Tutorial on
Immersive Imaging Technologies:
from Capture to Display

Dr. Martin Alain, Research Fellow — V-SENSE, SCSS, Trinity College Dublin, Dublin, Ireland
Dr. Cagri Ozcinar, Senior Research Engineer — Samsung R&D Institute UK (SRUK), UK
Dr. Emin Zerman, Research Fellow — V-SENSE, SCSS, Trinity College Dublin, Dublin, Ireland

VCIP 2021 Tutorial | Munich, Germany | 2021-12-05

About the Presenters

Martin Alain: Postdoctoral Research Fellow at Trinity College Dublin

- PhD at INRIA Rennes and Technicolor, France
- Tutorial on light field imaging for ITN RealVision at Erlangen, Germany
- Research interests: Light field imaging, image processing, computer vision

Cagri Ozcinar: Senior Research Engineer at Samsung R&D Institute UK

- PhD in multimedia signal processing and communication at University of Surrey
- Research interests: Signal/image processing, machine learning, virtual and augmented reality, coding and streaming, visual attention/saliency

Emin Zerman: Postdoctoral Research Fellow at Trinity College Dublin

- PhD in signal and images at Télécom ParisTech, France
- Research interests: Image and video quality assessment, human visual perception, human computer interaction

About V-SENSE

V-SENSE project team

- Extending Visual Sensation through Image-Based Visual Computing
- 20+ researchers

- Light Field Imaging
- 3DoF – 360 VR Video
- Visual Effects & Animation
- 6DoF – AR/VR & Volumetric Video
- Deep Learning for Visual Computing

Outline

1. Immersive Imaging Technologies

- Immersion & tele-immersion
- Imaging modalities
- Applications

2. Acquisition and Data Format

- Single-camera systems
- Multi-camera systems

3. Content Delivery

- Coding
- Adaptive streaming

4. Rendering and Display Technologies

- Immersive imaging on 2D screens
- HMDs for VR
- HMDs for AR

5. Perception & Quality Evaluation

- Visual perception
- Quality assessment
- Visual attention

Outline

13:30	Part I: Immersive Imaging Technologies	~20 minutes
	Part II: Acquisition and Data Formats	~40 minutes
	5-10 mins Q&A	
	Part III: Content Delivery	~20 minutes
15:00	Coffee Break	
15:30	Continuing Part III: Content Delivery	~20 minutes
	5-10 mins Q&A	
	Part IV: Rendering and Display Technologies	~20 minutes
	5 mins Q&A	
	Part V: Perception & Quality Evaluation	~40 minutes
	5-10 mins Q&A	
17:00		

Part I: Immersive Imaging Technologies

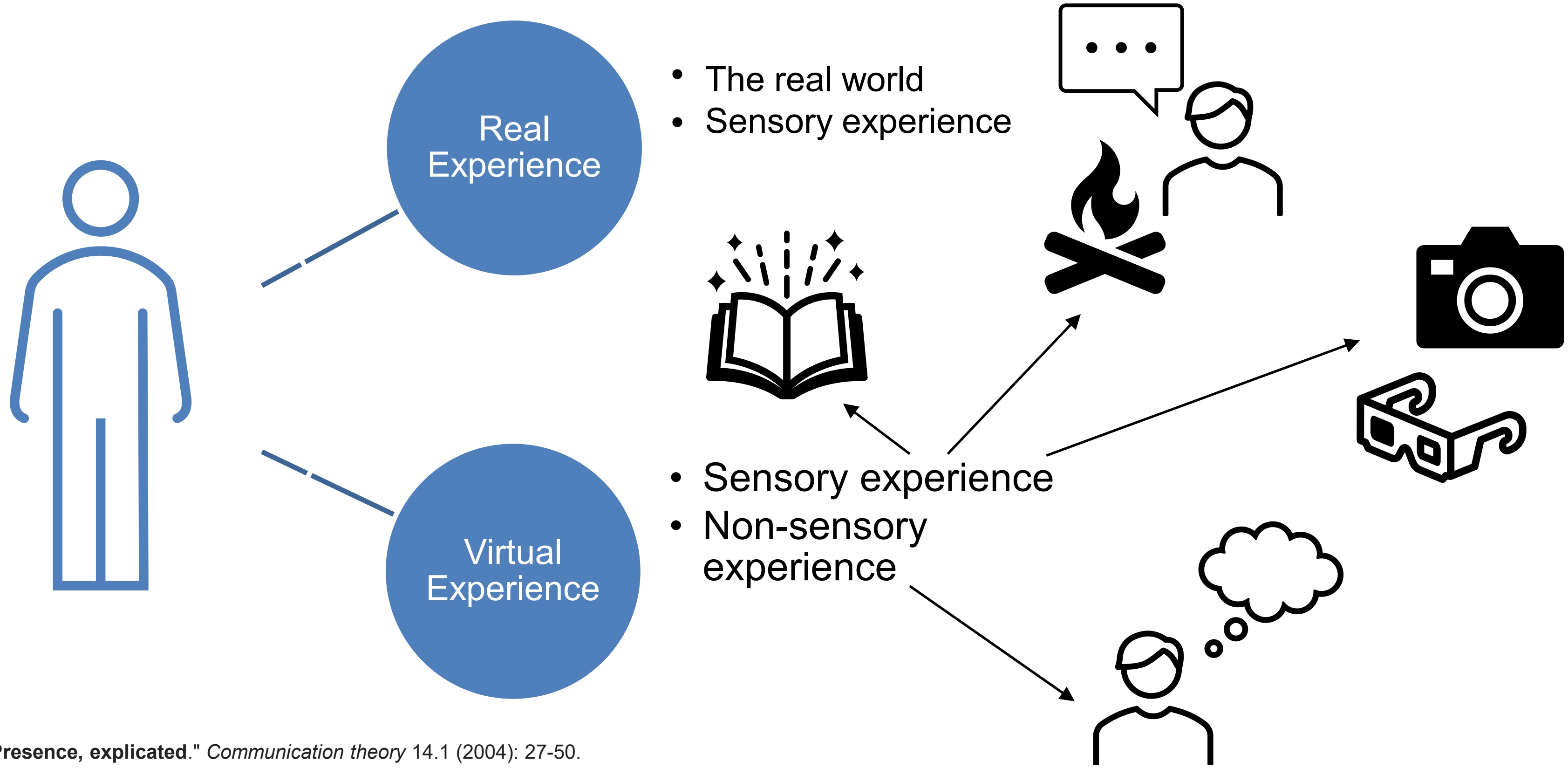
What is immersion? What are immersive imaging technologies? What are the applications?

Part I: Immersive Imaging Technologies

Three main points:

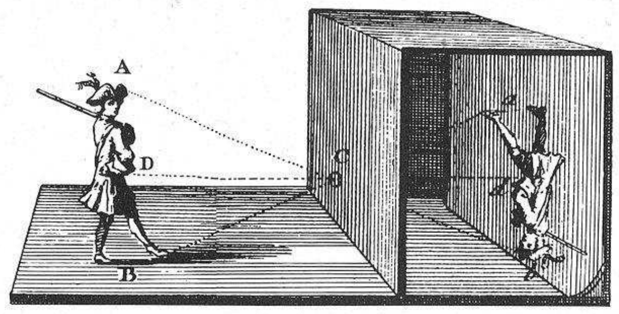
- Immersion & tele-immersion
- Imaging modalities
  - Using traditional imaging
  - Light fields
  - Omnidirectional imaging
- Applications

Human Experience



Lee, Kwan Min. "**Presence, explicated**." *Communication theory* 14.1 (2004): 27-50.

Imaging Technologies



Imaging

- Storing visual information on a semi-permanent or permanent medium

- Pinhole camera
- Photography
  - Photographic plates
  - Film
  - Digital

- Enable telepresence & information transmission in great distances

Immersion & Tele-Immersion

Telepresence (Minsky, 1980)

- "sense of being physically present at a remote location through interaction with the system's human interface,"

Presence (Ijsselsteijn, 2000)

- "***being there*** in a mediated environment"
- Lombard and Ditton (1997)
  - Realism
  - Transportation
  - Immersion, etc.

Immersion (Takatalo, 2008)

- "concentration to the VE instead of the real world, loss of time"

Tele-Immersion (Mulligan, 2001)

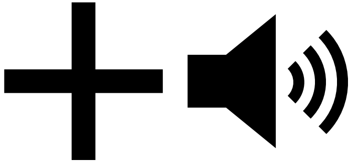
- "interaction (presence) of people from different places in virtually the same environment."

Minsky, Marvin. "**Telepresence**," *Omni*, (1980): 45-51.
Mulligan, Jane, and Kostas Daniilidis. "**Real time trinocular stereo for tele-immersion**." *Proceedings 2001 International Conference on Image Processing*, Vol. 3. IEEE, 2001.
IJsselsteijn, Wijnand A., et al. "**Presence: concept, determinants, and measurement**." *Human vision and electronic imaging V*. International Society for Optics and Photonics, 2000.
Takatalo, Jari, Göte Nyman, and Leif Laaksonen. "**Components of human experience in virtual environments**." *Computers in Human Behavior* 24.1 (2008): 1-15.
Lombard, Matthew, and Theresa Ditton. "**At the heart of it all: The concept of presence**." *Journal of computer-mediated communication* 3.2 (1997).

Imaging Technologies & Immersion

Timeline of imaging and immersion

- Lumière brothers’ train
- Silent films → Audio
- Color films
- High dynamic range
- Immersive imaging technologies



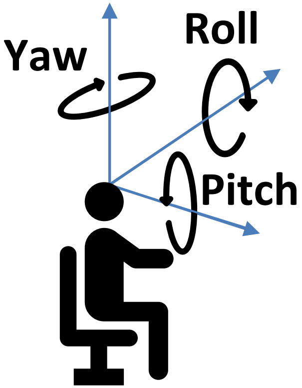
Imaging Technologies & Immersion

Immersive Imaging Technologies

- Extend visual* sense
- Augment the “presence”
- Provide the viewer with a higher degree of freedom

* Although different modalities can be also included, immersive imaging technologies mostly focus on extending visual sensation

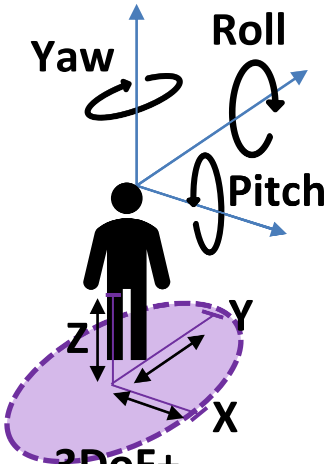
Immersion & Degrees of freedom (DoF)



3DoF

- Rotation around 3 axes
- <u>**No**</u> spectator movement

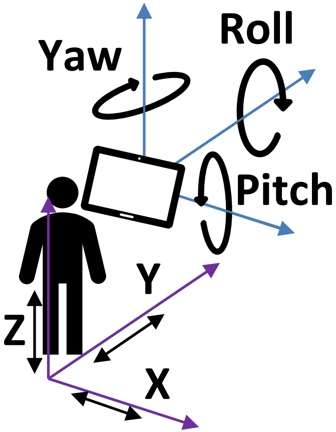
- 3 degrees-of-freedom in total



3DoF+

- Rotation around 3 axes
- <u>**Limited**</u> spec. movement

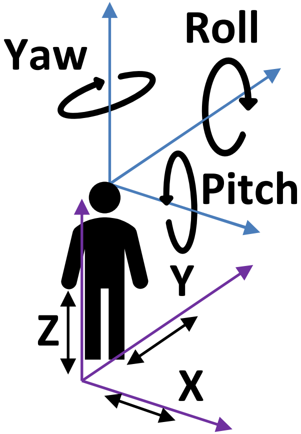
- More than 3 DoF in total, but within limits



