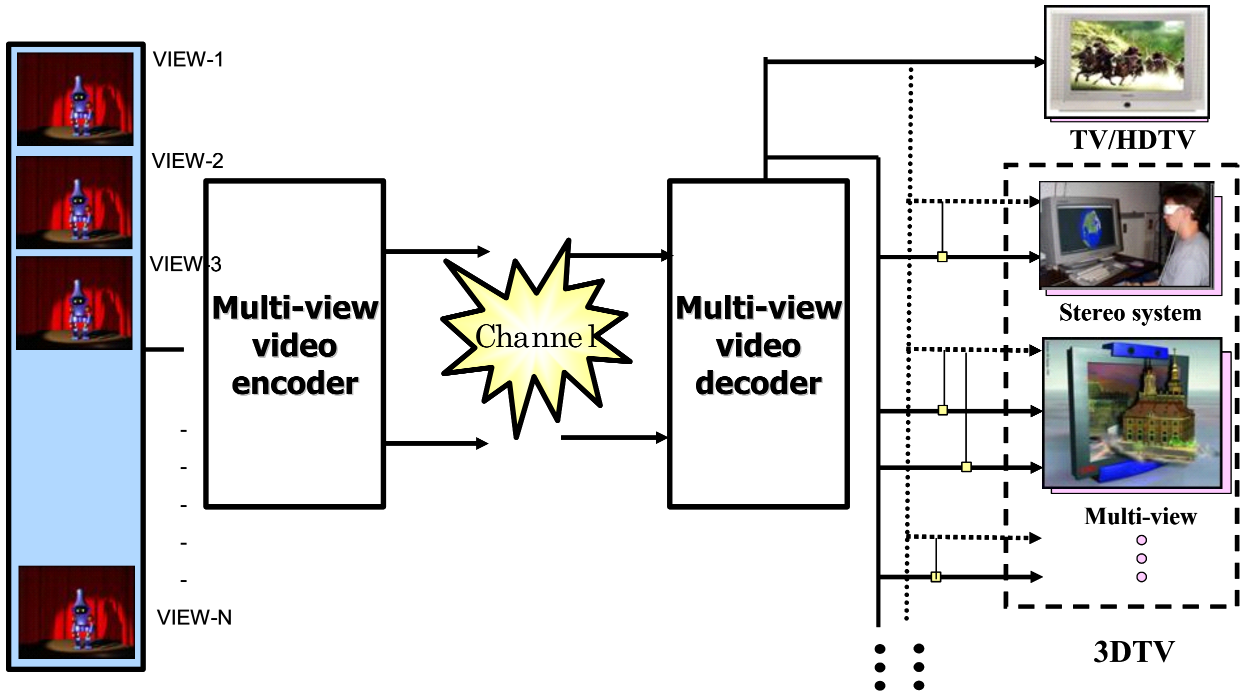
Conventional Stereo Coding



Combined temporal and interview prediction

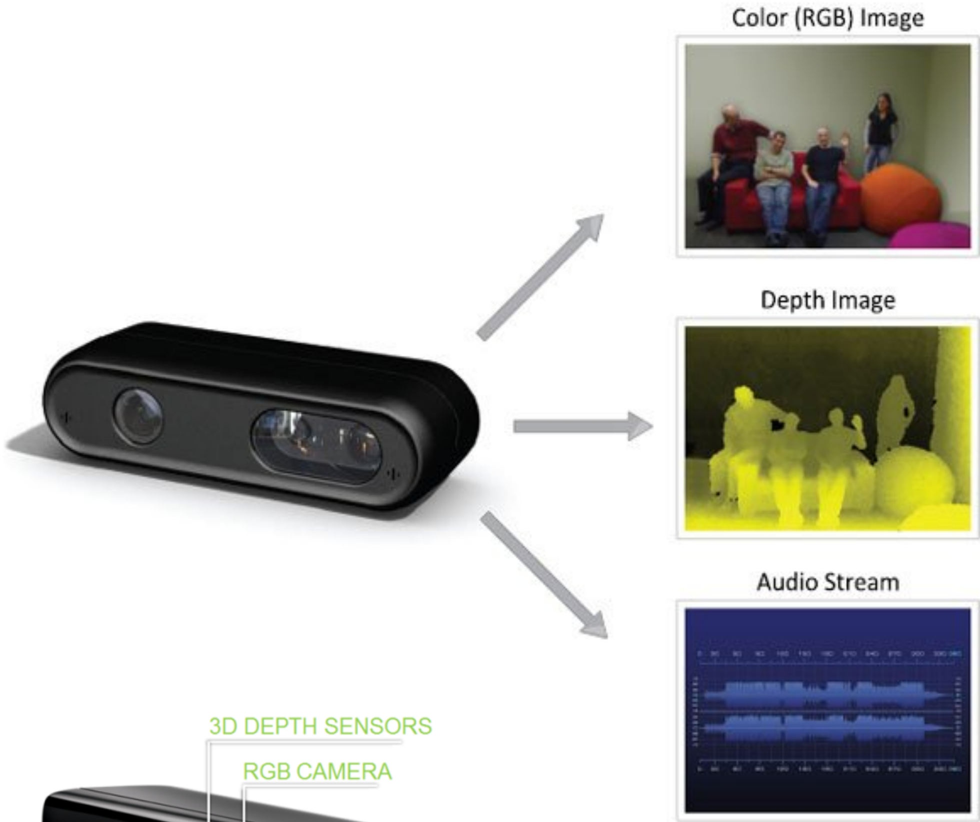
- Conventional stereo coding refers to the case where two full resolution stereo views are coded exploiting their interview redundancy.
- MPEG-2 Video, MPEG-4 Visual and the MVC standards offer full stereo coding solutions with increased compression efficiency.

Multi-View Video Coding



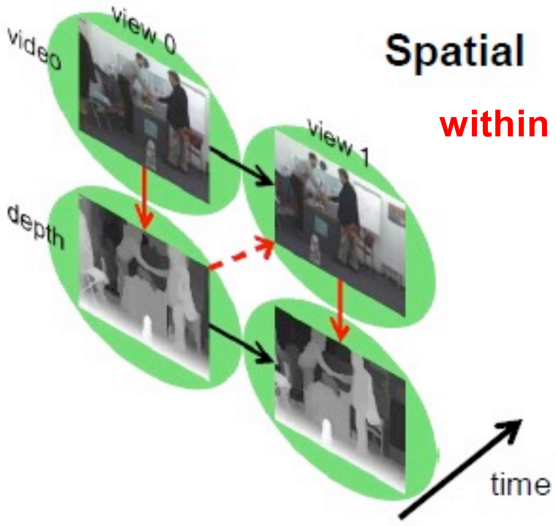
Multi-view video (MVV) refers to a set of *N* temporally synchronized video streams coming from *N* cameras capturing the same real scenery from different viewpoints.

Sensing More with Depth …



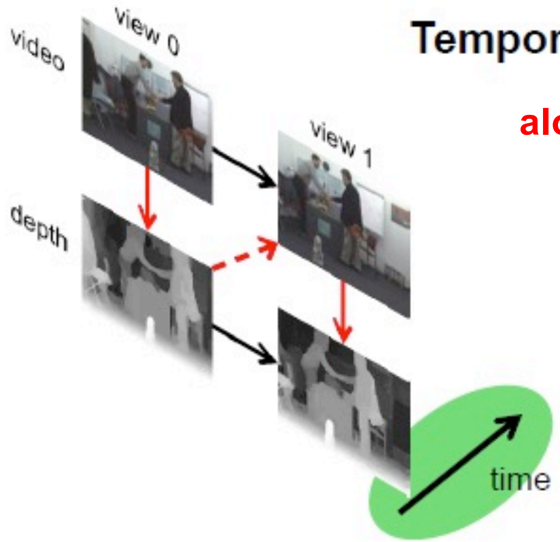
- **A depth map is a 'gray image' containing information about the distance from the scene objects to the camera.**
- **Depth maps may be obtained by:**
  - Special range cameras
  - Extraction from texture
  - Inherent to the content, e.g. computer-generated imagery
- ***Depth maps provide important information about the scene geometry.***

Redundancies in 3D Video ...



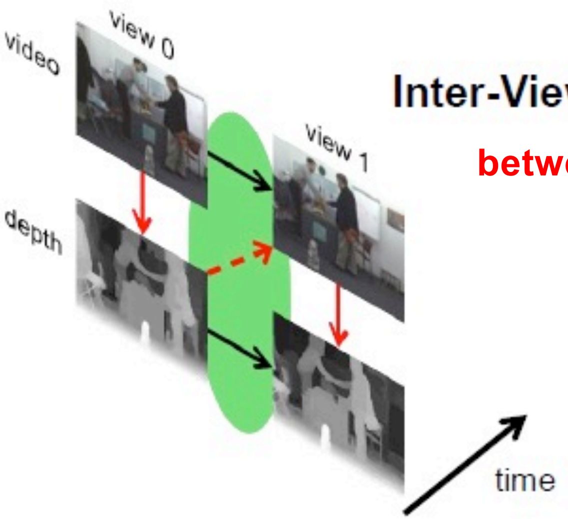
Spatial

within frames



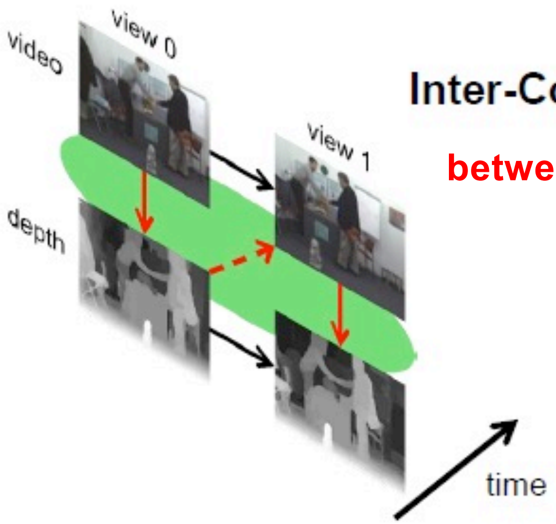
Temporal

along frames



Inter-View

between views



Inter-Component

between components

Multiview Video Coding: the Menu …

Texture only based

- ✶ Multiview Simulcasting
- ✶ Frame Compatible Stereo
- ✶ Conventional Stereo Video
- ✶ Multiview Video, MVC and MV-HEVC standards

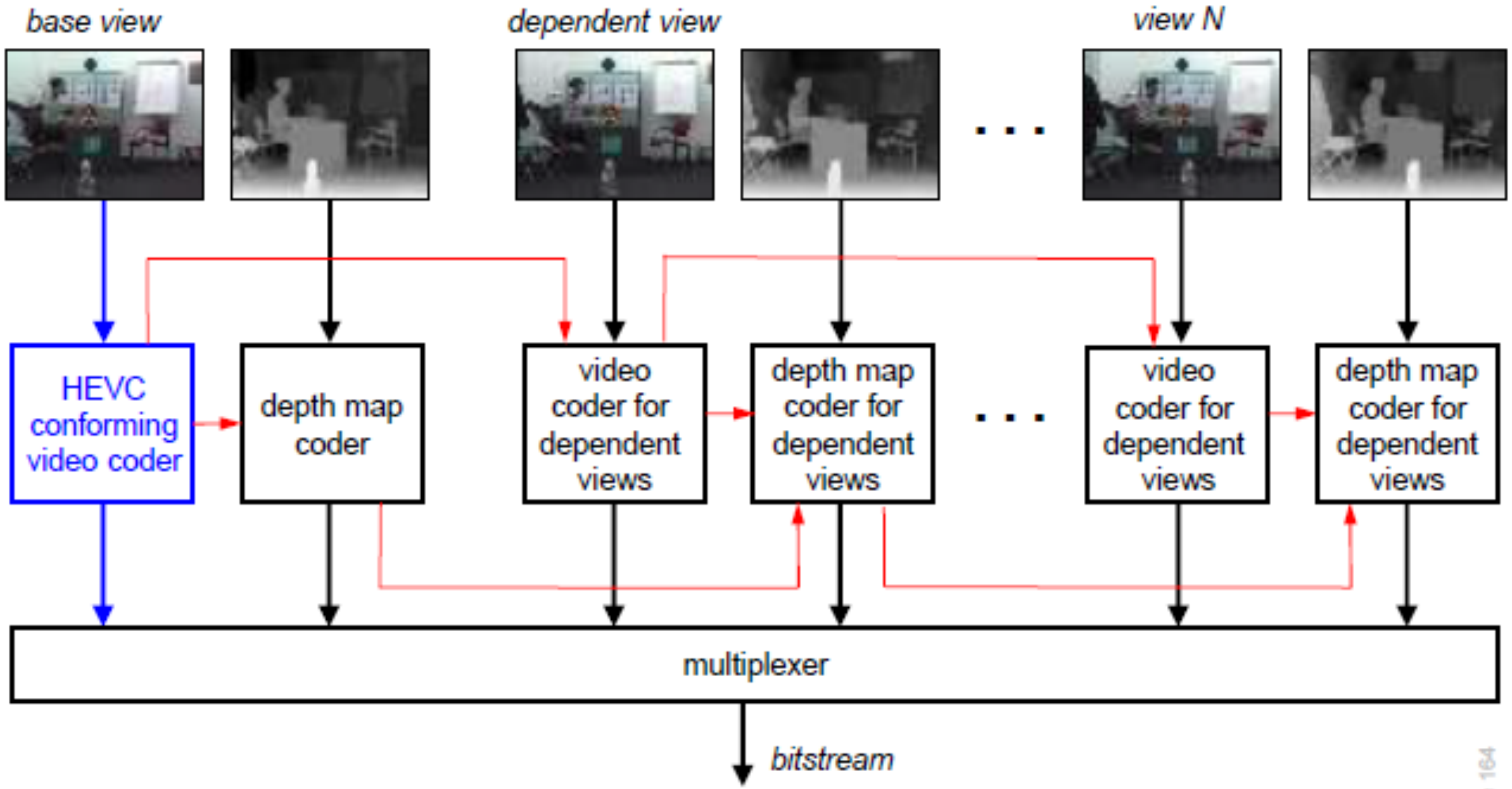
Texture plus Depth based

- ✶ 2D (Texture)+Depth, MPEG-C standard
- ✶ Multiview+Depth (MVD), 3D-HEVC standard

Multiview Video plus Depth (MVD)

- The MVD format encodes both the texture and the depth data for the same number of views.
- MVD is the reference format for some MPEG 3D Video formats where the texture and depth views are independently encoded with MVC.

3D-HEVC: the Most Powerful Available Coding Standard



from K. Muller slides, Fraunhofer HHI

3D-HEVC Limitations

The current visual representation standards only provide efficient multiview video coding solutions for

- Linear and horizontal-only parallax camera arrangements
- Reduced viewing ranges

Moreover

- 3D-HEVC reference software considers a limited number of horizontal-only parallax views (up to 64)
- Some evidence indicates that 3D-HEVC does not provide efficient enough performance for a scenario with many, high density views

Why Didn't Stereo Video Fly ?