Windowed 6DoF

- Rotation around 3 axes
- Spectator movement through 3 axes

- Essentially 6 DoF with the help of a <u>handheld display</u>



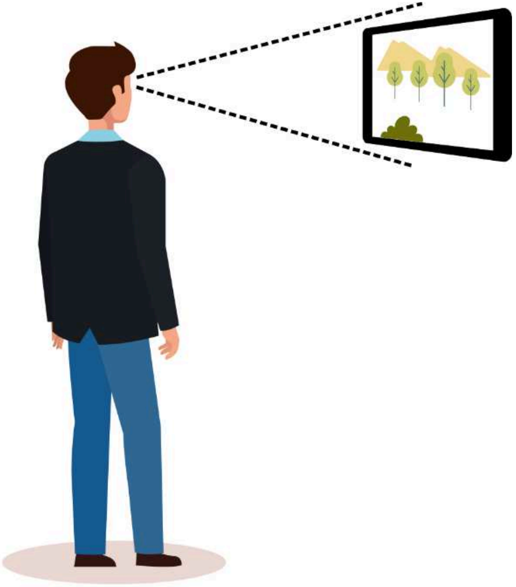
6DoF

- Rotation around 3 axes
- Spectator movement through 3 axes

- 6 degrees-of-freedom in total

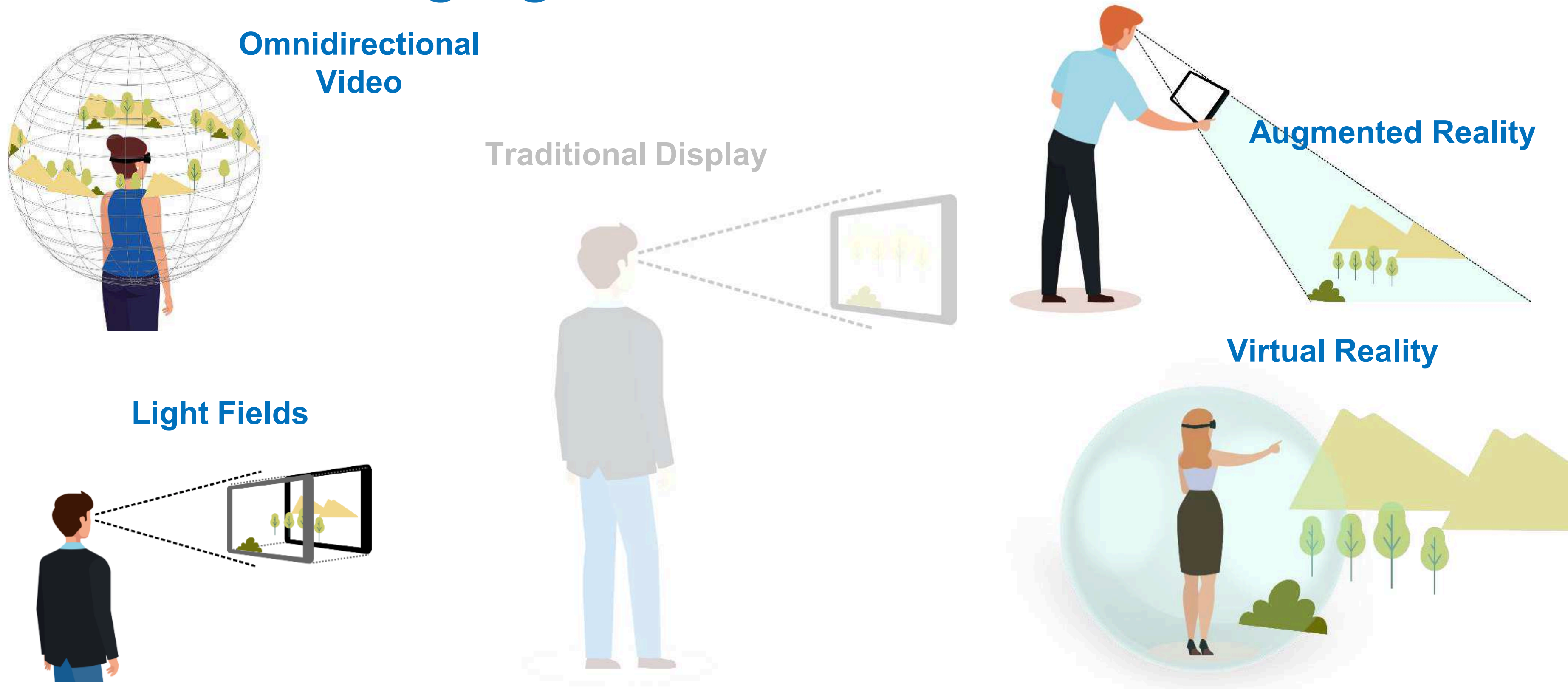
Different Imaging Modalities

Traditional Display



Attribution: People illustrations were created by studiogstock - www.freepik.com

Different Imaging Modalities

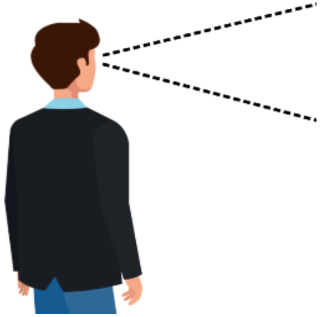


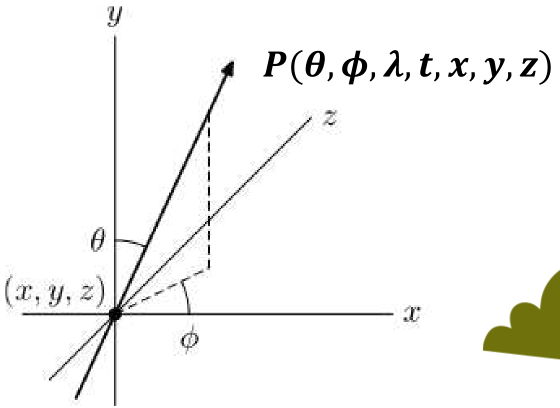
Attribution: People illustrations were created by studiogstock - www.freepik.com

Different Imaging Modalities

Theoretical background

The plenoptic function





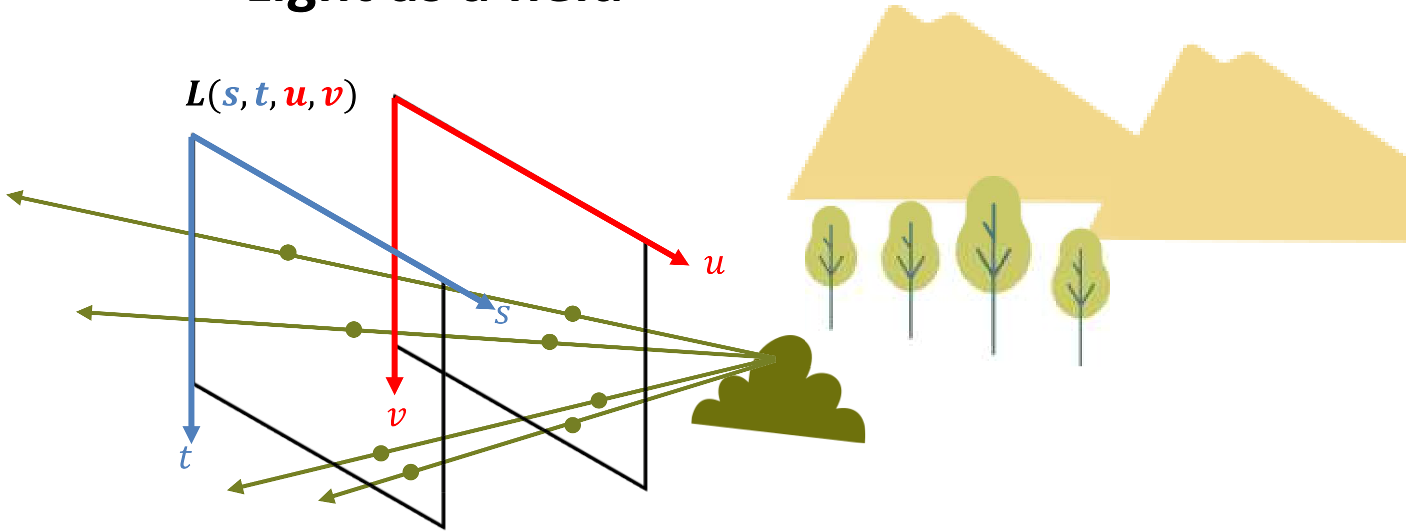
Adelson, Edward H., and James R. Bergen. “**The Plenoptic Function and the Elements of Early Vision**.” *Computational Models of Visual Processing*, edited by Michael Landy and J. Anthony Movshon, MIT Press, 1991, pp. 3–20.

Represents all the information available to an observer at any point in space and time

Different Imaging Modalities

Light fields

Light as a field

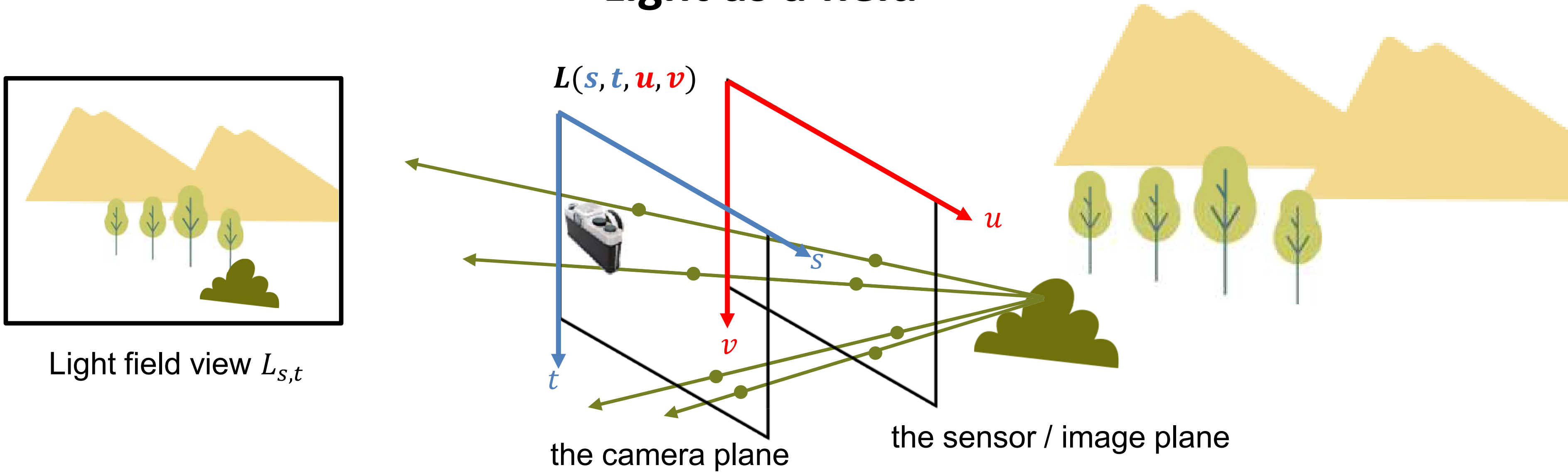


Light rays parameterized by their intersection with two parallel planes

Different Imaging Modalities

Light fields

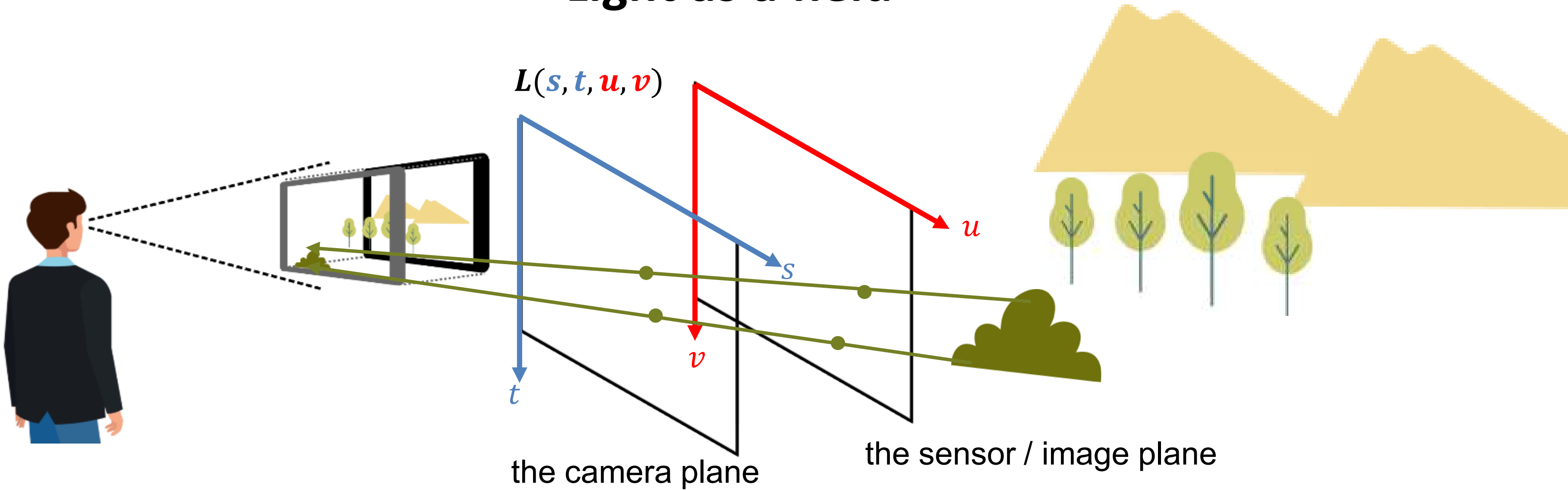
Light as a field



Different Imaging Modalities

Light fields

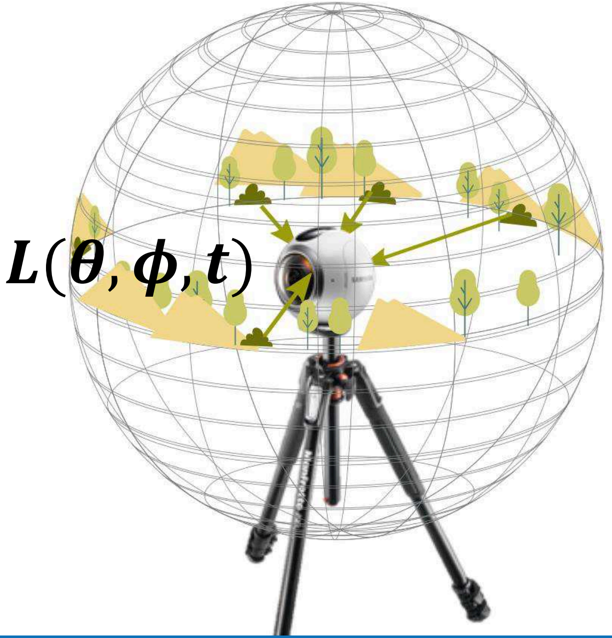
Light as a field



Different Imaging Modalities

Omnidirectional image and video

Omnidirectional (360-degree) image and video



Different Imaging Modalities

Omnidirectional image and video

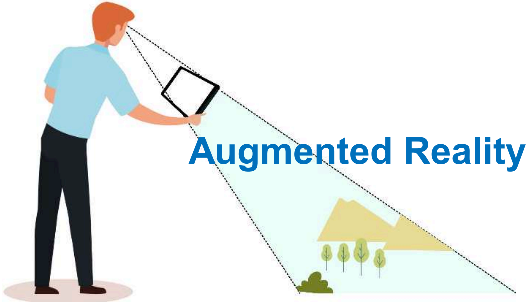
Omnidirectional (360-degree) image and video

Omnidirectional Video

Different Imaging Modalities

Volumetric video, augmented reality, and virtual reality

Volumetric Video or Free-Viewpoint Video

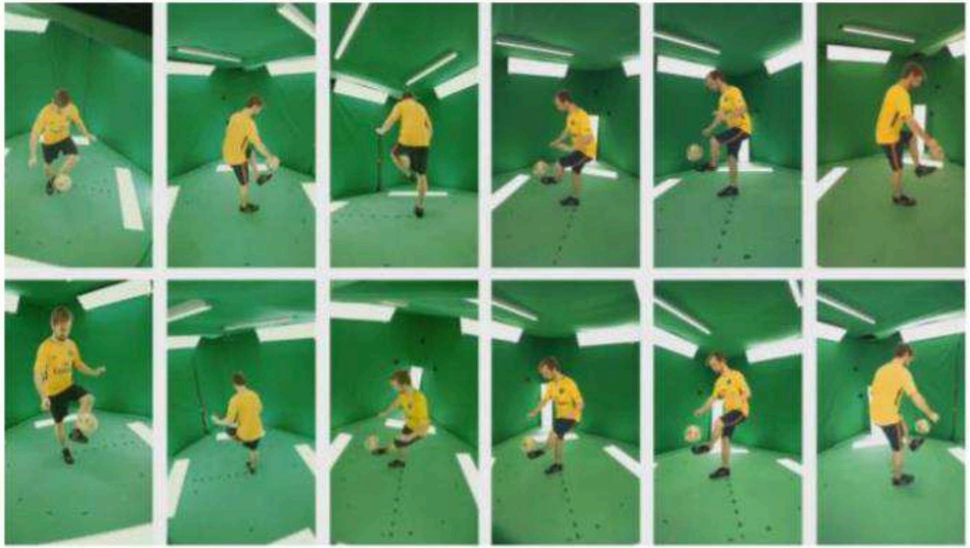


Different Imaging Modalities

Volumetric video, augmented reality, and virtual reality

Volumetric video content creation

Image taken from: Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., ... & Sullivan, S. (2015). **High-quality streamable free-viewpoint video**. *ACM Transactions on Graphics (ToG), 34(4)*, 69.

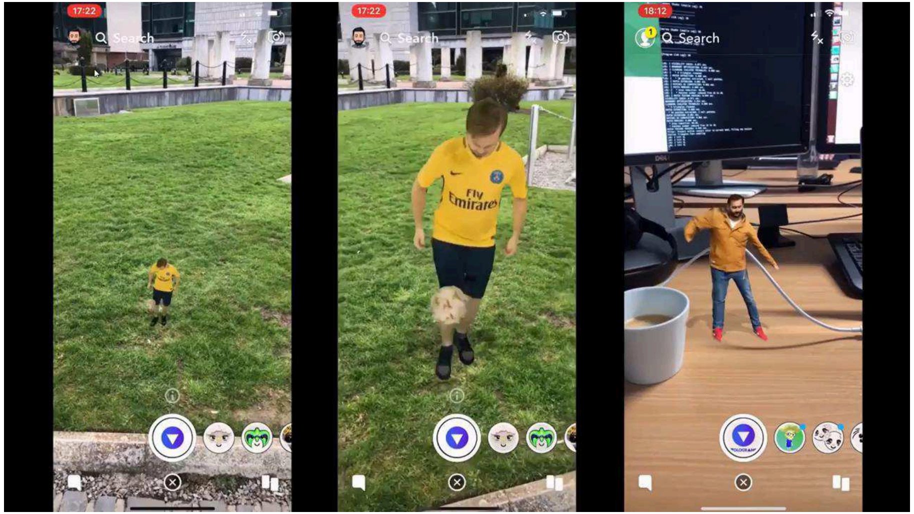


Pagés, Rafael, et al. "**Affordable content creation for free-viewpoint video and VR/AR applications**." *Journal of Visual Communication and Image Representation* 53 (2018): 192-201.

Different Imaging Modalities

Volumetric video, augmented reality, and virtual reality

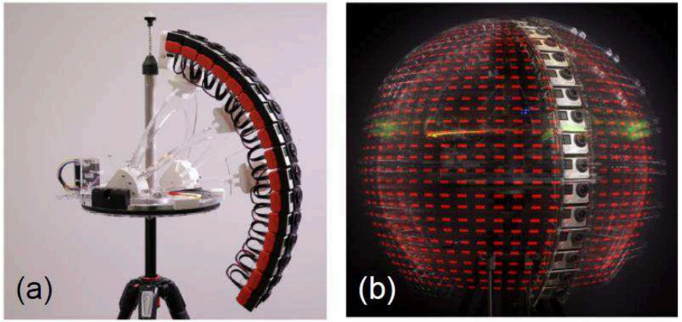
VV can be used in AR & VR applications



Different Imaging Modalities

The boundaries between the different modalities are disappearing

"Welcome to lightfields" (Google)



Overbeck, Ryan S., et al. "**A system for acquiring, processing, and rendering panoramic light field stills for virtual reality**." *ACM Transactions on Graphics (TOG)* 37.6 (2018): 1-15.

Applications & Creative Experiments

Where do we use these immersive imaging technologies?

- Entertainment
- Education
  - ‘Augmenting’ education
  - Cultural heritage
- Communication
  - Immersive communication
  - Remote collaboration
- Novel storytelling
- New internet medium

Applications & Creative Experiments

Entertainment

- Sports
- Drama & Theatre
  - Augmented Play
  - Virtual Play
  - Awake One
- Post-production
- Computer games

Applications & Creative Experiments

Entertainment

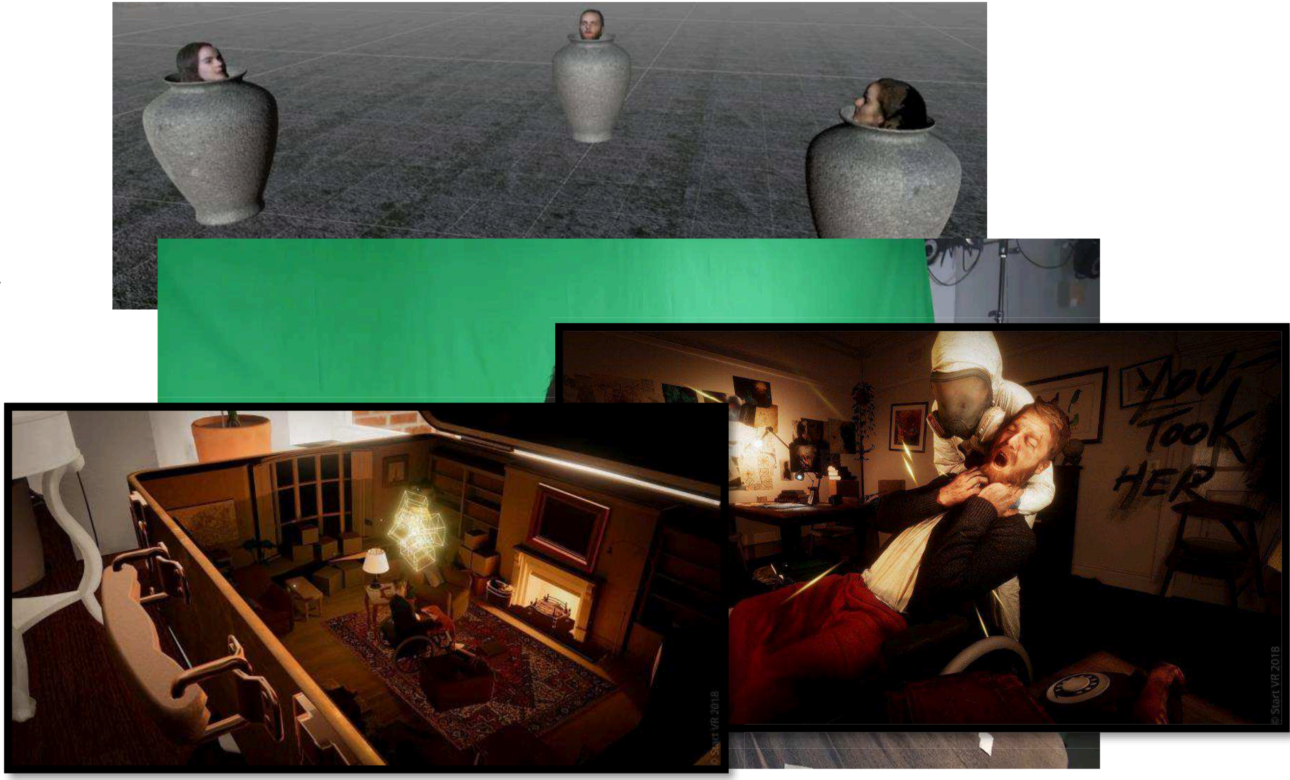
- Sports
- Drama & Theatre
  - Augmented Play
  - Virtual Play
  - Awake One
- Post-production
- Computer games