It simply did not deliver the Quality of Experience that users expected …

Immersion is poor … Only stereo parallax … No freedom to move/rotate … Far from real world experiences …

TÉCNICO LISBOA

Let’s move forward

Wait ! Glasses, Again ?



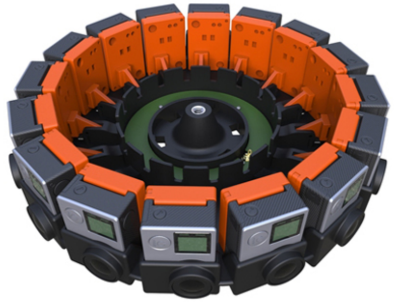




oculus

TÉCNICO LISBOA

360º Video Cameras … For All Tastes …

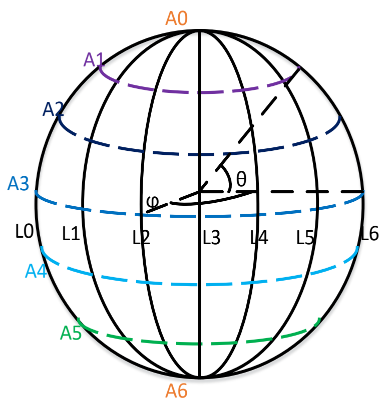


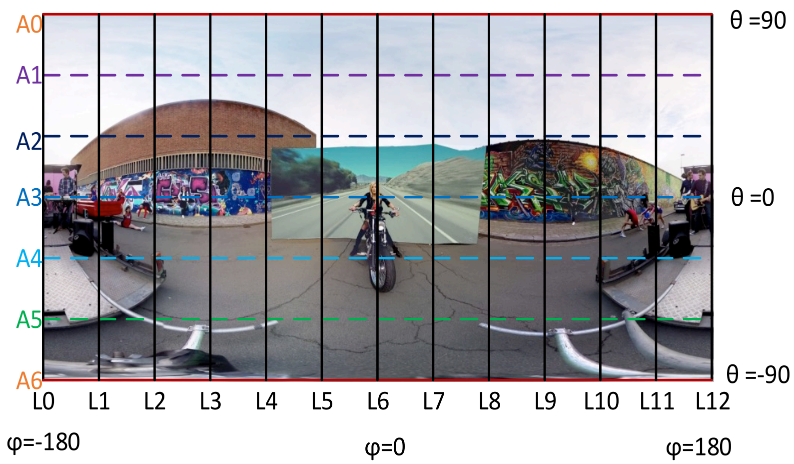
360º or Omnidirectional or Spherical Video

- 360-degree videos, also known as immersive videos or spherical videos, are video recordings where a view in every direction is recorded at the same time, acquired using an omnidirectional camera or a collection of cameras (likely with stitching).

- Viewers can pan and rotate a 360º video's perspective to watch it from different angles.

- Spherical media enables a range of immersive viewing experiences and is currently an essential VR building block.



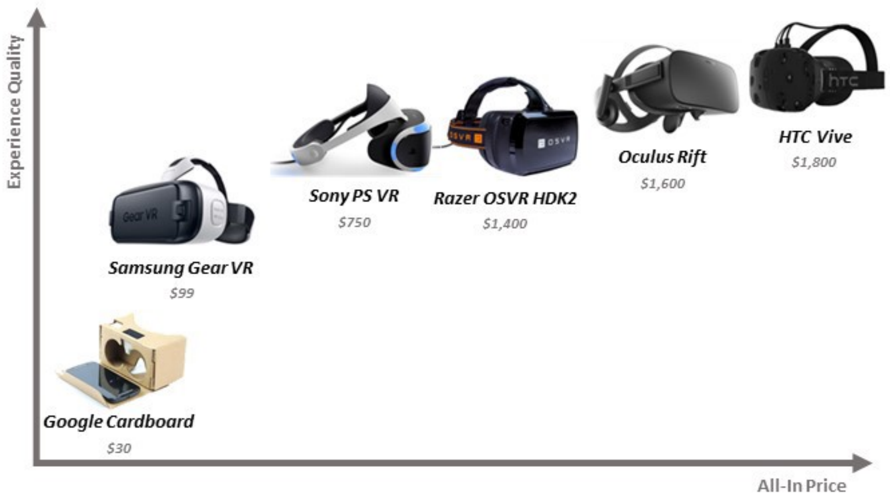


Coding with Tiling …

- Coding of the rectangular projection may be performed with any of the usual image and video codecs, e.g. JPEG, H.264/AVC, HEVC.
- Tiling is particularly relevant when streaming to avoid having to send the full projection map, including the parts that are not being seen at all.

0-DoF & 3-DoF Experiences

3-DoF Quality of Experience



- ✶ **Low spatial resolution**
- ✶ **Coding artifacts**
- ✶ **Motion-to-photon delay (time needed for a user movement to be fully reflected on a display screen)**
- ✶ **Motion sickness**
- ✶ **<u>Limited degrees of freedom</u>**

Quality of Experience ...

- Current 3-DoF experiences have major cyber sickness issues, especially when moving cameras are used.

- 6-DoF visual content (and not only 3-DoF) seems to be a critical step to achieve cyber sickness-free visual experiences.

- However, the 3-DoF to 6-DoF jump will significantly increase the amount of data.

TÉCNICO LISBOA

Let's move forward

again!

TÉCNICO LISBOA

2

Plenoptic Function based Imaging

TÉCNICO LISBOA

2.1 Plenoptic Function based Imaging: The Basics

Genius, Simply Genius ...

Leonardo di ser Piero da Vinci (1452 – 1519)
his areas of interest included invention, painting, sculpting, architecture, science, music, mathematics, engineering, literature, anatomy, geology, astronomy, botany, writing, history, and cartography !!!!!

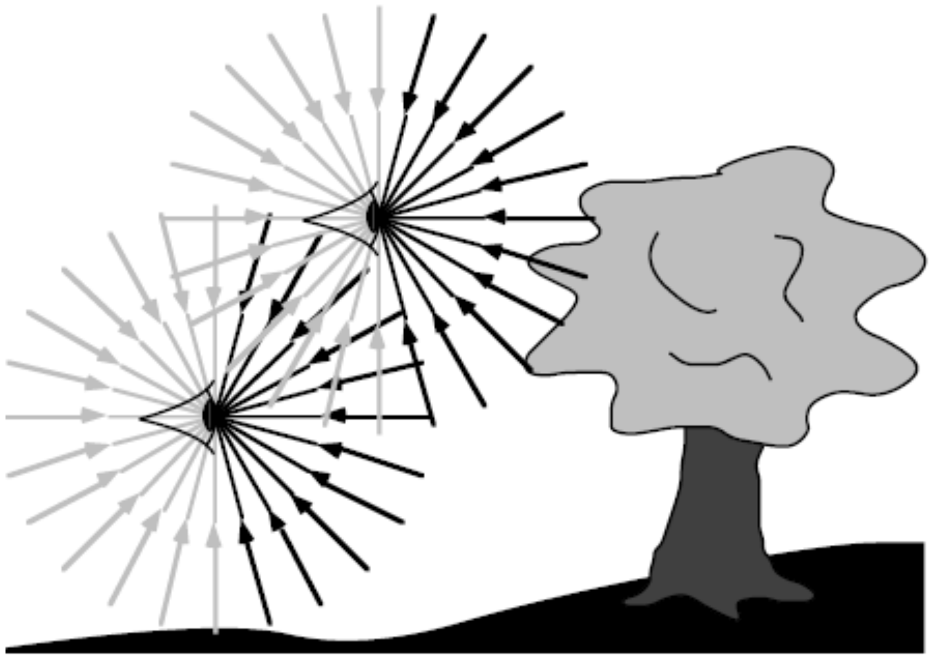
*"**The air is full of an infinite number of radiant pyramids caused by the objects located in it.** These pyramids intersect and interweave without interfering with each other….*

***The semblance of a body is carried by them as a whole into all parts of the air**, and each smallest part receives into itself the image that has been caused."*

Leonardo and I. Richte, "The Notebooks of Leonardo Da Vinci", Oxford University Press, 1980

The World and its Observers

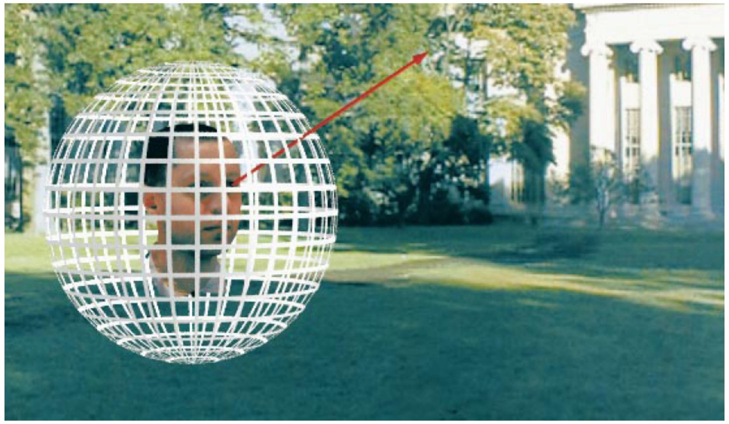
- ✶ The world is made of 3D objects …
- ✶ These objects do not communicate their properties directly to an observer ...
- ✶ The objects fill the space around them with a pattern of light rays that constitutes the so-called <u>plenoptic function</u> …
- ✶ The observer’s eyes take samples from this function …

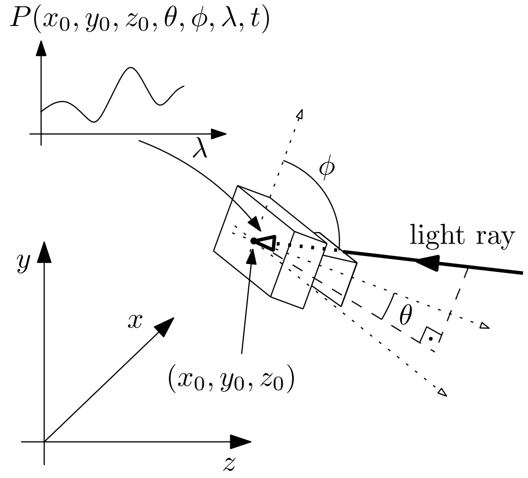


from Adelson and Bergen, “The Plenoptic Function and the Elements of Early Vision”, 1991

TÉCNICO LISBOA

The 7D Plenoptic Function …





- **The Plenoptic Function serves as the sole communication link between the world and the eye/vision/observer.**
- **The Plenoptic Function, *P(x, y, z, θ, φ, t, λ)*, measures the intensity of light seen from**
  - any viewpoint, camera centre 3D spatial position *(x,y,z)*
  - any angular viewing direction (*θ,ϕ*)
  - over time *(t)*
  - for each wavelength (*λ*)
- **The Plenoptic Function can represent every possible view, from every position, at every moment, and at every wavelength.**

Restricting the Plenoptic Function …

- **It is possible to reduce the high dimensionality of the Plenoptic function by adopting various restrictions:**
  - **BW static images -** *fixed time, 1 wavelength component, no angular data*
  - **Color static images** *– fixed time, 3 wavelength components, no angular data*
  - **Color mono video** *– variable time, 3 wavelength components, no angular data*
  - **Color stereo video** *– variable time, 2 angular perspectives, 3 wavelength components*
  - **Color array of cameras** *– fixed time or not, 3 wavelength components, angular data for multiple 2D positions*

- **Depending on the application scenario and associated display, it is essential to measure/sample the Plenoptic function using appropriate sensor devices.**

- **Image-based rendering regards the sampling and reconstruction of the Plenoptic function, e.g. creating new views from sampled views.**

The 6-DoF Experience Challenges

1. **IMMERSION - How to offer 6-DoF immersive experiences, i.e. the perfect pixel for any motion ?**

2. **DATA MODEL - What is the best visual data representation model to offer 6-DoF experiences ?**

3. **COMPRESSION - How to design efficient coding algorithms to achieve 6-DoF immersion with a selected representation model ?**

4. **RENDERING - How to design rendering to achieve as transparent as possible 6-DoF immersion?**

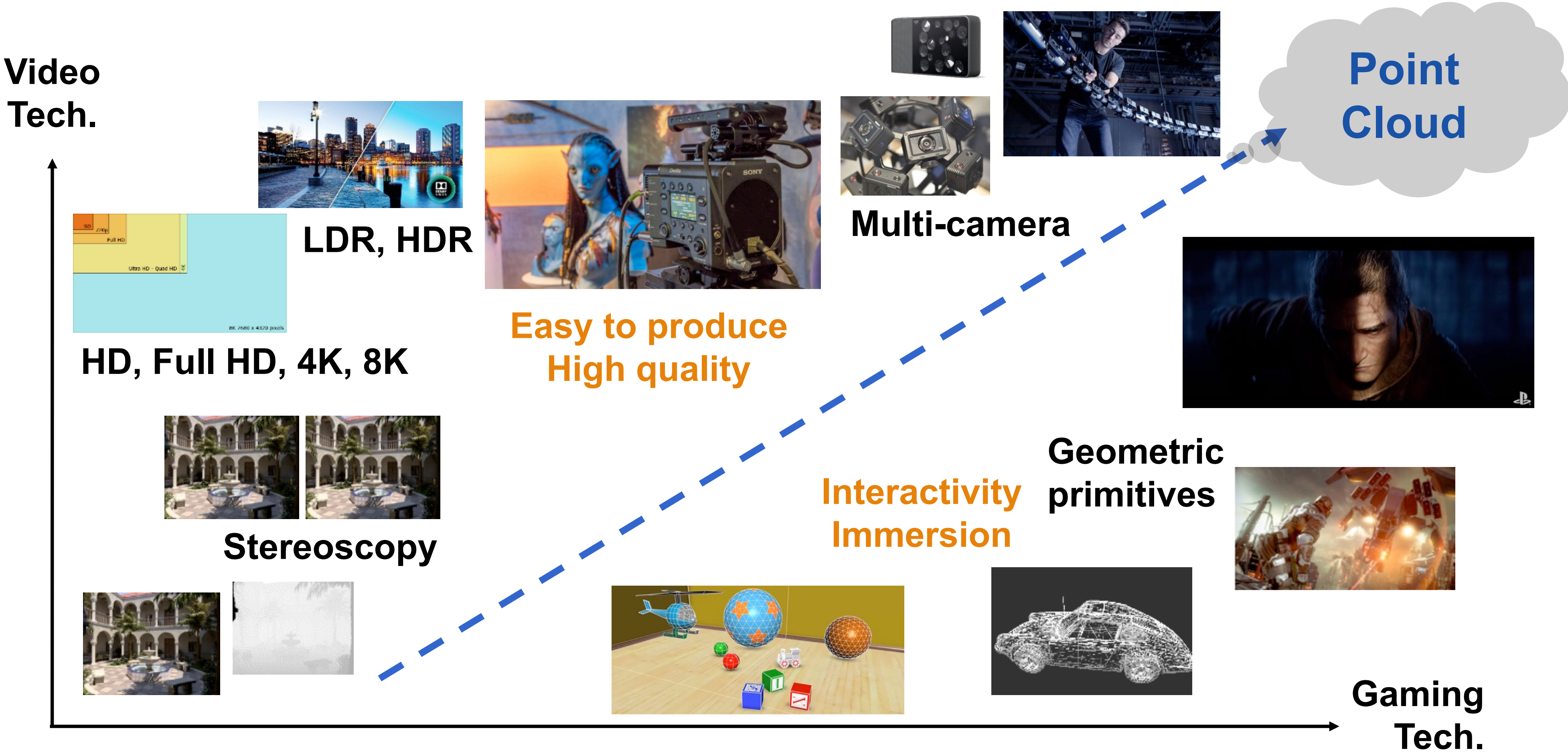
5. **QUALITY OF EXPERIENCE - How to avoid motion sickness, e.g. no shift between what is displayed and the expectation of the brain ?**

inspired from J. Jung

2.2 Plenoptic Function based Imaging:

Light Fields versus Point Clouds

Trends for Immersive Content ...



from Danillo Graziosi

The Acquisition: Sampling the Plenoptic Function …

How do we Sample/Acquire this type of scene ?