O'Dwyer, Néill, and Nicholas Johnson. "**Exploring volumetric video and narrative through Samuel Beckett's Play**." *International Journal of Performance Arts and Digital Media* 15.1 (2019): 53-69.

Applications & Creative Experiments

Entertainment

- Sports
- Drama & Theatre
  - Augmented Play
  - Virtual Play
  - Awake One
- Post-production
- Computer games



O'Dwyer, Néill, and Nicholas Johnson. "**Exploring volumetric video and narrative through Samuel Beckett's Play**." *International Journal of Performance Arts and Digital Media* 15.1 (2019): 53-69.

Attribution: VR Focus - https://www.vrfocus.com/2018/12/cinematic-experience-awake-episode-one-now-available-for-htc-vive/

Applications & Creative Experiments

Entertainment

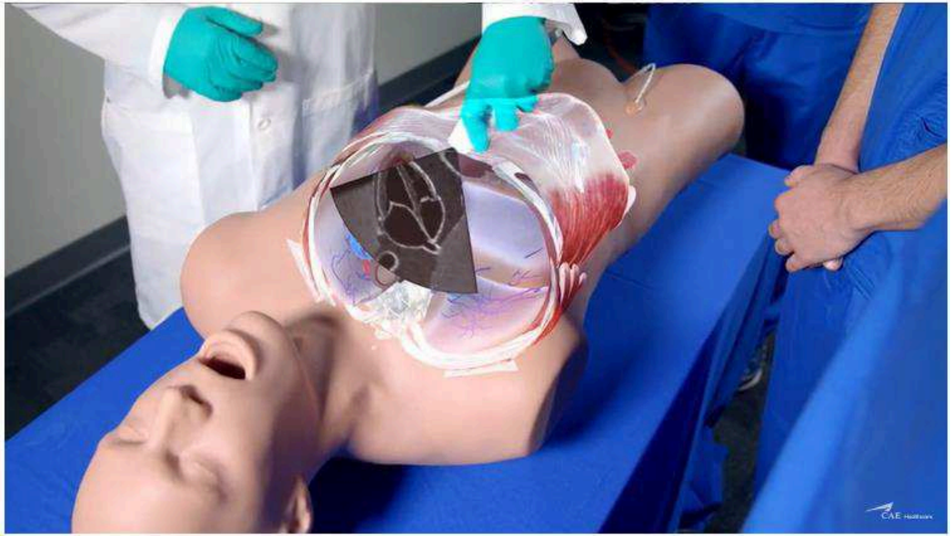
- Sports
- Drama & Theatre
  - Augmented Play
  - Virtual Play
  - Awake One
- Post-production
- Computer games

Fraunhofer IIS – "Coming home" breakdown

Applications & Creative Experiments

Education

- ‘Augmenting’ Education
  - Realistic representation
  - Simulations
  - Employee training
  - Drama education
    - Beckett’s “Play”
- Cultural Heritage
  - Museum guide
    - Dean Jonathan Swift in the Old Room of TCD Library



Attribution: AR in Healthcare education - https://jasoren.com/ar-in-healthcare-education/#:~:text=Augmented%20reality%20is%20a%20powerful,unnecessary%20risk%20for%20the%20patients.

Applications & Creative Experiments

Education

- ‘Augmenting’ Education
  - Realistic representation
  - Simulations
  - Employee training
  - Drama education
    - Beckett’s “Play”
- Cultural Heritage
  - Museum guide
    - Dean Jonathan Swift in the Old Room of TCD Library

Amplianitis, Konstantinos, and Aljoša Smolić. "**Jonathan Swift: Augmented Reality Application for Trinity Library’s Long Room**." *Interactive Storytelling: 11th International Conference on Interactive Digital Storytelling, ICIDS 2018*, Dublin, Ireland, December 5–8, 2018. Springer, 2018.

Applications & Creative Experiments

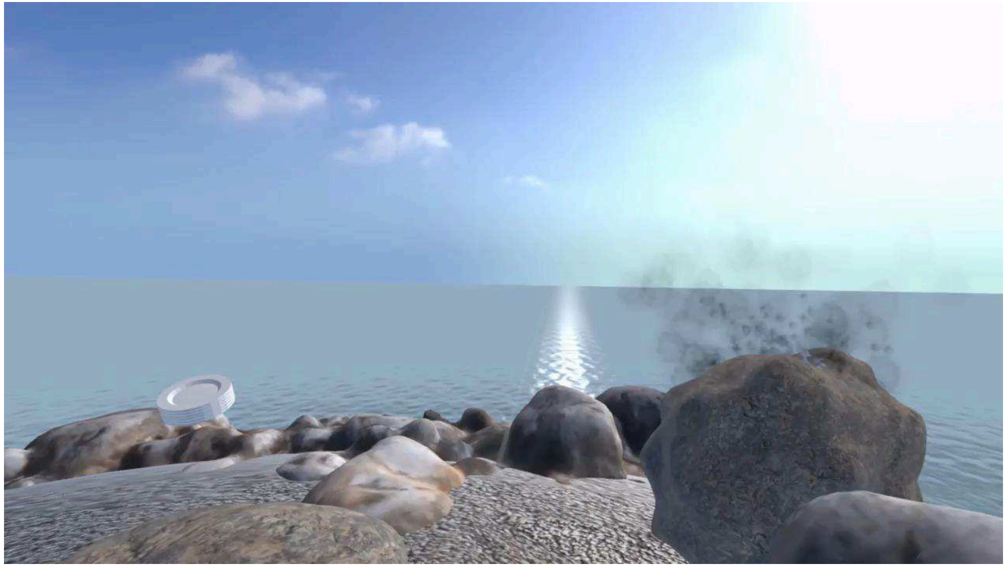
Communication

- Immersive communication
  - “Holoportation”
    - Real-time acquisition and display of volumetric videos
    - https://www.microsoft.com/en-us/research/project/holoportation-3/
- Remote collaboration
  - “Spatial”
    - Remote collaboration platform using many devices: HoloLens, MagicLeap, PC, phone, etc.
    - https://spatial.io/

Applications & Creative Experiments

Novel Storytelling

- V-SENSE - Storytelling
  - “Bridging the Blue” - an immersive creative experiment that explores virtual reality (VR) as “the ultimate empathy machine” where users can explore an imaginary world and experience personal representations of clinical depression.
  - mediated perspective-taking experience of VR



Attribution: V-SENSE, Trinity College Dublin - https://v-sense.scss.tcd.ie/research/bridging-the-blue/

Applications & Creative Experiments

Medium of the Future

- An immersive version of the Internet
- Combination of social media & daily internet use in AR/VR
  - Second Life
  - Altspace VR
  - Facebook Horizon/Metaverse
- More prevalent in fiction as well
  - "Oasis" from Ready Player One
  - "V-World" from Caprica TV series
- Creation of human visual 3D models
- Creation of objects and scenes

Summary: Immersive Imaging Technologies

Key concepts of immersive imaging

- Immersion
- Plenoptic function
- Degrees-of-freedom (DoF)

Different imaging modalities

- Light fields
- Omnidirectional imaging
- Volumetric videos

Applications for these technologies

Next part:

- How to capture immersive images and videos in practice?

Part II: Acquisition and Data Format

How are the different modalities of immersive imaging captured? Which data formats and representations are used to store them?

Part II: Acquisition and Data Format

Light fields

- Both single and multi-camera systems are used

Omnidirectional imaging

- Both single-camera & multi camera systems are used
- More limited in single camera case (e.g., omnidirectional video)

3D models & Volumetric video

- Mostly require multi-camera solutions

Imaging strategies are diverse!

Let’s consider two categories:

- Single-camera systems
- Multi-camera systems

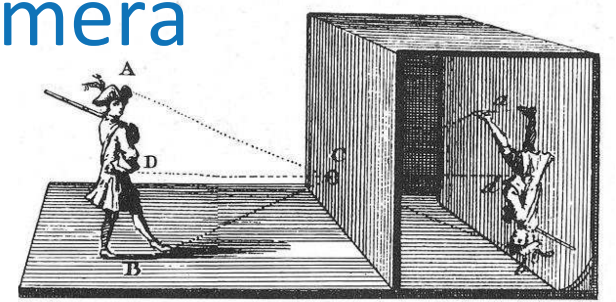
Single-Camera Systems

Imaging with single cameras

- Light fields
  - Robotized 2D cameras
  - Lenslet plenoptic cameras
- Omnidirectional imaging
  - Rotating 2D cameras
  - Single-device systems
- 3D models, point clouds
  - Moving 2D cameras and simultaneous localization and mapping (SLAM)

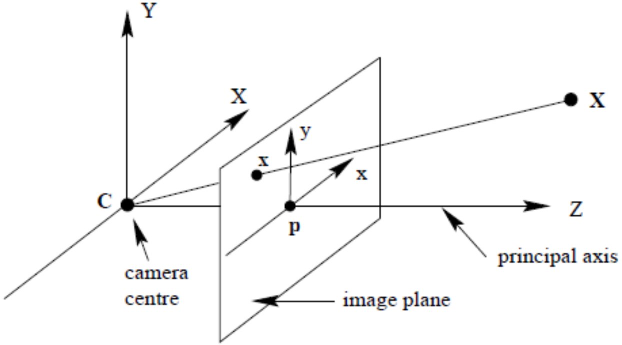


Based on a single 2D camera



2D imaging: the pinhole camera model

- Intrinsics parameters
  - Focal length
  - Principal point
- Extrinsics parameters
  - Position
  - Rotation
- Intrinsics and Extrinsics parameters allow to connect pixels to points in the 3D world

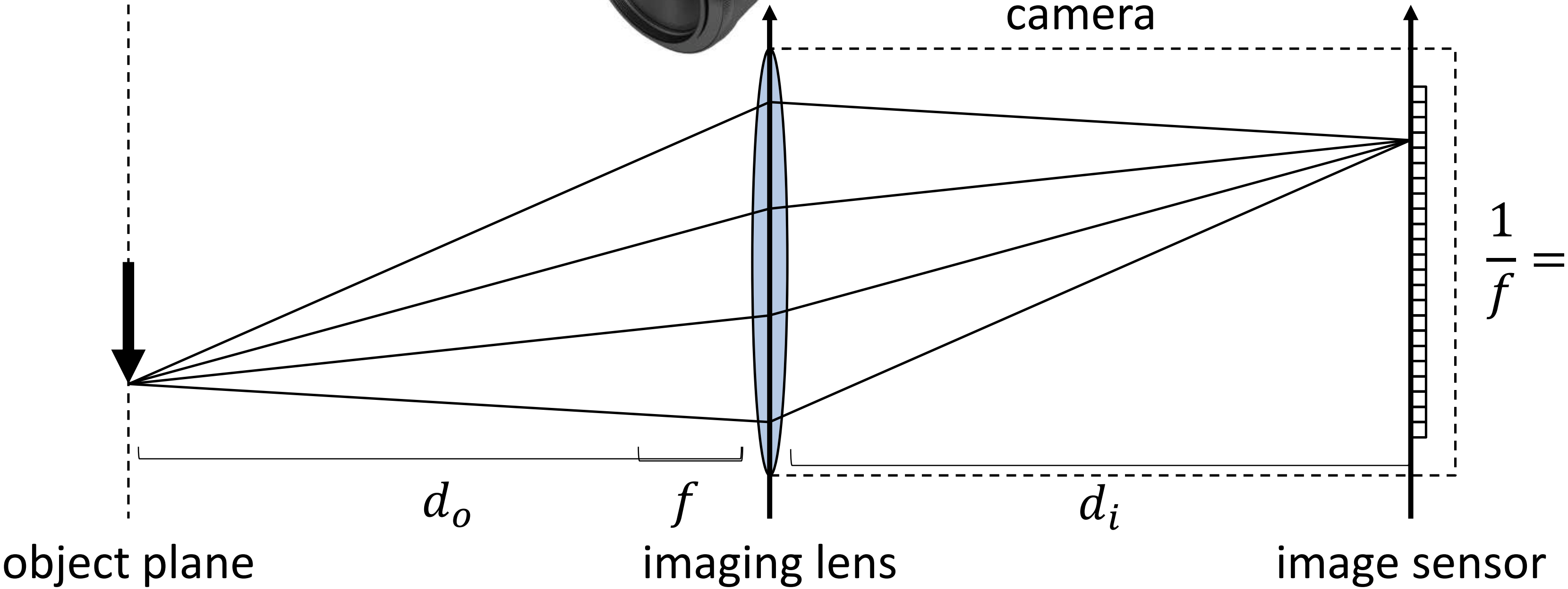


$$K = \begin{bmatrix} f_x & & x_0 \\ & f_y & y_0 \\ & & 1 \end{bmatrix} \qquad P = K \times [R \mid t] \qquad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \times \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Hartley, Richard, and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

Based on a single 2D camera

2D imaging: camera lens

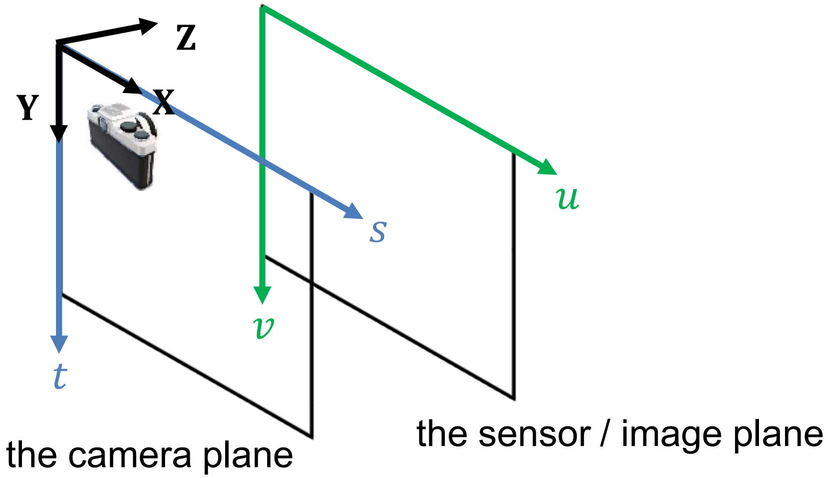


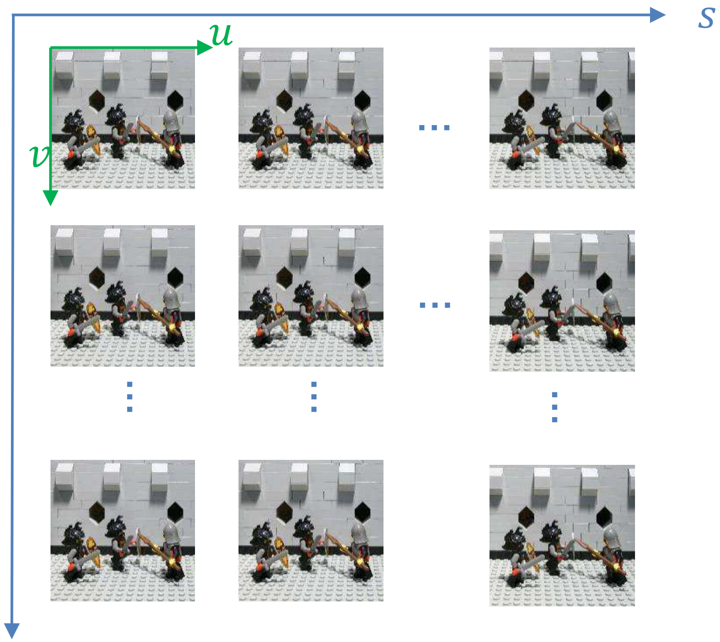
$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$$

Attribution: Weston Aenchbacher

Based on a single 2D camera

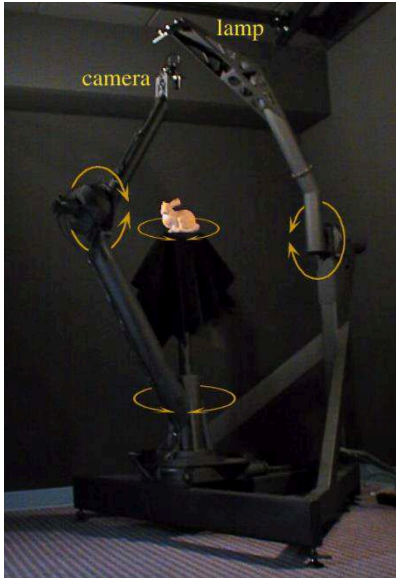
Light field capture with robotized camera





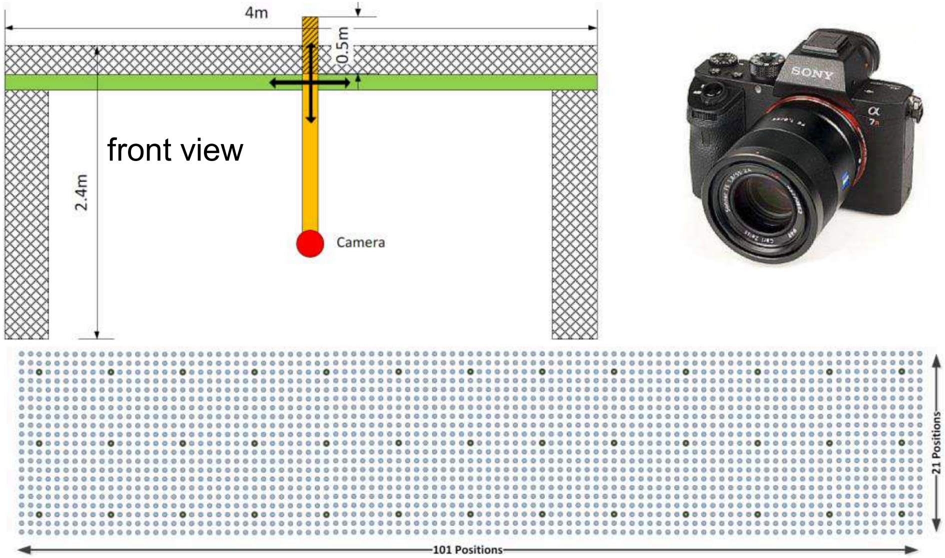
Based on a single 2D camera

Light field capture with robotized camera



Stanford, 2002

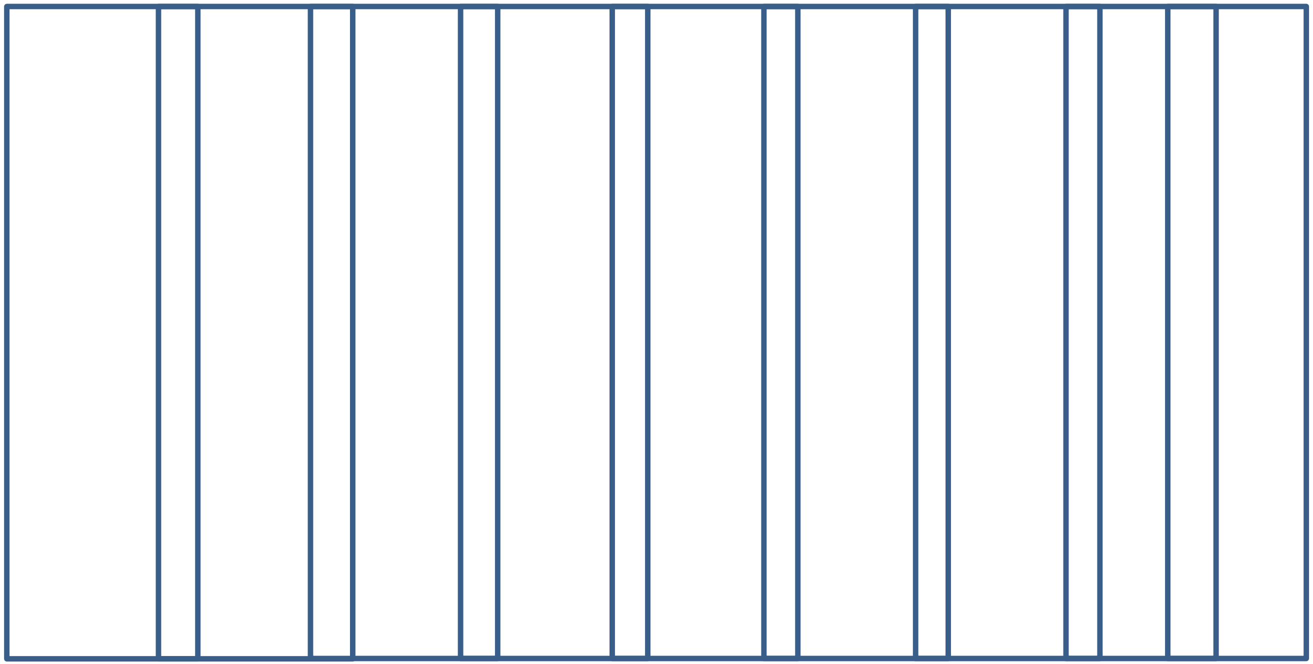
Stanford, 2008



Fraunhofer IIS , 2016

Based on a single 2D camera

Omnidirectional imaging capture with rotating camera



Stitching

Based on a single 2D camera

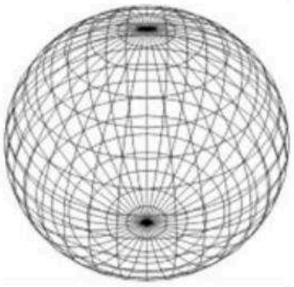
Omnidirectional imaging capture with rotating camera