Again Sensors in the lead...

Targetting increased Immersion !

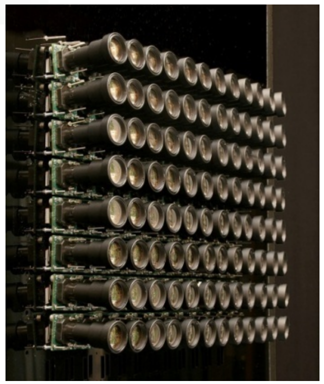
TÉCNICO LISBOA

Sensors … We Need New Sensors …

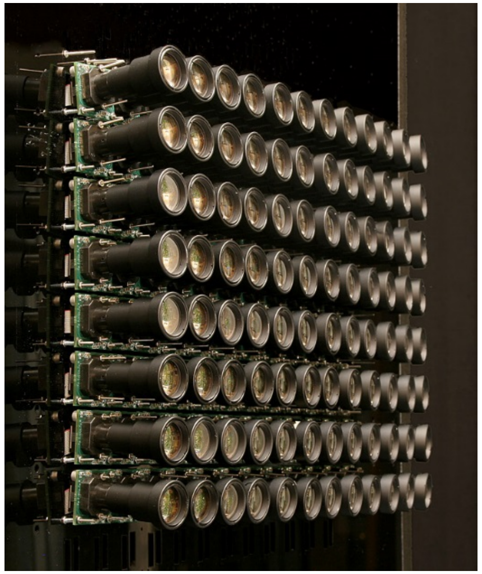
A sensor is a transducer whose purpose is to sense some characteristic of its environment. It detects events or changes in quantities and provides a corresponding output.

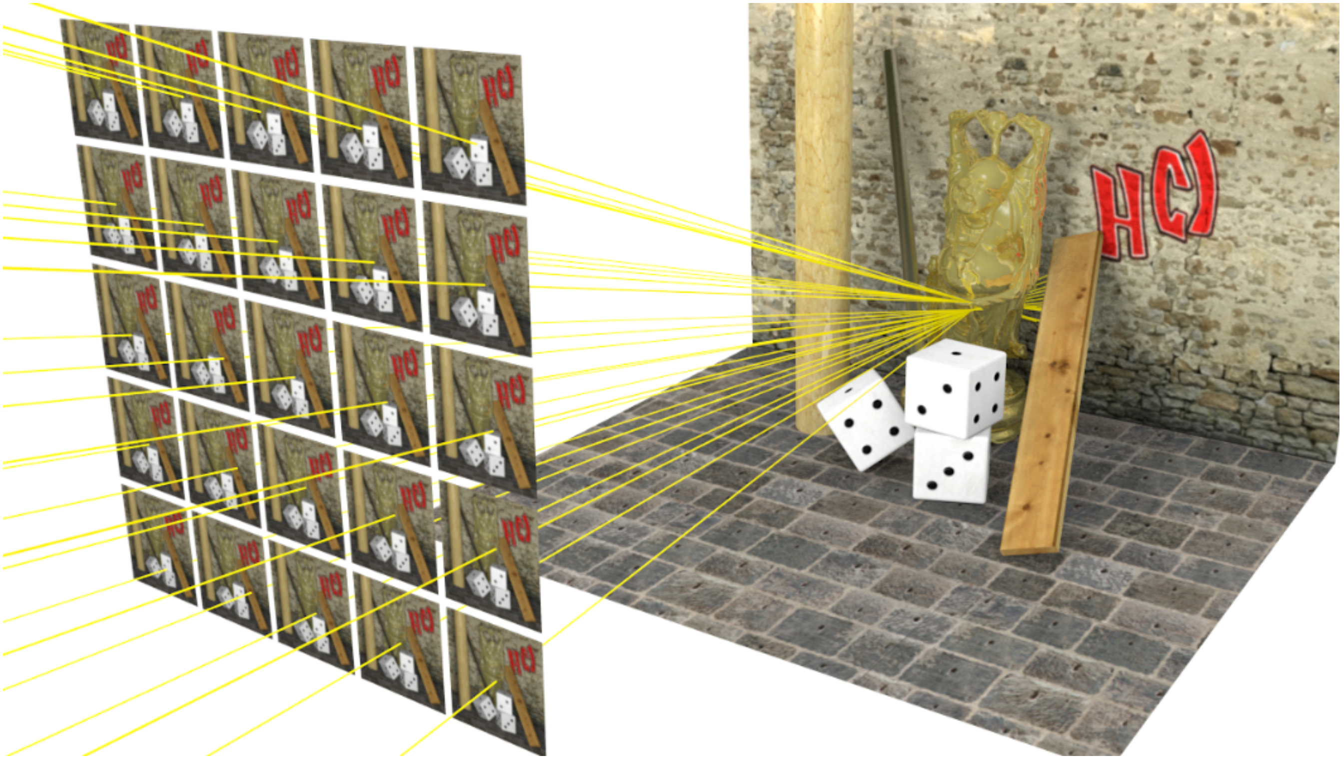
- **Up to Now … mostly video sensors ...**
  - With increased spatial resolution
  - With increased frame rate
  - With increased dynamic range
  - …

- **More recently …**
  - Arrays of cameras
  - Camera as array of lens
  - 360º cameras
  - Depth cameras
  - LIDAR
  - ...

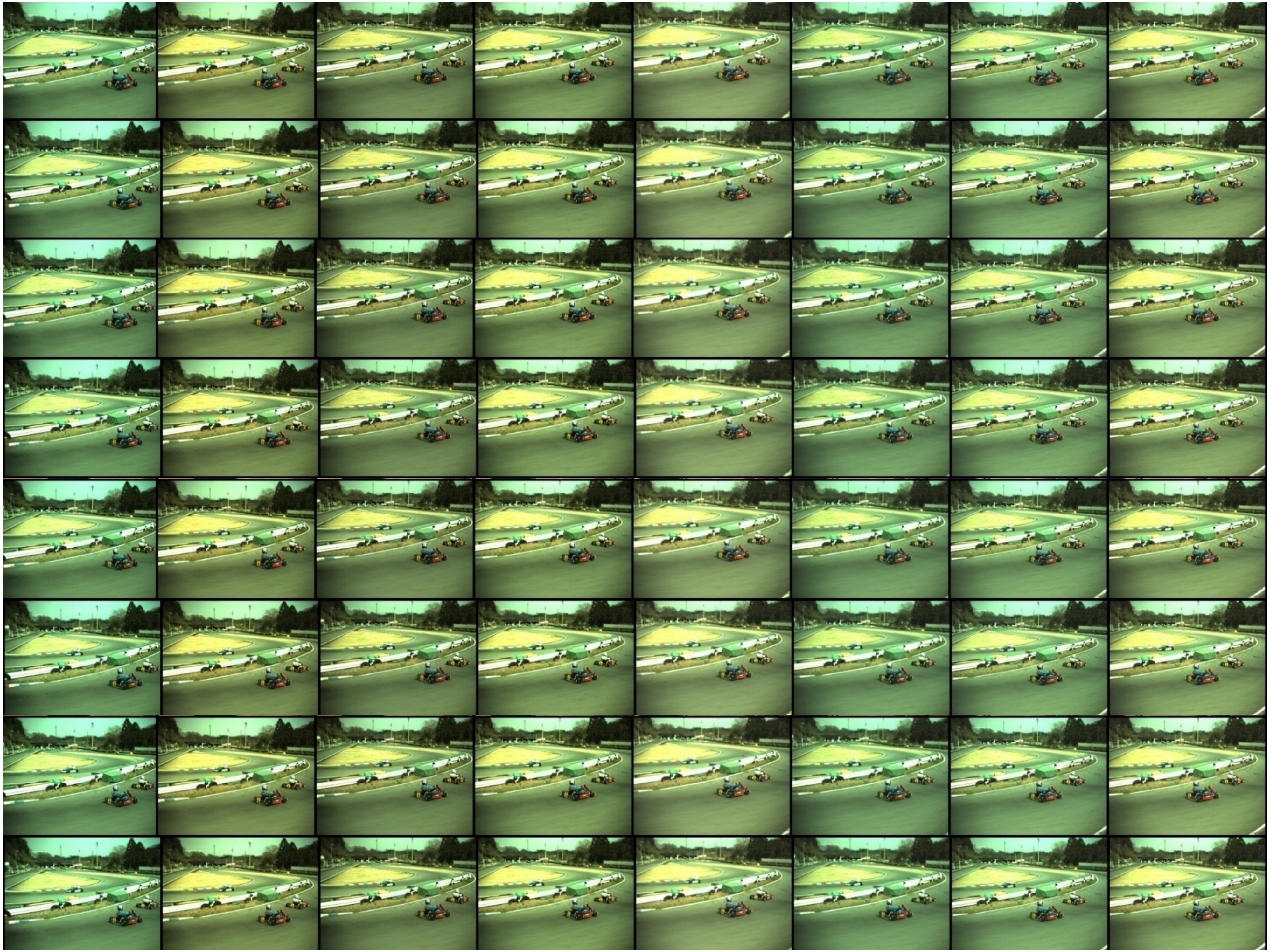


High Density Camera Array Light Field Acquisition



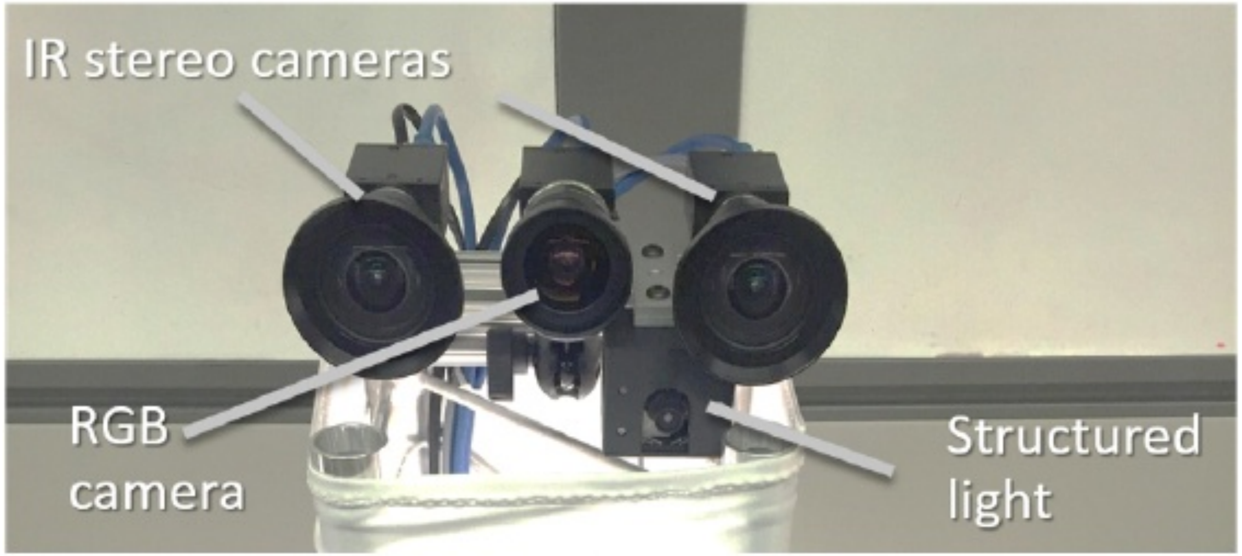


6-DoF Cinematic (very realistic) VR needs light fields !