- Stitching

1) Registration

2) Warping

3) Blending



Sphere

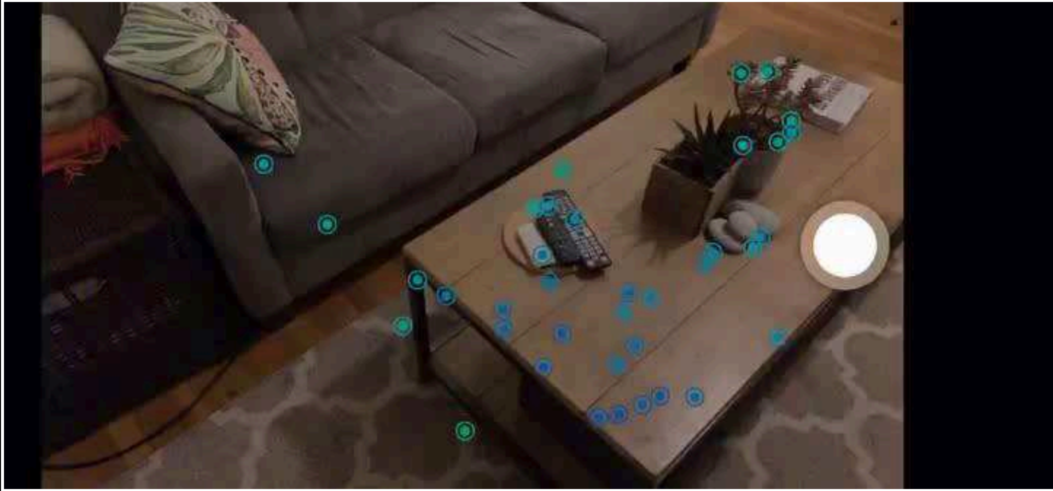
Attribution: Adrian Rosebrock

Based on a single 2D camera

Smartphone capture

- simultaneous localization and mapping (SLAM)

Light field



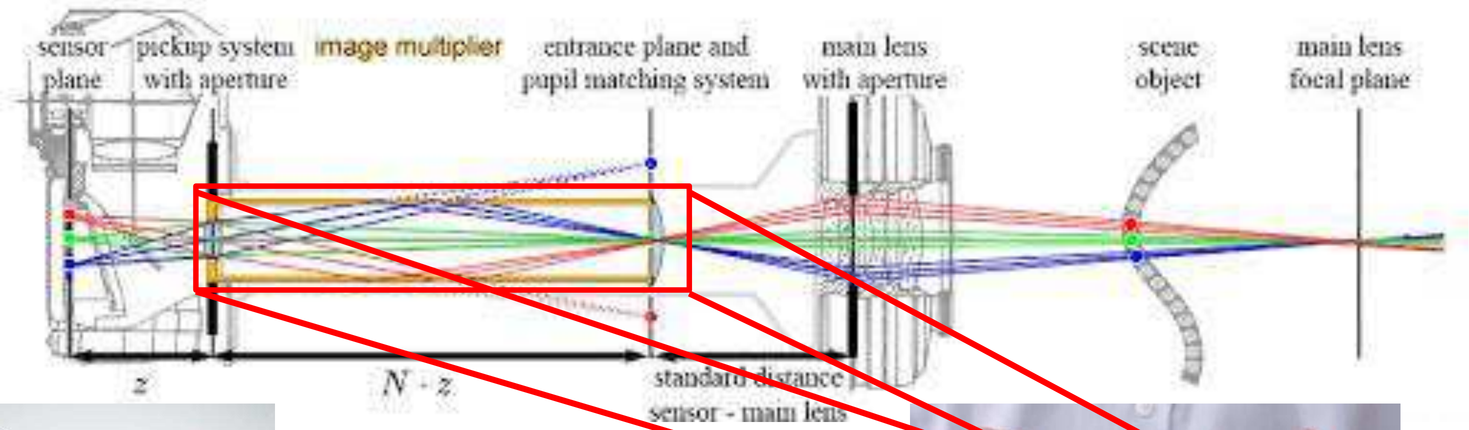
3D mesh model

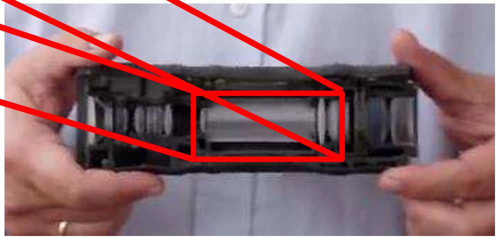
Birklbauer, Clemens, and Oliver Bimber. "**Active guidance for light-field photography on smartphones**." *Computers & Graphics* 53 (2015): 127-135.

Attribution: www.aboundlabs.com

Advanced camera design

Light field capture with a kaleidoscope lens

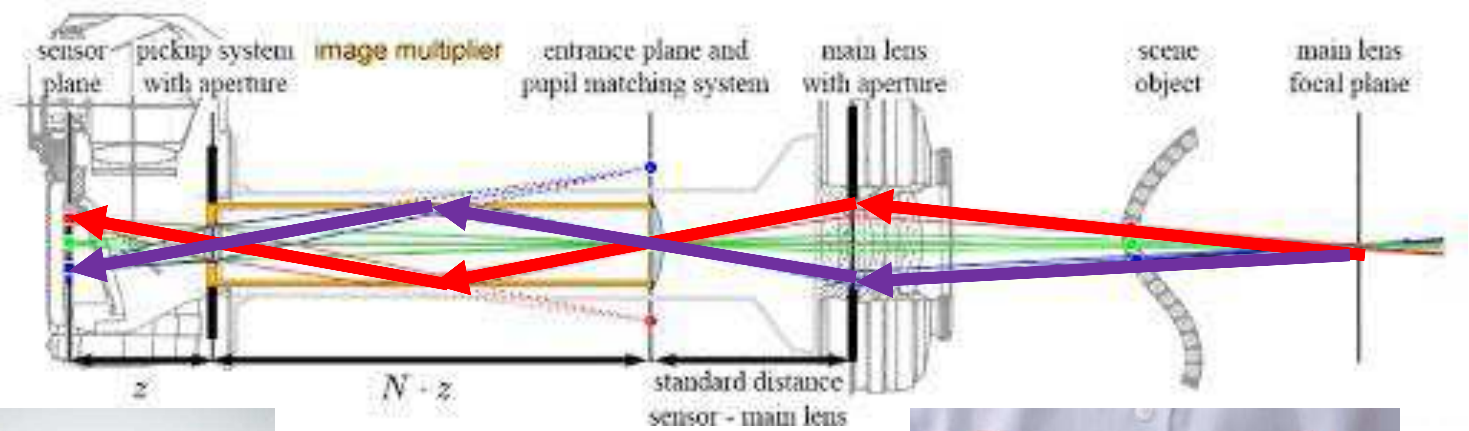




Manakov, Alkhazur, et al. "**A Reconfigurable Camera Add-On for High Dynamic Range, Multispectral, Polarization, and Light-Field Imaging**." *ACM Transactions on Graphics* 32.4 (2013): 47-1.

Advanced camera design

Light field capture with a kaleidoscope lens



Manakov, Alkhazur, et al. "**A Reconfigurable Camera Add-On for High Dynamic Range, Multispectral, Polarization, and Light-Field Imaging**." *ACM Transactions on Graphics* 32.4 (2013): 47-1.

Single-Camera Systems

Light field capture with a kaleidoscope lens

Sensor image

Manakov, Alkhazur, et al. "**A Reconfigurable Camera Add-On for High Dynamic Range, Multispectral, Polarization, and Light-Field Imaging**." *ACM Transactions on Graphics* 32.4 (2013): 47-1.

Advanced camera design

Light field capture with a kaleidoscope lens

Sensor image

Manakov, Alkhazur, et al. "**A Reconfigurable Camera Add-On for High Dynamic Range, Multispectral, Polarization, and Light-Field Imaging**." *ACM Transactions on Graphics* 32.4 (2013): 47-1.

Advanced camera design

Light field capture with lenslet plenoptic camera

Lytro Illum
Plenoptic 1.0

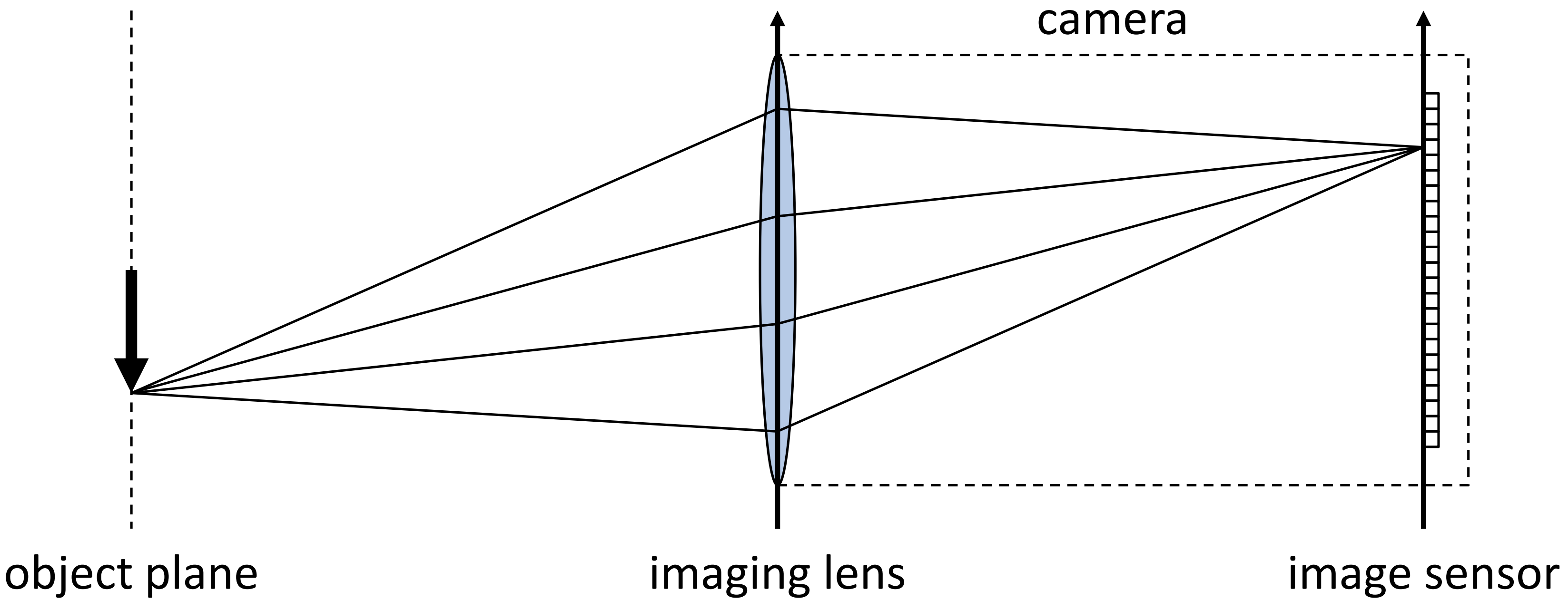
Raytrix
Plenoptic 2.0

Ng, Ren. *Digital light field photography*. Stanford: stanford university, 2006.

Perwass, Christian, and Lennart Wietzke. "**Single lens 3D-camera with extended depth-of-field**." *Human Vision and Electronic Imaging XVII*. Vol. 8291. International Society for Optics and Photonics, 2012.