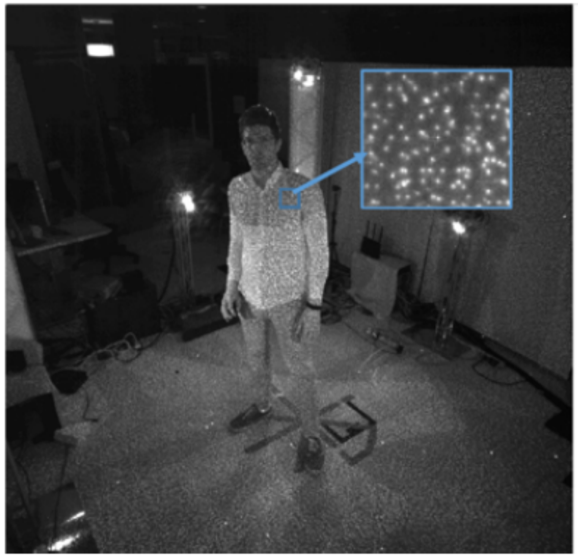


TÉCNICO LISBOA

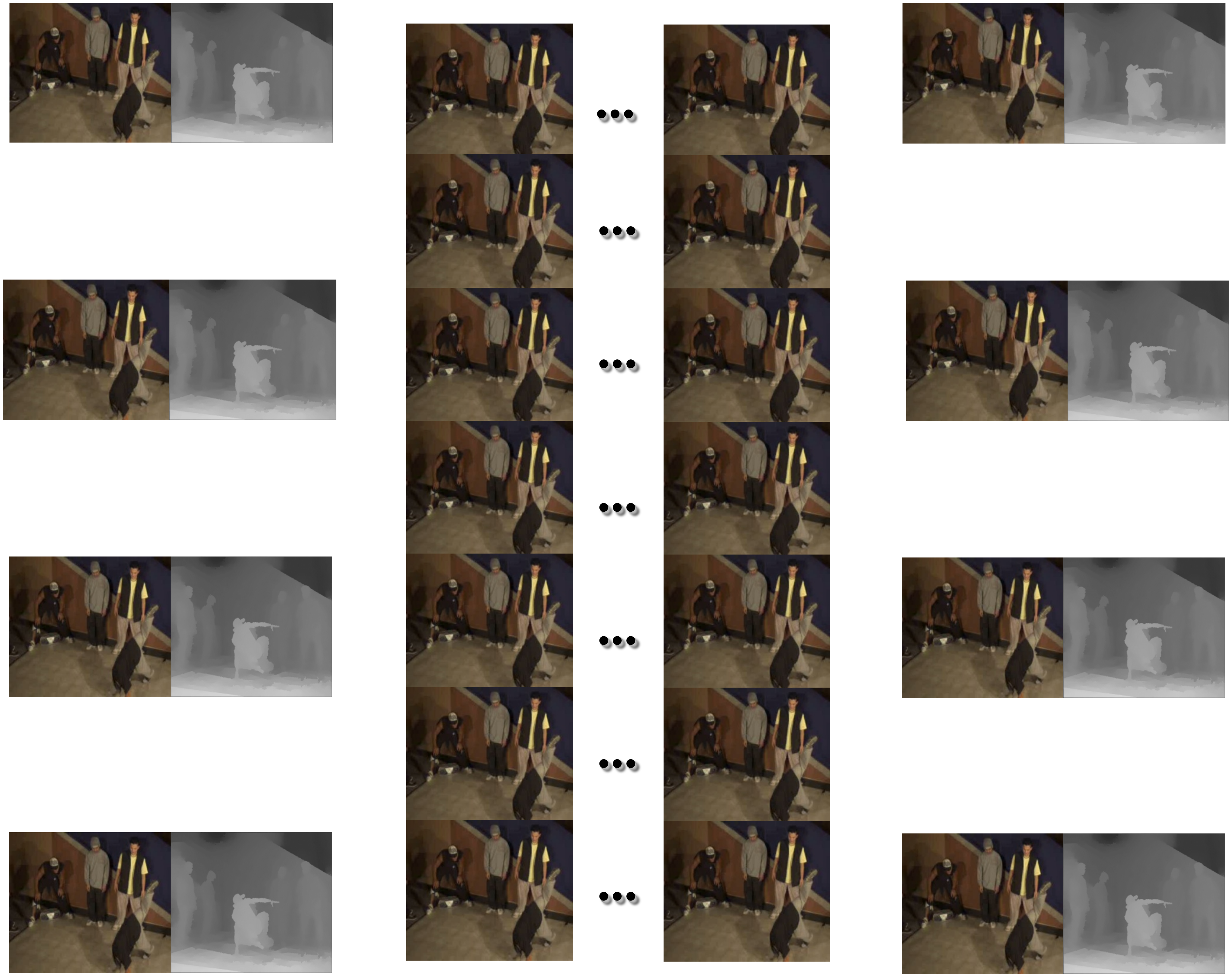
Texture + Depth: a Powerful Cockail …



NIR (Near Infra-Red) image

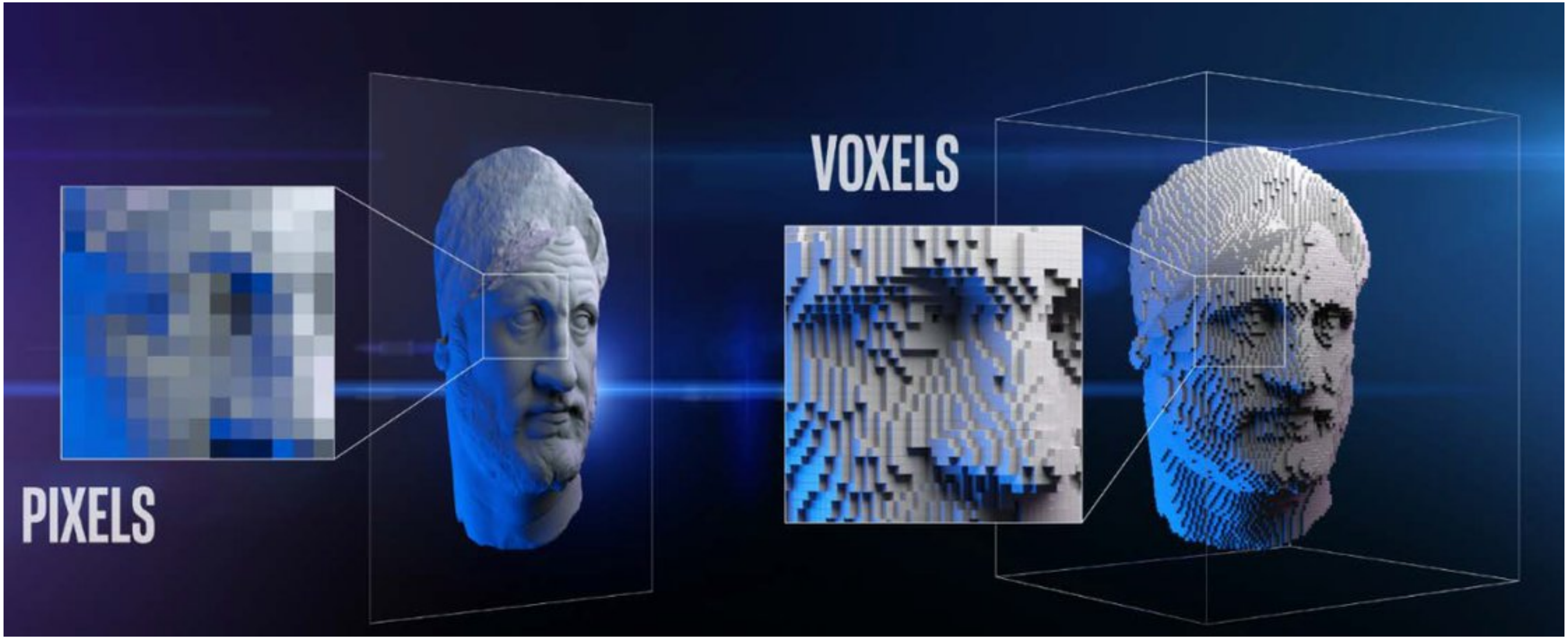


RGB image



Light Field Representation: the Pixel based Approach

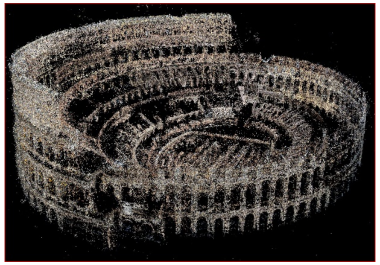
From Pixels to Voxels



Point Cloud Definition

- A point cloud is a set of points in the 3D space with coordinates (x,y,z), located on the surface of objects, the geometry.
- Each point may have attributes associated, e.g. color, normals, reflectance, etc.
- Point clouds are generally produced by 3D scanners, which measure a large number of points on the external surfaces of objects.
- Point clouds may also be generated from texture & depth data … this means from light fields ...
- One voxel (position of a regular 3D grid) may correspond to one or more points depending on the adopted precision.

Point Cloud Acquisition …



The geometry information may be obtained using two types of methods:

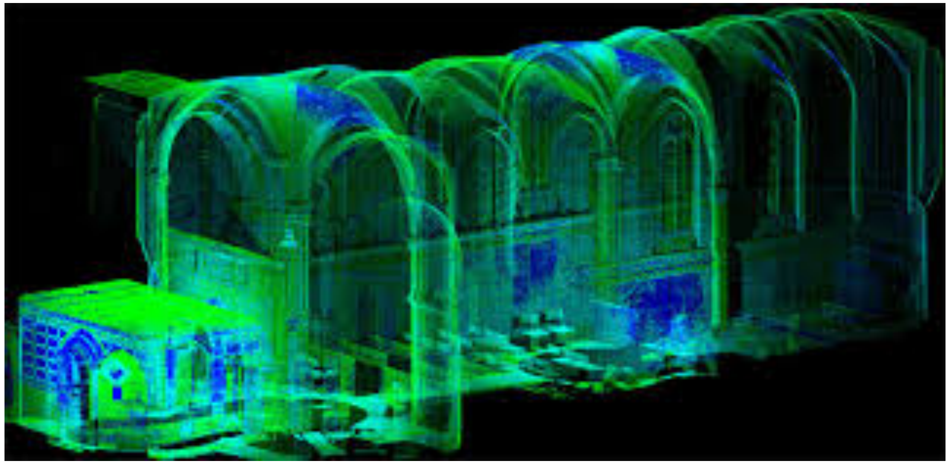
* Passive methods use multiple cameras and perform image matching and spatial triangulation to infer the distance between the captured objects in 3D space and the cameras.

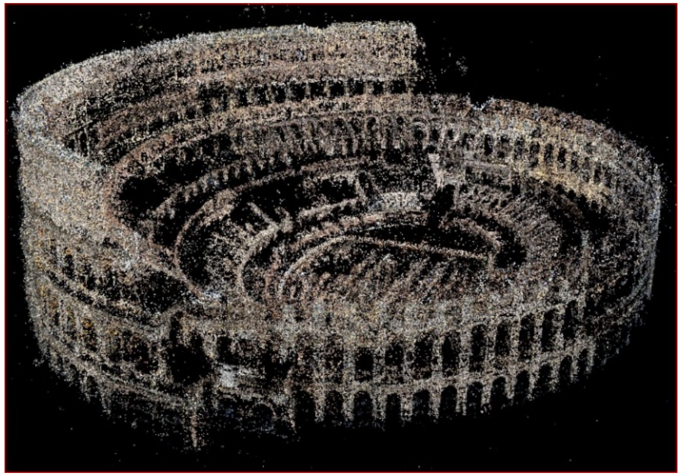
* Active methods use light sources (e.g. infra-red or lasers) and backscattered reflected light to measure the distances between the objects and the sensor.

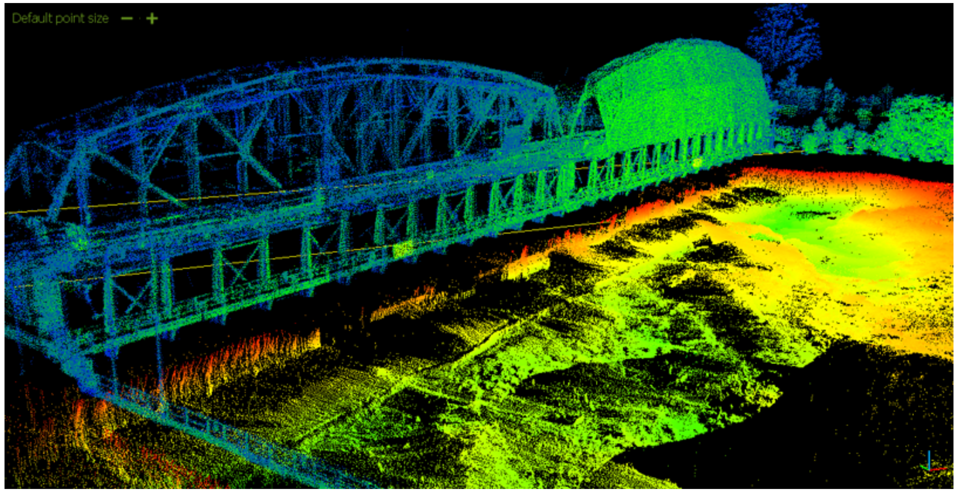
Both active and passive depth acquisition methods can be used in a complementary manner to improve the generation of point clouds.

The latest trend in capture technology is volumetric studios, where either passive methods (using RGB cameras only) or a combination of passive and active methods (using RGB and depth cameras) create a high-quality point cloud.

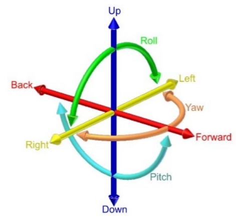
Point Cloud Representation: the Voxel based Approach







Light Fields versus Point Clouds



Light Fields and Point Clouds are associated to specific different representation models and coding algorithms.

It is possible to convert light fields into point clouds and vice-versa …

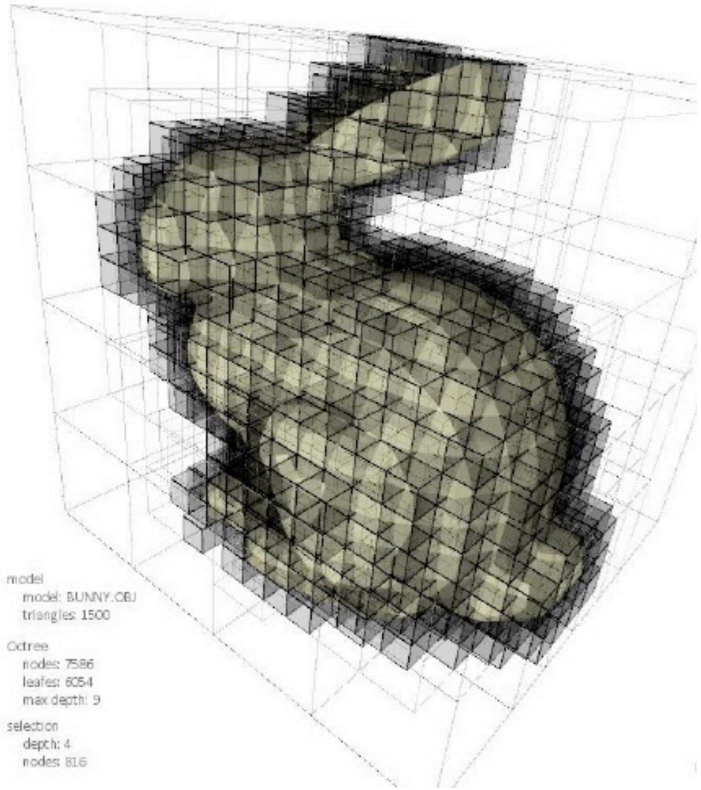
Light Fields

- 6-DoF
- Image based representation
- Tens or hundreds of cameras
- Does not distinguish the objects in the scene
- Not so easy interaction with objects in the scene
- Easy reuse of advanced video coding technology

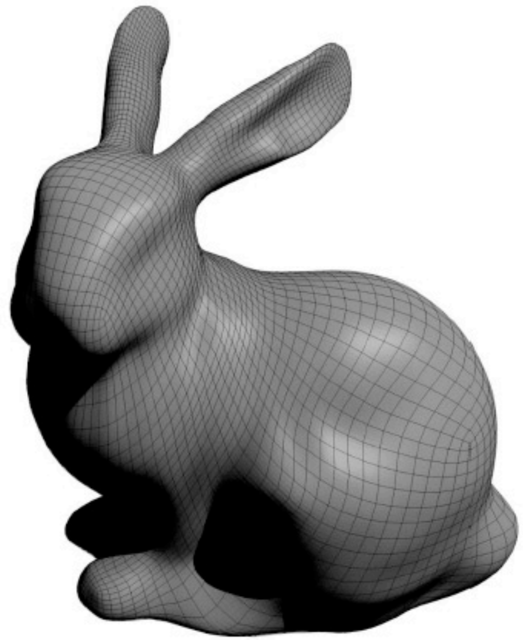
Point Clouds

- 6-DoF
- Geometry based representation
- Millions of 3D points
- Easy separation of the objects in the scene
- Easier interaction and manipulation with objects
- Less obvious reuse of advanced video coding technology, although possible (see later)

From Point Clouds to Meshes ...



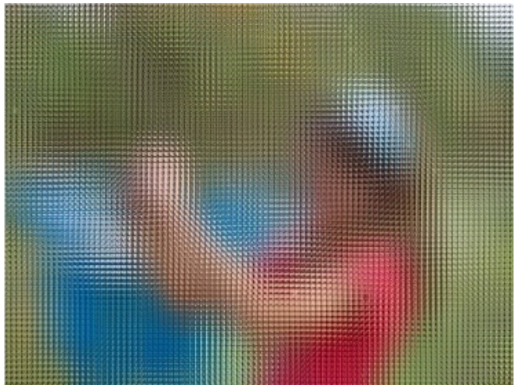
Irregular sampling !

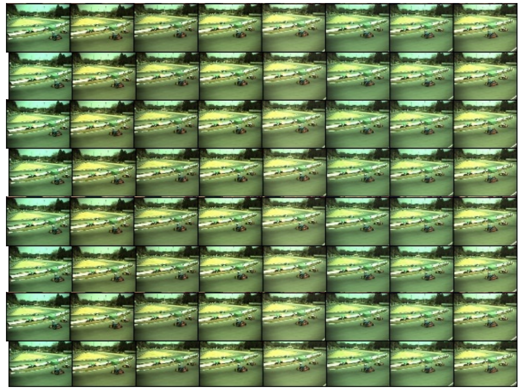


Meshes are similar to Point clouds but with connectivity information (points and edges). This makes them slightly more complex to process and represent than Point clouds.

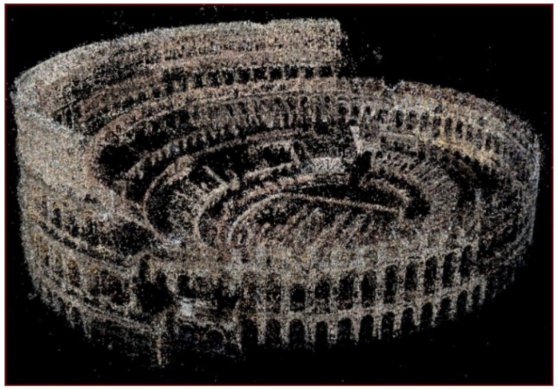
Courtesy of P.Chou, Microsoft

Plenoptic Representation Models: the Main Candidates

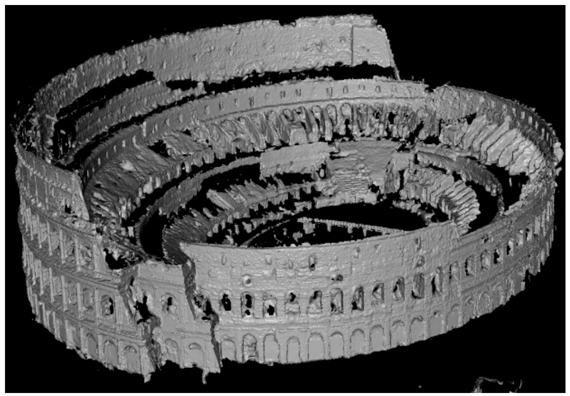




- Light Fields
- Light Fields with Depth
- Point Clouds (with attributes)
- Meshes (with attributes/texture)



These are rather complementary representation models, and each of them may be better suited for specific application domains.



TÉCNICO LISBOA

Light Fields and Point Clouds Data: It is Really Too Much !

TÉCNICO
LISBOA

TÉCNICO
LISBOA

Tutorial Outline

1. 3D Visual Representation and Coding (F.Pereira)
2. Plenoptic Function based Imaging (F.Pereira)

Questions

3. Point Cloud Coding: Basic Approaches (F.Pereira)

Questions

4. Point Cloud Coding: Standardization, part 1 (J.Ascenso)

Questions

5. Point Cloud Coding: Standardization, part 2 (J.Ascenso)
6. Point Cloud Quality Assessment (J.Ascenso)
7. Summary and Trends (J.Ascenso)

Questions

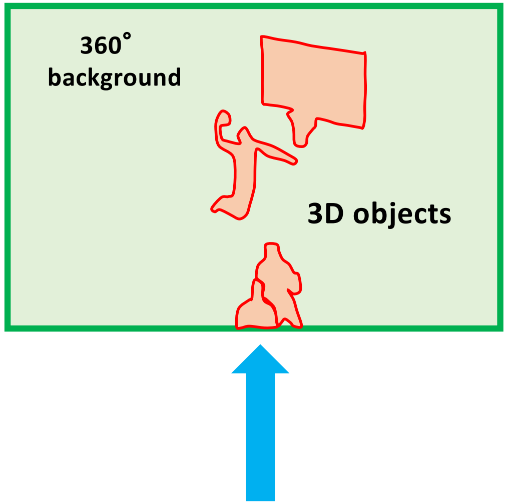
TÉCNICO LISBOA

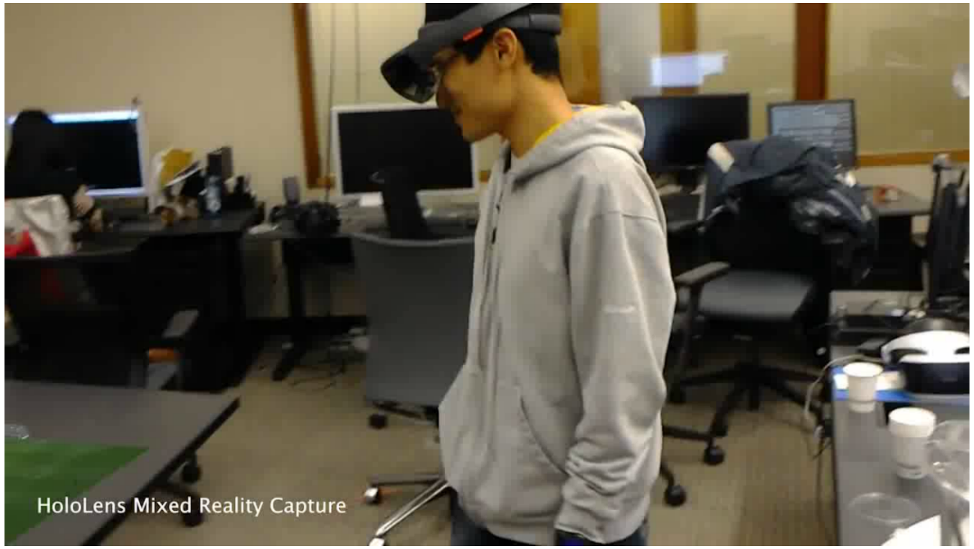
3 Point Cloud Coding: Basic Approaches

Sport Viewing with Point Clouds

https://www.youtube.com/watch?v=Q-LNA9KIHhw

https://www.youtube.com/watch?v=eRGAB4QBS6U



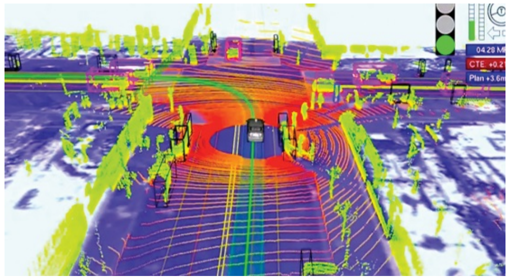


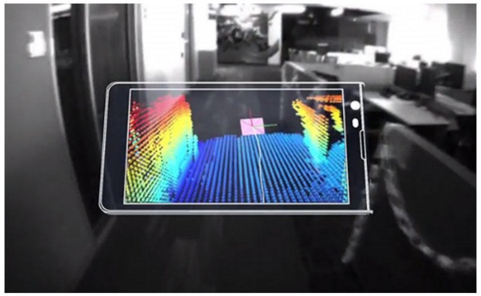
3 Gbps per object

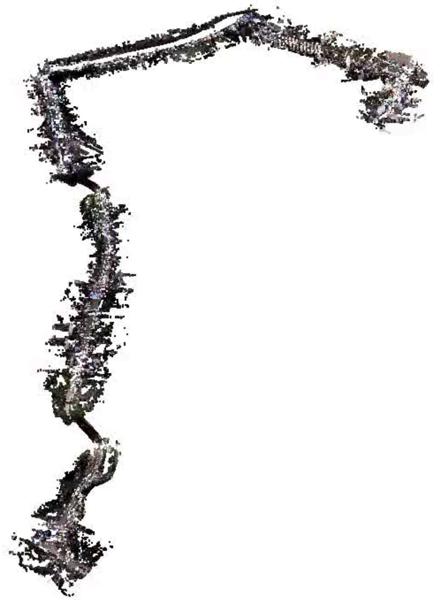
Courtesy of Danillo Graziosi

TÉCNICO LISBOA

Environment Mapping for Autonomous Driving







~20 million points - 2,020,734,515 bytes

Courtesy of Danillo Graziosi

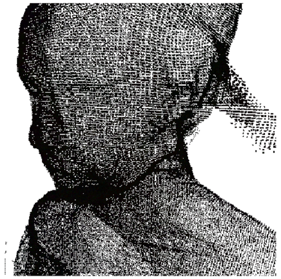
Point Cloud Coding is a Must ...

2D media content

HD @ 30fps → 1.5 Gbps

(fixed viewpoint)

3D media content

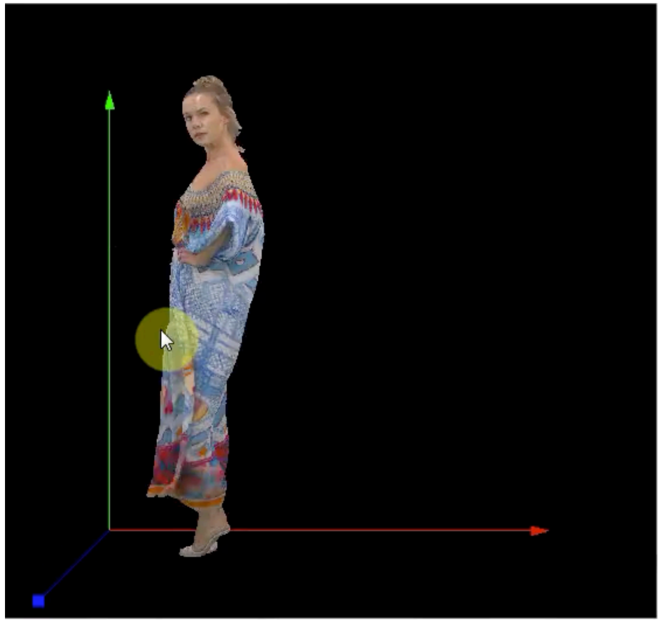


3D point cloud

⇩

A set of **unordered** 3D points:

- (X, Y, Z)
- (R, G, B) or (Y, U, V)
- reflectance, transparency, …



800,000 points @ 30fps → 2.88 Gbps

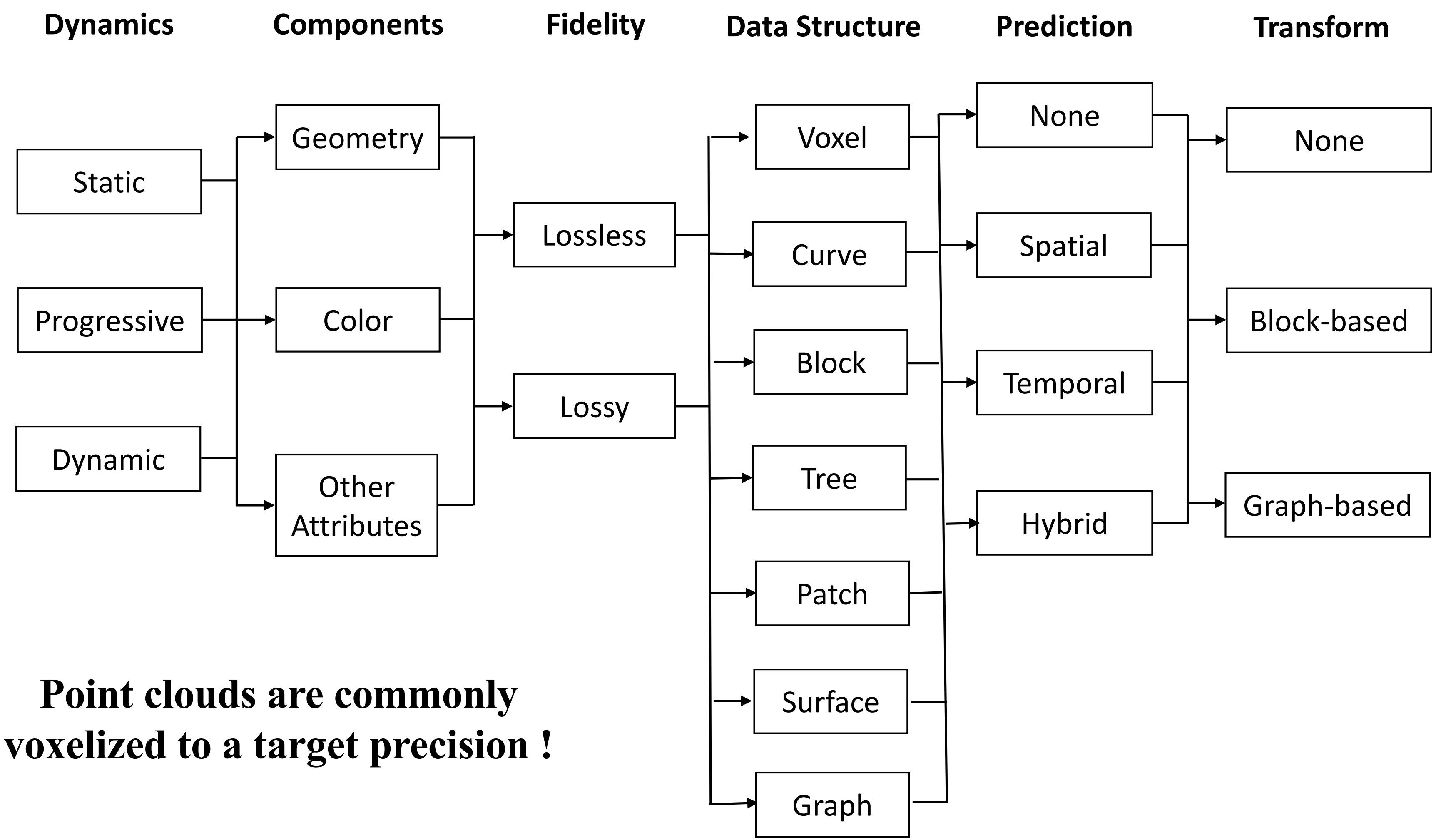
Compression is required in order to make PC useful

Courtesy of Danillo Graziosi

TÉCNICO LISBOA

3.1 Point Cloud Coding: Taxonomy

PC Coding Classification Taxonomy



Point clouds are commonly voxelized to a target precision !

from F. Pereira, et al., "Point cloud coding: A privileged view driven by a classification taxonomy", Signal Processing: Image Communication, 2020

Taxonomy Classification Dimensions

- **Dynamics** – Refers to the type of dynamics the data involves, notably in terms of temporal evolution.
- **Components** – Refers to the specific type of data being coded, e.g. the 3D point positions or attributes.
- **Fidelity** – Refers to the fidelity with which the data is coded.
- **Data Structure** – Refers to the way the raw data is structured to be then coded while exploiting the available spatial and temporal redundancies; this may involve segmentation and partitioning of the data.
- **Prediction** – Refers to the way the spatial and temporal correlations in the structured data are exploited to create a lower energy signal, the so-called residue.
- **Transform** – Refers to the way the spatial correlation in the original signal or the remaining correlation in the residue signal are exploited to reach a more compact energy representation, usually in some type of frequency domain.