Advanced camera design

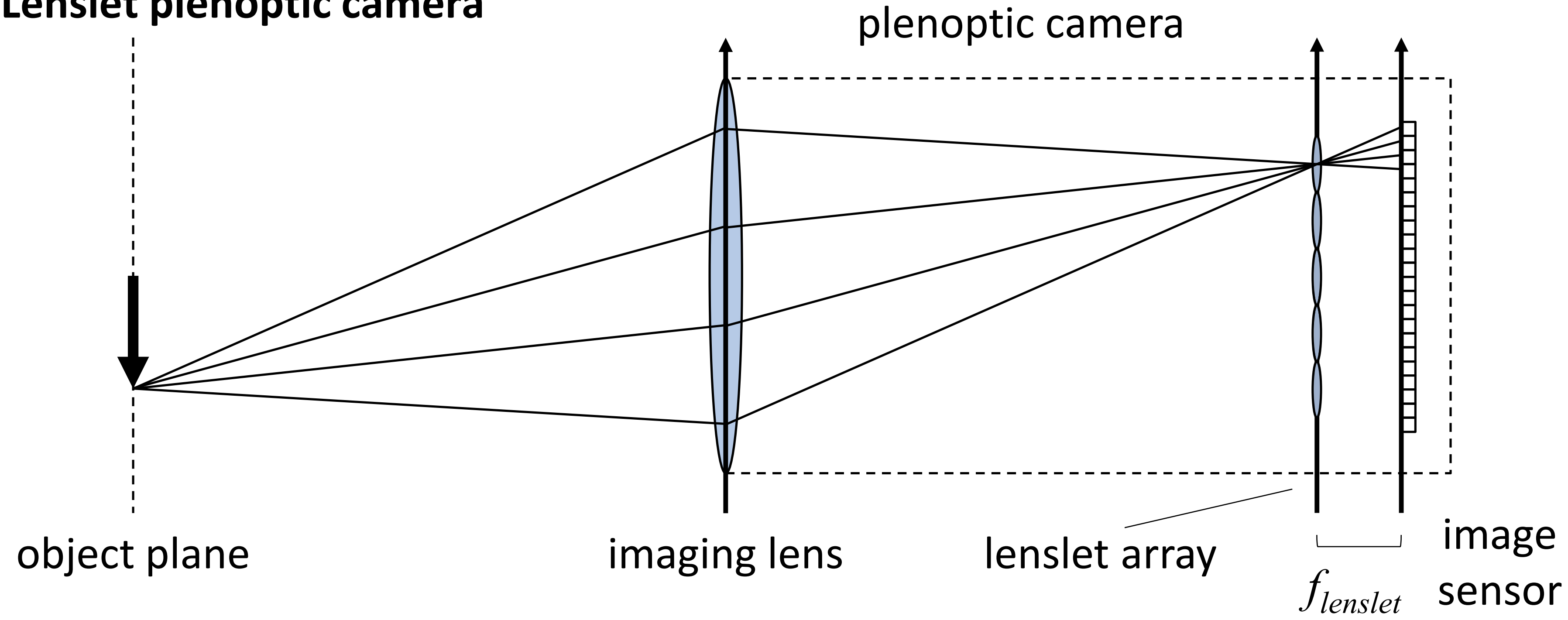
Lenslet plenoptic camera



Attribution: Weston Aenchbacher

Advanced camera design

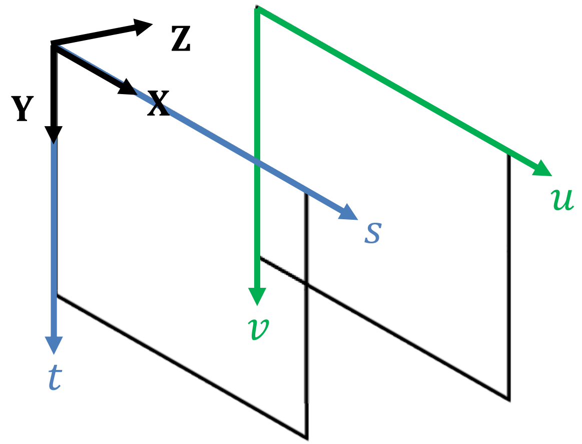
Lenslet plenoptic camera

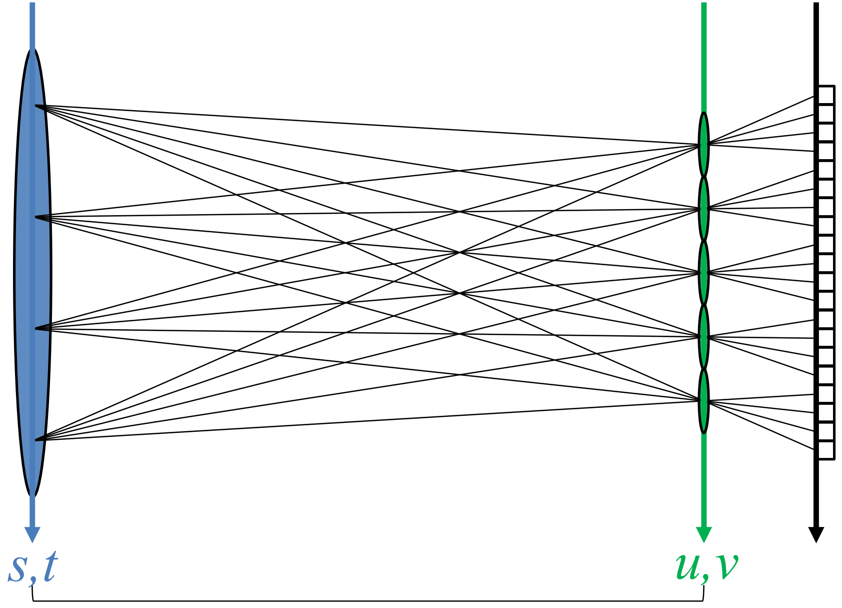


Attribution: Weston Aenchbacher

Advanced camera design

Lenslet plenoptic camera

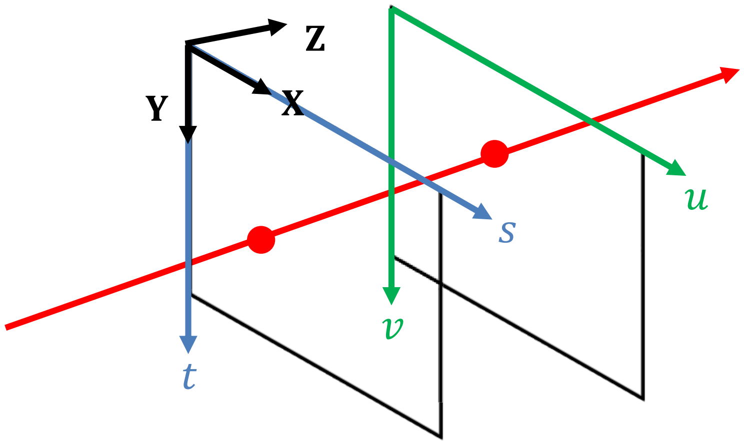




Attribution: Weston Aenchbacher

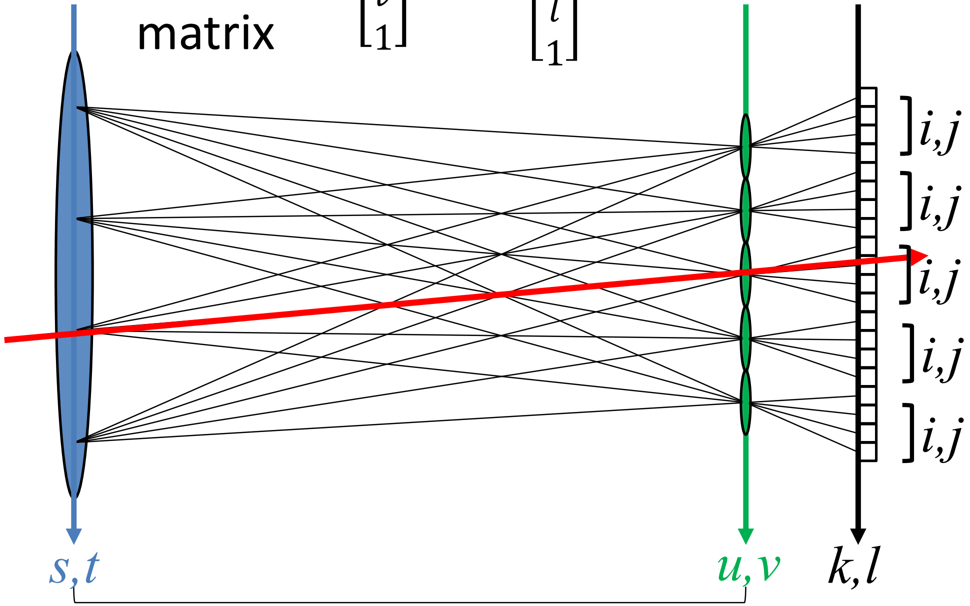
Advanced camera design

Lenslet plenoptic camera



Plenoptic intrinsics matrix

$$\begin{bmatrix} s \\ t \\ u \\ v \\ 1 \end{bmatrix} = H \times \begin{bmatrix} i \\ j \\ k \\ l \\ 1 \end{bmatrix}$$



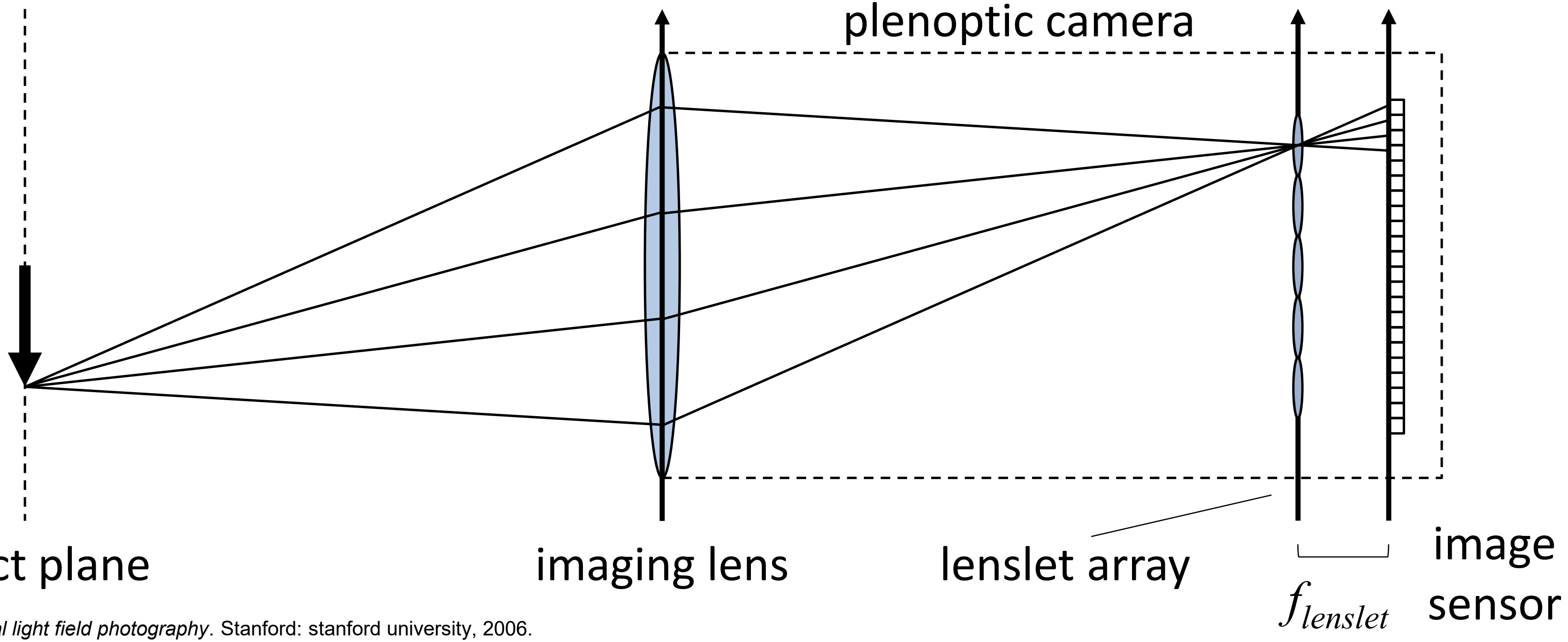
light field $L(s, t, u, v)$

Dansereau, Donald G., Oscar Pizarro, and Stefan B. Williams. "**Decoding, calibration and rectification for lenselet-based plenoptic cameras**." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013.