Dynamics Dimension

- **Static** – PC corresponding to a single time instant.
- **Progressive** – Large scale PC that is typically not consumed all at once, thus corresponding to complementary parts of a scene, which are coded as refinements or extensions, and involving changes/improvements in a spatial navigation.
- **Dynamic** – PC evolving along time, thus corresponding to a sequence of PC frames, each corresponding to a static PC, therefore involving changes associated to motion and deformations.



Components Dimension

- **Geometry –** Set of 3D coordinates (x,y,z) representing the positions of the points in the 3D space.

- **Color –** Set of color values associated to each PC point in some color space, e.g. RGB or YUV; each point may have a single color or a set of colors associated with a predefined set of directions.

- **Other Attributes –** Set of local features (other than color), associated to each PC point that may be used to further describe the PC. The main PC attributes include the following:
  - ***Normals*** – Set of normal vectors associated to each PC point; the normal at a certain 3D point is the vector perpendicular to a local plane representing the neighborhood of that point.
  - ***Reflectance*** – Set of reflectance values associated to the PC points; the reflectance can be defined as the effectiveness of a material in reflecting radiant energy and is expressed as the fraction of incident radiant flux that is reflected by that material.
  - ***Additional attributes*** – Any other PC attributes that may be relevant for the target application, e.g. distances of the points to the acquisition device

Fidelity Dimension

- **Lossless -** Codecs keeping the original fidelity, meaning that the decoded and original data are mathematically equal up to a certain precision; it is possible that a lossless codec also offers lossy decoding in a scalable manner.
- **Lossy** – Codecs that do not keep the original fidelity, typically to increase the compression factor; high fidelity, notably perceptually lossless quality, may still be achieved with the appropriate coding parameters configuration.

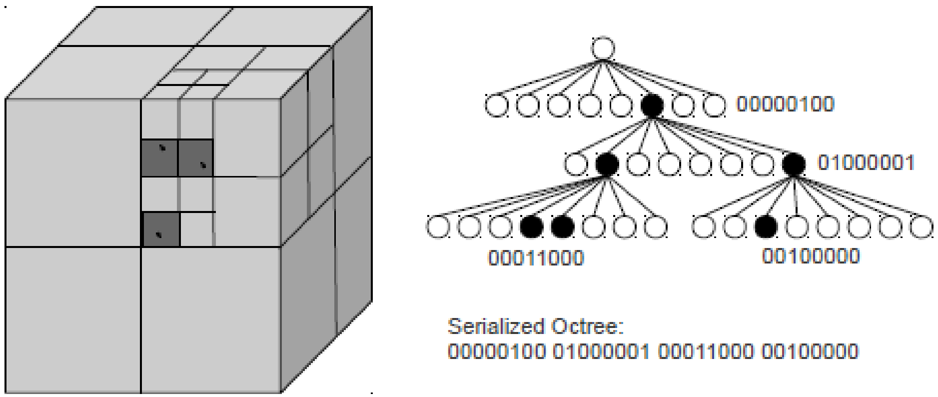


Courtesy of Danillo Graziosi

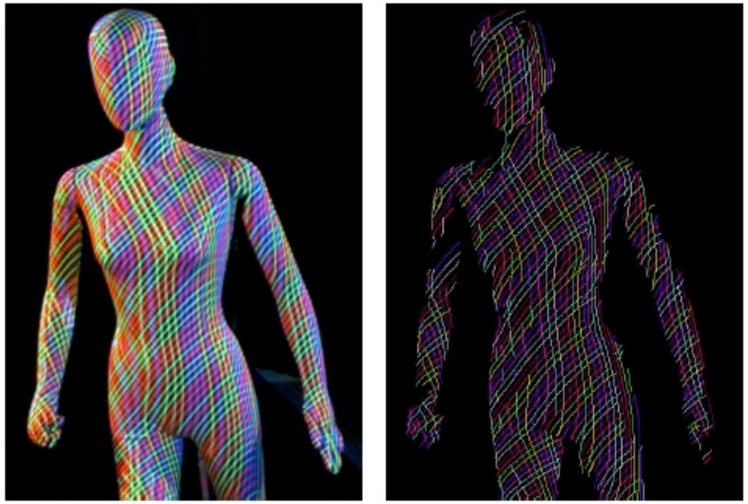
Data Structure Dimension

- **Voxel** – One or more points are grouped into voxels which are directly coded. A voxel is said to be occupied if it contains at least one point of the PC.
- **Curve** – Points are grouped into voxels and after organized into 3D curves with some length and shape.
- **Block** – Points are grouped into voxels and after organized into regular 3D blocks with $a \times b \times c$ voxels size, including both filled and empty voxels. The full PC may be coded as a single block or divided in multiple blocks as common in image and video coding standards.
- **Tree** – Points are organized in a tree, e.g. octrees, kd-trees or spanning trees; while octrees and kd-trees divide the 3D space into blocks that can contain zero, one, or several points, spanning trees generally order the points by minimizing the distance between successive points.
- **Patch** – Points are organized into local groups with the same or arbitrary sizes, defined using a criterion associated to the similarities between neighboring points and to the surface they belong to.
- **Surface** – Points are organized as some type of surface that can be represented by a more or less complex set of parameters, e.g. represented as a set of triangles.
- **Graph** – Points are organized in a graph, this means into a set of nodes/vertices and corresponding connections/edges, using some criteria; since a PC may be rather large, the full structure may be divided into multiple graphs to control their size.

Data Structure Examples

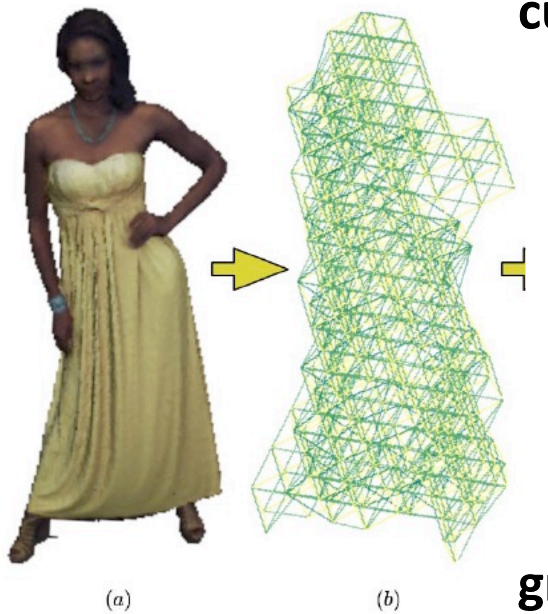


tree



curve

patch



graph

Prediction Dimension

- **None** – No prediction is performed at all.
- **Spatial** – The component data, structured in a specific way, is coded while exploiting the correlation within a single time instant; this is called Intra-coding.
- **Temporal** – The component data, structured in a specific way, is coded while exploiting the correlation along time, considering or not the motion; this is called Inter-coding.
- **Hybrid** – The component data, structured in a specific way, is coded while exploiting both the spatial and temporal correlations; this may involve the definition of (Intra/Inter) coding modes, which are adaptively selected for different parts of the content.

Transform Dimension

- **None** – No transform is applied at all.

- **Block-based** – A transform is applied to some appropriate signal or residual signal, structured as a regular block; this includes for example, both DCT and wavelet, where the block may include part or the full PC. These transforms may be fixed or hand-crafted (e.g. DCT), adaptive (e.g. KLT), or learned (e.g. deep learning-based).

- **Graph-based** – A graph-based transform is applied to some appropriate signal or residual signal, structured as one or more graphs.

TÉCNICO LISBOA

3.2 Point Cloud Coding: Main Coding Approaches

Static Point Cloud Coding

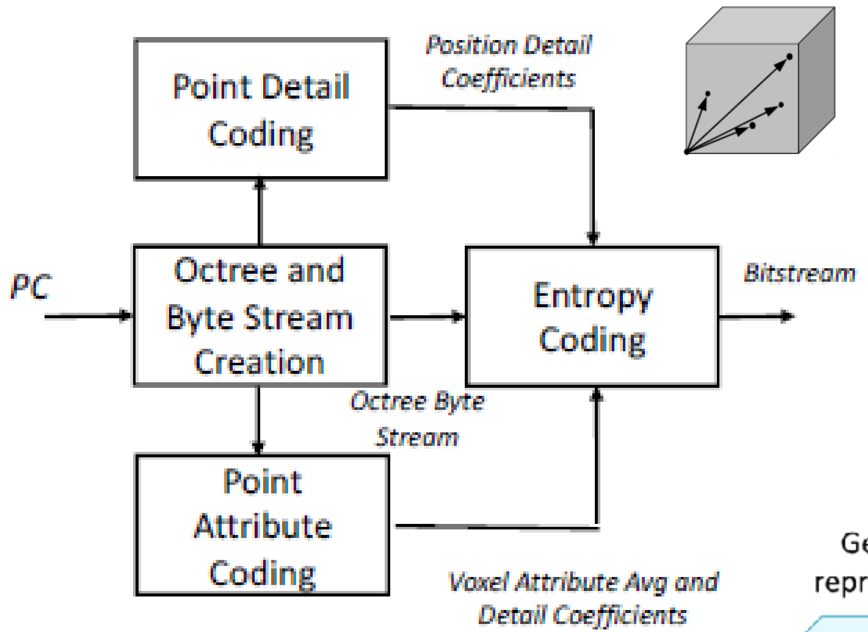
   

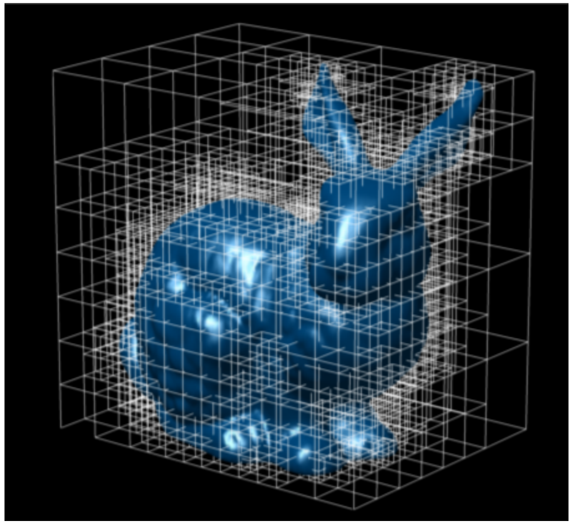
Point Cloud Library (PCL) Octree-based Coding

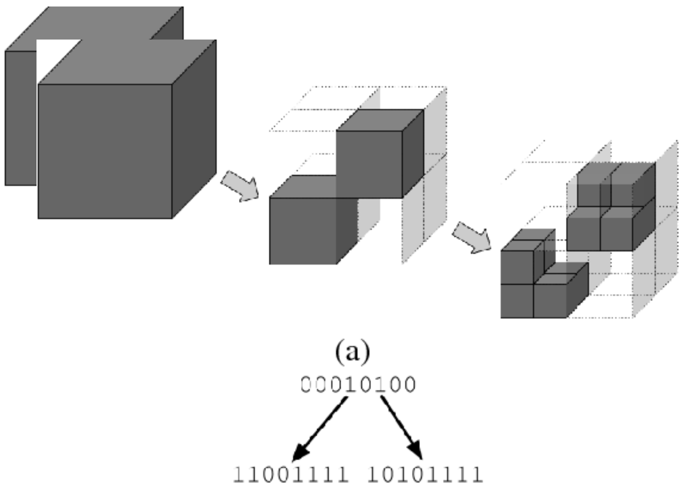
See G-PCC

- PCL coding is octree-based and can handle unorganized point clouds of arbitrary size and density.
- PC geometry is coded as the occupancy symbols of the nodes (or voxels) on an octree; this coding may be lossy or lossless depending on the octree depth/resolution.
- It is possible to code the exact position of one or more points within a leaf voxel by coding the distance between each point and the voxel origin, generating the so-called *point detail coefficients.*
- The color for each leaf voxel is coded as the average of the colors of the points in the voxel; if further precision is required, the color difference for each point in the voxel regarding the average voxel color is coded.
- PCL PC coding solution is rather popular due to the public availability of software and is often used as benchmark for novel PC coding solutions.
- *e.g. J. Kammerl et al., "Real-time compression of point cloud streams", in IEEE Int. Conf. Robotics Automation, Saint Paul, MN, USA, May 2012, pp. 778–785.*

Octree-based Encoding

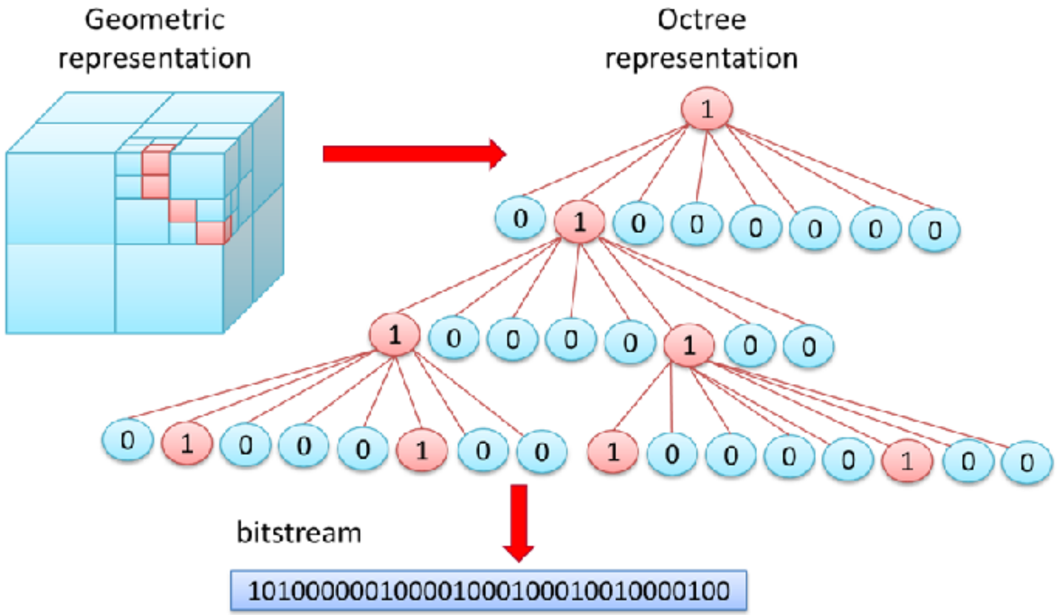






(a)

(b)



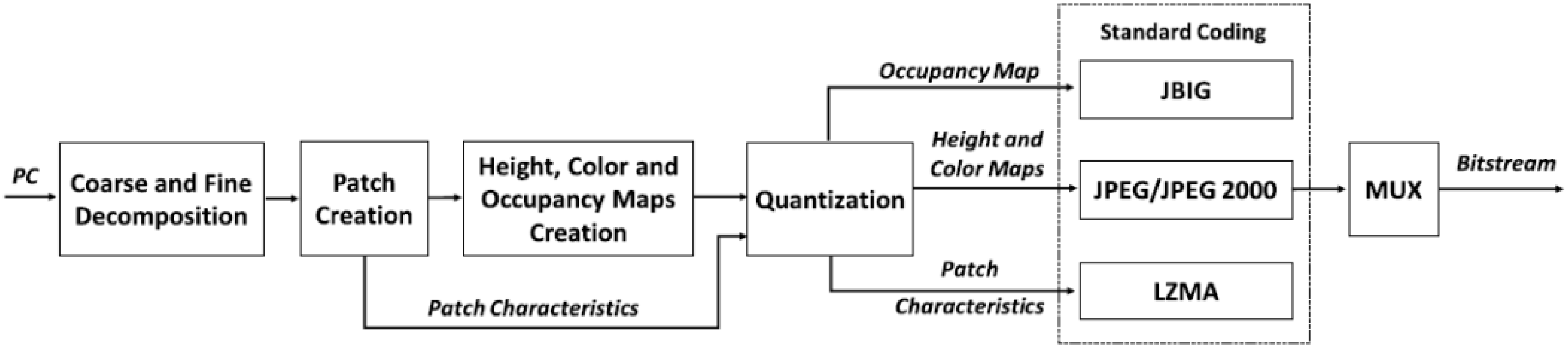
Patch-based Point Cloud Coding

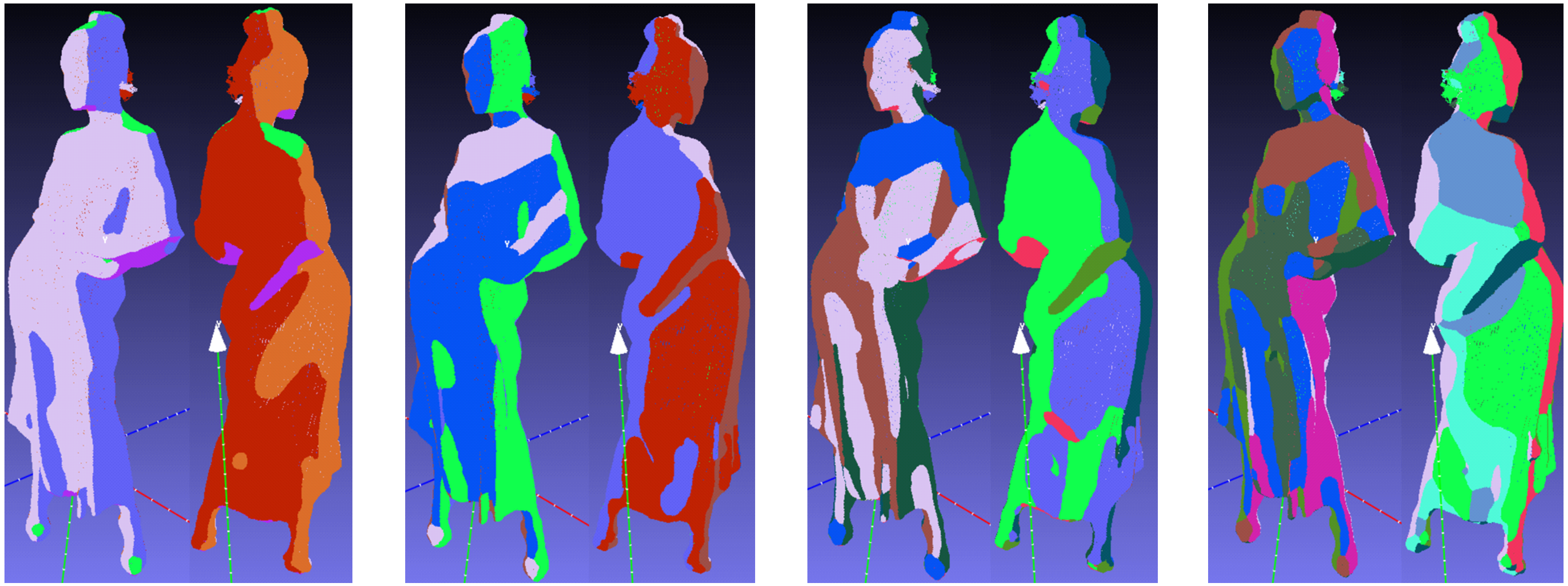
See V-PCC

- PC coding solution exploiting the spatial correlation, both in terms of color and geometry, using a lossy patch-based approach.
- Main idea is to create patches, e.g. associated to leaf voxels within chunks, characterized by position, orientation and size, AND three 2D maps.
- Patches are projected into a 2D grid so that standard image/video codecs can be applied to the height, color and occupancy maps.
- Height map indicates the position of the points in the 3D world while the color map indicates the color for each point.
- Occupancy map indicates which points from the 2D maps really exist in the 3D point cloud.
- Acquired data may be incrementally coded to offer random access with local decoding, and subsampled reconstructions decoding.
- *e.g. T. Golla and R. Klein, "Real-time point cloud compression", in IEEE Int. Conf. Intelligent Robots Systems, Hamburg, Germany, Oct. 2015.*

TÉCNICO LISBOA

Patch-based Point Cloud Coding



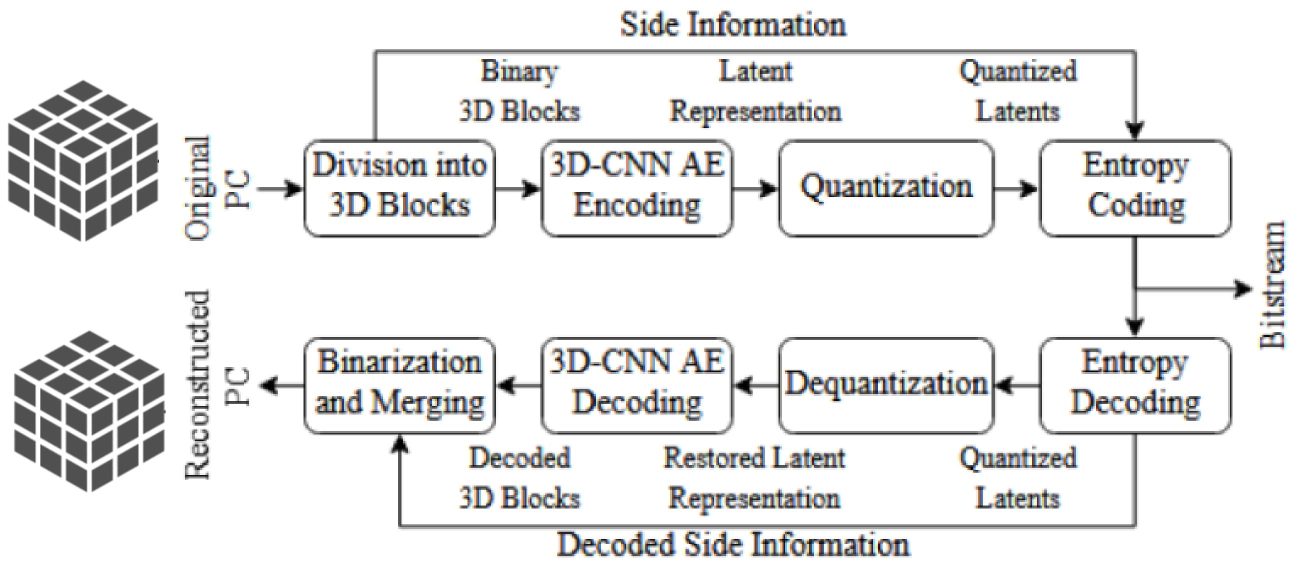


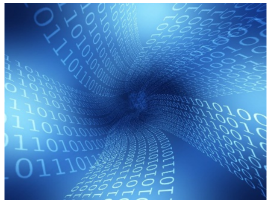
TÉCNICO LISBOA

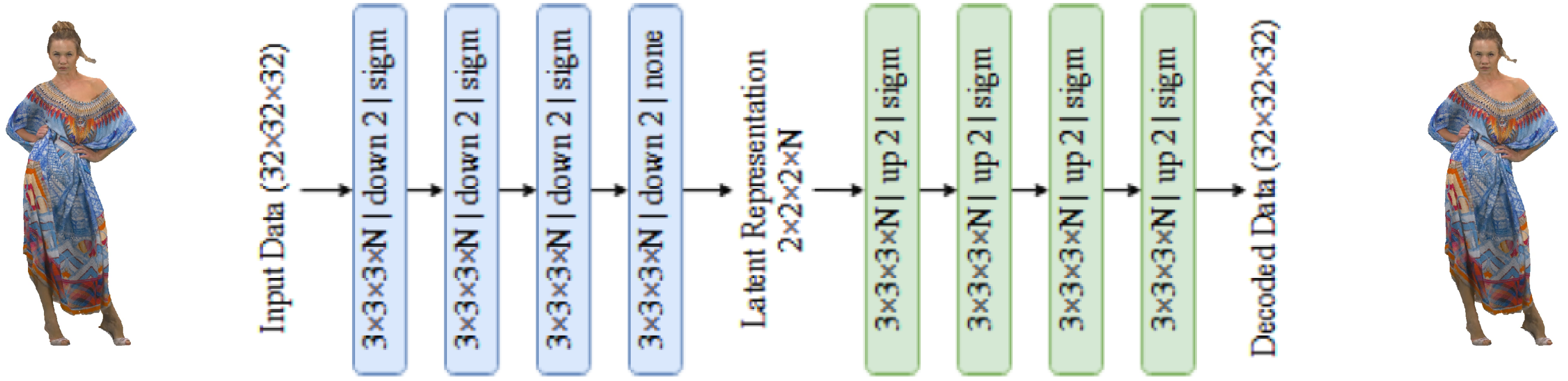
Block-based Point Cloud Geometry Coding

- **PC is interpreted as a binary signal defined over the voxel grid, structured as 3D blocks.**
- **For the deep learning-based solutions, the PC data may be represented by a set of so-called *latents*, generated using a Convolutional Neural Network (CNN) based autoencoder model, after training with a substantial amount of PC data.**
- **This data-driven training process allows learning appropriate analysis and synthesis convolutional transforms in opposition to using hand-crafted transforms.**
- **Decoding process is understood as a binary classification of the PC voxel occupancy, either filled or empty.**
- **Full PC may be coded as a single block with equal edge sizes, i.e. a cube, or segmented into 3D blocks, typically cubical volumes, which are independently coded.**
- ***e.g. A. Guarda et al., "Point cloud coding: adopting a deep learning-based approach", in Picture Coding Symposium, Ningbo, China, November 2019.***

Block-based Point Cloud Geometry Coding

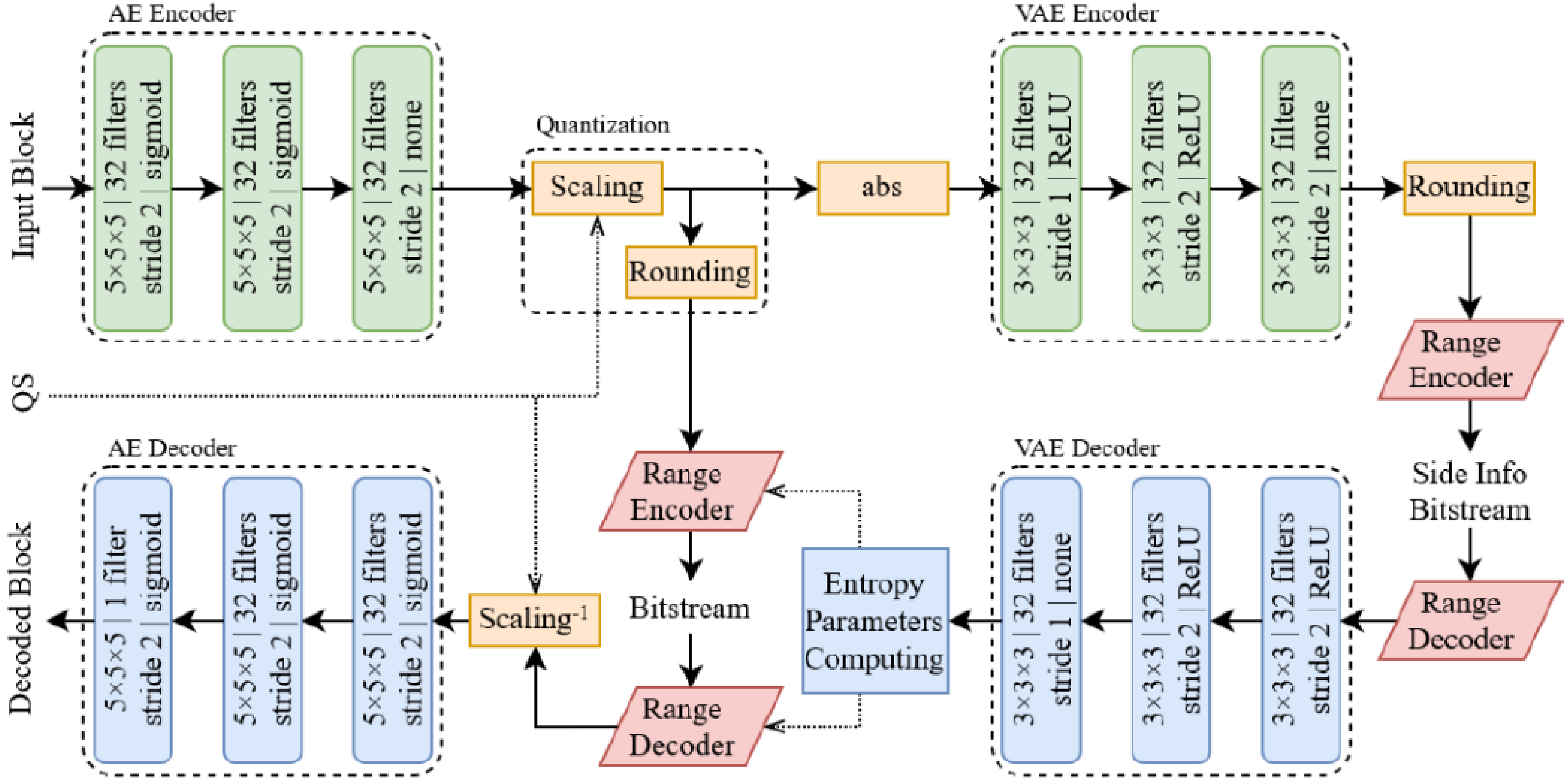






from A. Guarda et al., "Point cloud coding: adopting a deep learning-based approach", in Picture Coding Symposium, Ningbo, China, November 2019.

Combining CNNs …



from A. Guarda et al., "Deep learning-based point cloud geometry coding: RD control though implicit and explicit quantization", ICMEW'2020, London, United Kingdom, July 2020.

TÉCNICO LISBOA

Dynamic Point Cloud Coding

Dynamic Point Clouds ...

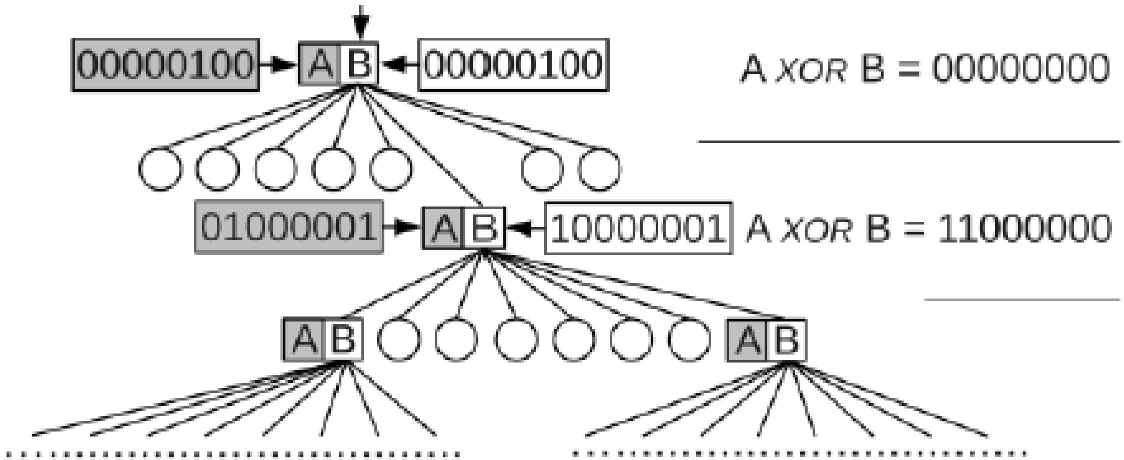
from D. Thanou et al., "Graph-based compression of dynamic 3D point cloud sequences", IEEE Trans. Image Process., Apr. 2016.

The successive frames of a dynamic PC have a different number of points (or filled voxels) and no explicit correspondence between the points is available.

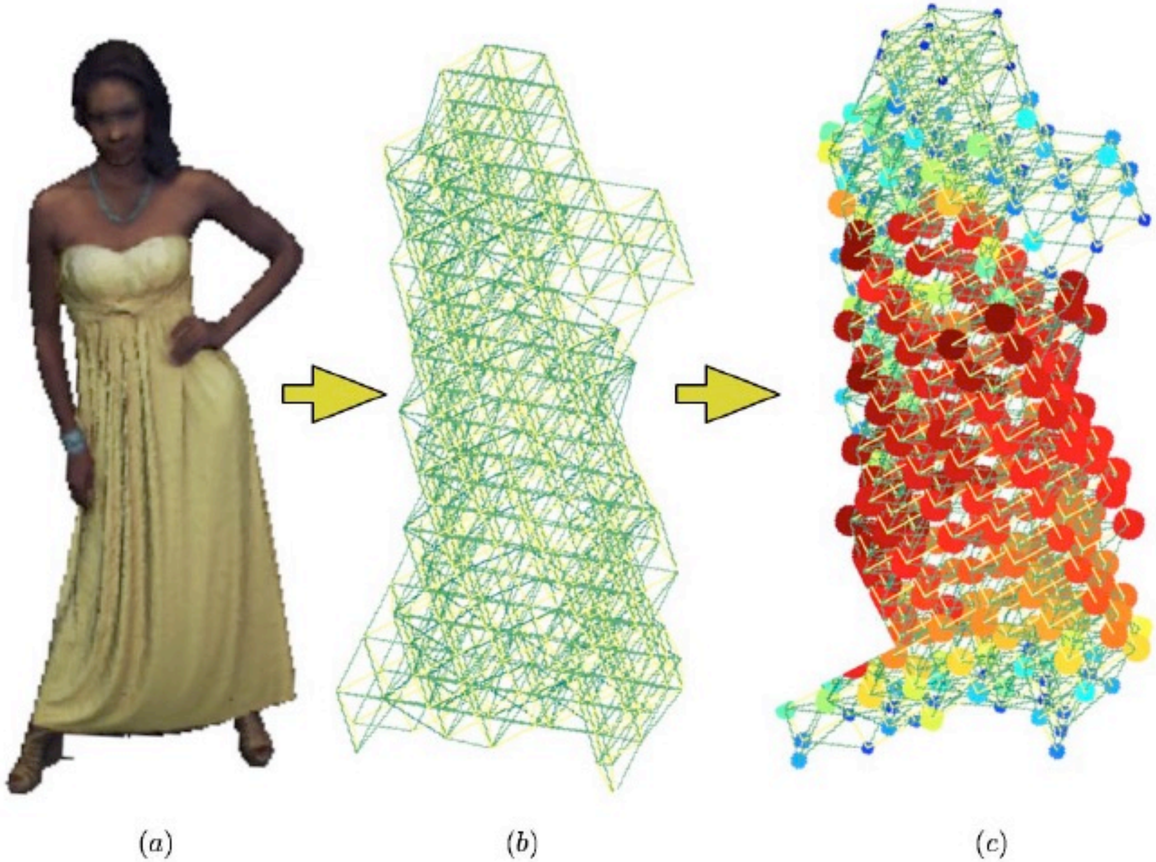
Estimating the motion for removing the temporal redundancy is a rather difficult task !

Point Cloud Library Octree-based Coding

- Dynamic geometry is coded by applying a XOR operation between the occupancy symbol streams of two successive octree frames, in practice predicting the geometry along time, although *without any motion compensation*.
- This process is called *double buffered octree prediction* and allows the PCL coding solution to exploit the temporal correlation in dynamic PC, captured at successive time instants.
- For attributes, notably color coding, no additional tools are used regarding the static coding solution.



Point Clouds and Graphs



A Graph G is defined by the voxels taken as graph vertices and edges between nearby vertices.

The attributes of each voxel, including 3D position and color components, are treated as signals residing on the graph vertices.

- (a) Example of a point cloud of the 'yellow dress' sequence.
- The geometry is captured by a graph (b) and the R component of the color is considered as a signal on the graph (c).
- The size and the color of each disc indicate the value of the signal at the corresponding vertex.

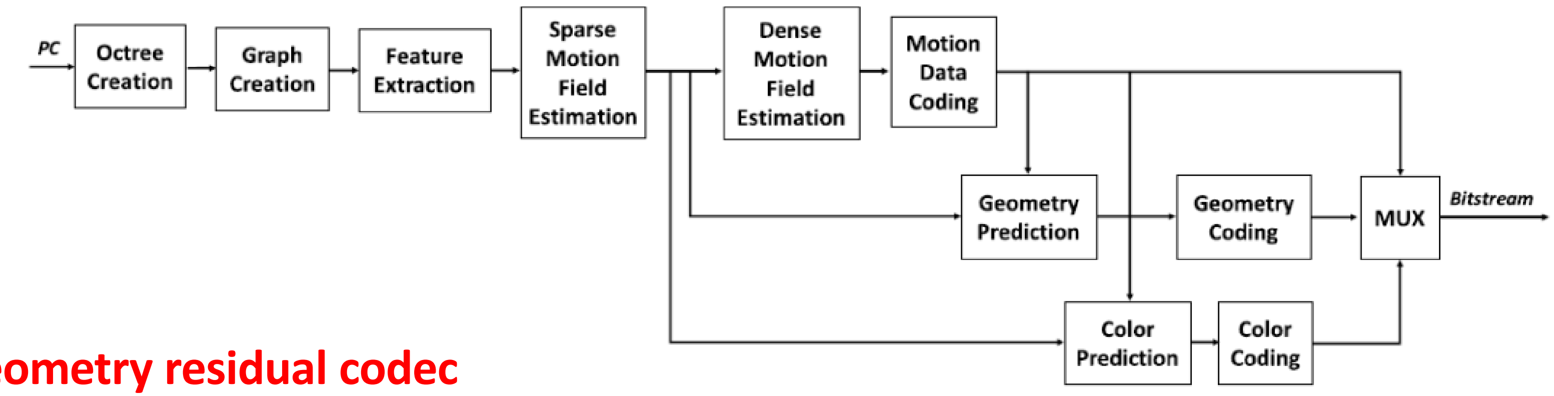
from D. Thanou et al., "Graph-based compression of dynamic 3D point cloud sequences", IEEE Trans. Image Process., vol. 25, no. 4, pp. 1765–1778, Apr. 2016.

Graph-based PC Coding

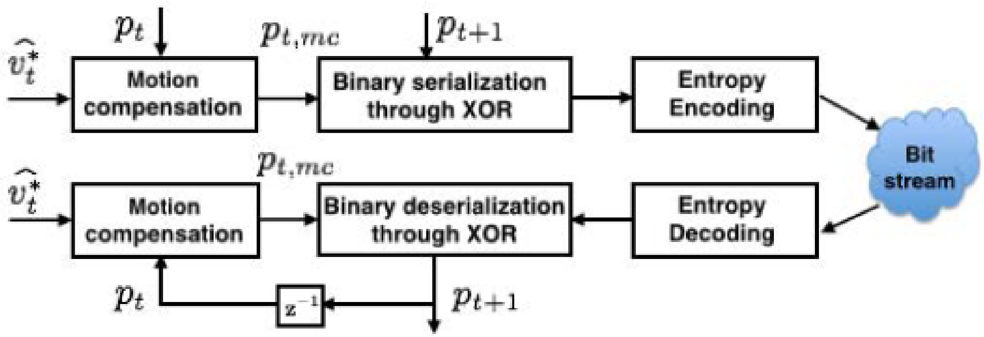
- **PC coding based on the exploitation of the temporal correlation using a graph-based representation for each frame (where each occupied voxel is mapped to a node in a graph) to extract the motion.**
- **The connectivity between nodes is established based on a KNN method, where *K* is the maximum number of adjacent neighbors for a node. Each connection gets a weight that is inversely proportional to the distance between the corresponding voxel centers.**
- **Motion estimation problem is formulated as a feature matching problem in dynamic graphs (wavelet coefficients).**
- **Motion vectors are used to motion compensate both the geometry and color, thus creating geometry and color predictions and the corresponding residues.**
- **A differential coding scheme is used to compress the geometry and color, thus exploiting the estimated motion information.**
- **While the geometry residual is coded using a double buffered octree approach, the color residual is coded with a graph-based transform.**
- ***e.g. D. Thanou et al., "Graph-based compression of dynamic 3D point cloud sequences", IEEE Trans. Image Process., vol. 25, no. 4, pp. 1765–1778, Apr. 2016.***

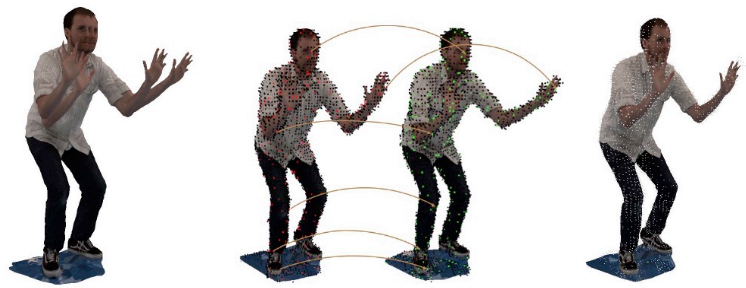
Graph-based PC Coding

Overall encoder



Geometry residual codec





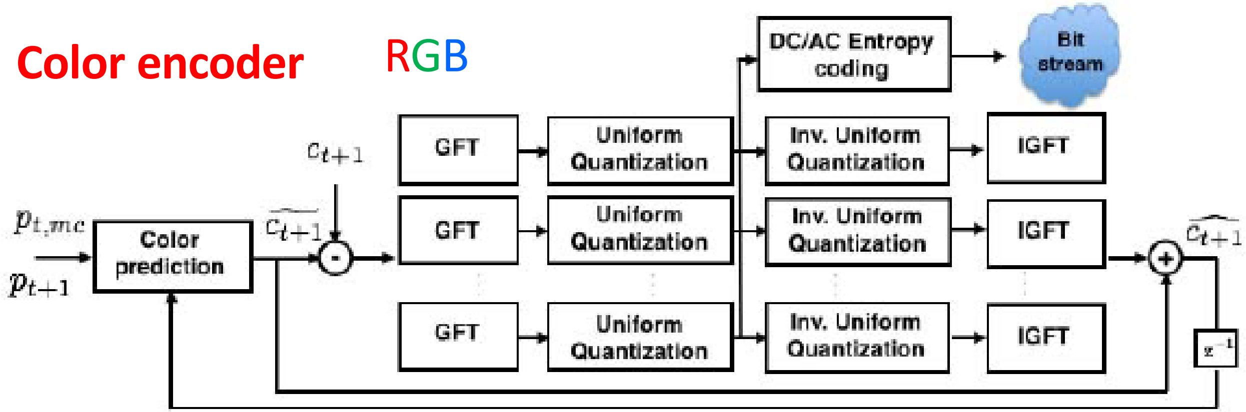
(d) $\mathcal{I}_t + \mathcal{I}_{t+1}$ (e) Correspondence between $\mathcal{I}_t$ and $\mathcal{I}_{t+1}$ (f) $\mathcal{I}_{t,mc} + \mathcal{I}_{t+1}$

v_t – motion vector field

GFT – Graph Fourier Transform

from D. Thanou et al., "Graph-based compression of dynamic 3D point cloud sequences", IEEE TIP, vol. 25, Apr. 2016.

Color encoder RGB



Take Home Messages …

- After 3-DoF, 6-DoF experiences are the next step towards visual immersion.
- Point clouds may offer 6-DoF immersion and are here to stay !
- For large deployment, efficient point cloud coding is a must !
- Since interoperability is critical, point cloud coding standards are needed …


ASK MORE
QUESTIONS
GET MORE ANSWERS
Anthony Burrill