Attribution: Weston Aenchbacher

Advanced camera design

Lenslet plenoptic camera: plenoptic 1.0

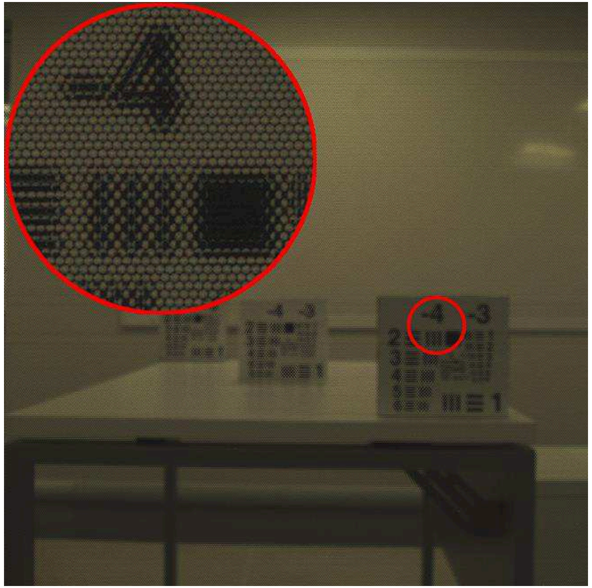


Ng, Ren. *Digital light field photography*. Stanford: stanford university, 2006.

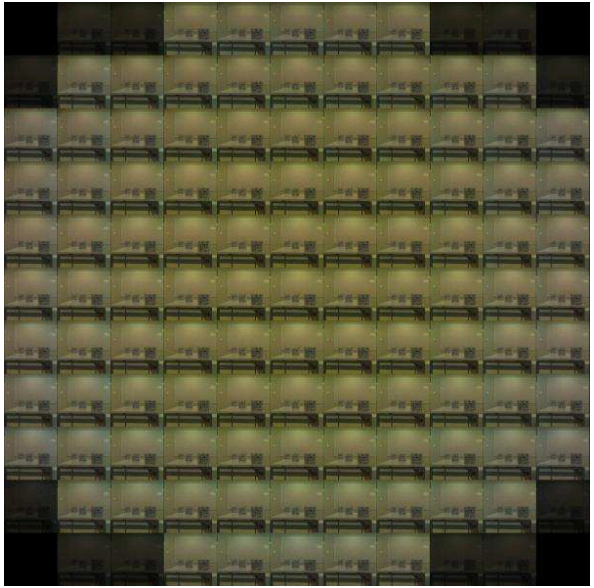
Attribution: Weston Aenchbacher

Advanced camera design

Lenslet plenoptic camera: plenoptic 1.0



Raw image

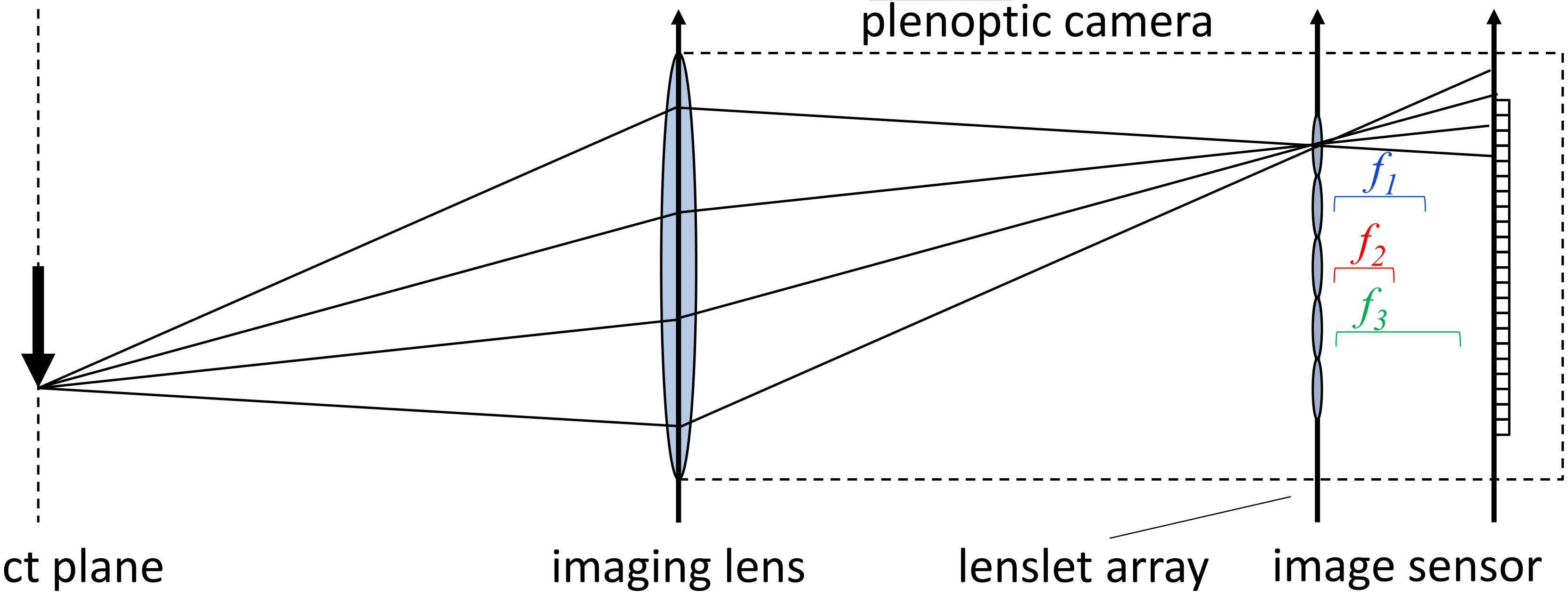


Light field views

Attribution: Weston Aenchbacher

Advanced camera design

Lenslet plenoptic camera: plenoptic 2.0



Perwass, Christian, and Lennart Wietzke. "**Single lens 3D-camera with extended depth-of-field**." *Human Vision and Electronic Imaging XVII*. Vol. 8291. International Society for Optics and Photonics, 2012.

Attribution: Weston Aenchbacher

Advanced camera design

Lenslet plenoptic camera: plenoptic 2.0

Raw video

Light field views

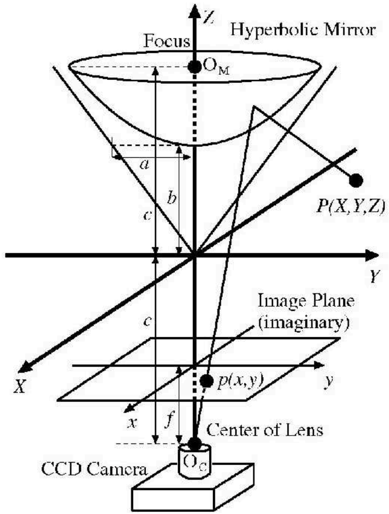
Center view

Guillo, Laurent, et al. "**Light field video dataset captured by a R8 Raytrix camera (with disparity maps)**." ISO /IEC JTC1/SC29/WG11 MPEG2018/m42468, ISO/IEC JTC1/SC29/WG1 JPEG2018/m79046, INTERNATIONAL ORGANISATION FOR STANDARDISATION,ISO/IEC JTC1/SC29/WG1 & WG11, April 2018, San Diego, CA, US

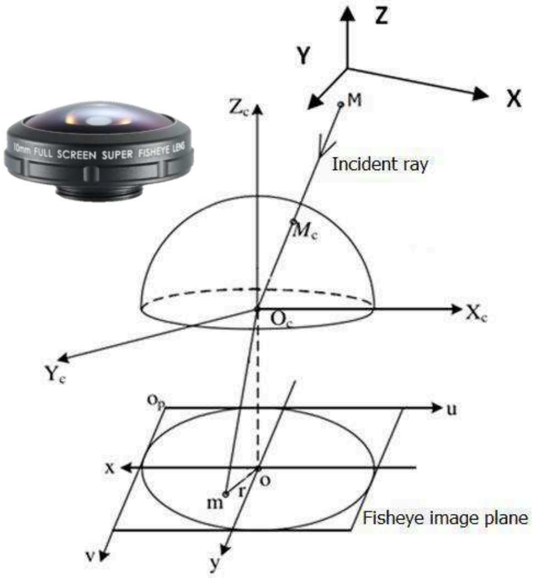
Advanced camera design

Omnidirectional imaging

- Mirror-based
  - FOV > 180°
- Fisheye / Dual fisheye
  - FOV ~ 180° for a single fisheye



Ukida, Hiroyuki, et al. "**Omni-directional 3d measurement by hyperbolic mirror cameras and pattern projection.**" *2008 IEEE Instrumentation and Measurement Technology Conference*. IEEE, 2008.



Chan, Sixian, et al. "**An improved method for fisheye camera calibration and distortion correction.**" *2016 International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2016.

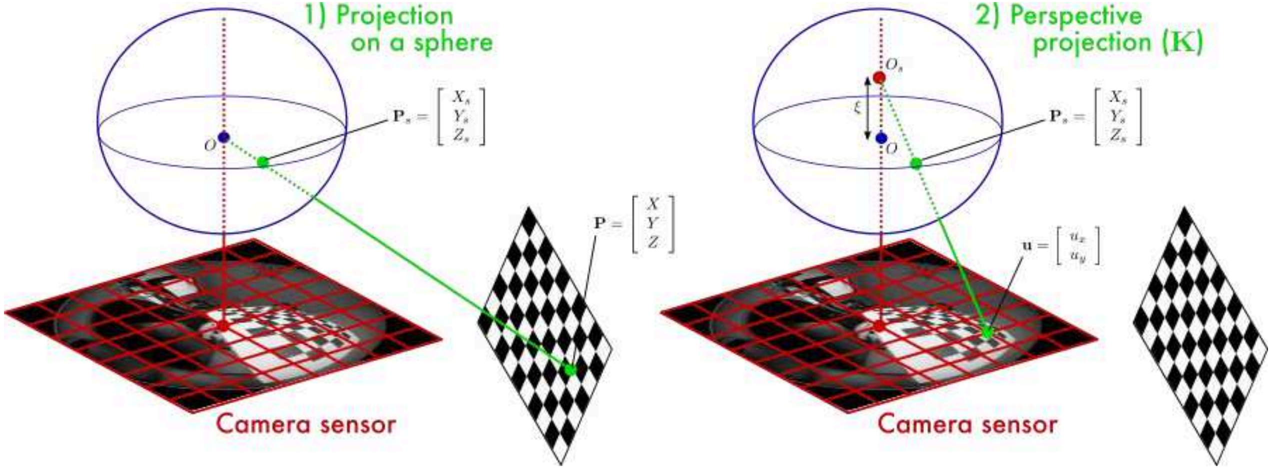
Advanced camera design

Omnidirectional imaging

- Unified Spherical model
  - Projection on a sphere followed by perspective projection

$$P_s = \frac{P}{\|P\|}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \times (P_s + \begin{bmatrix} 0 \\ 0 \\ \xi \end{bmatrix})$$



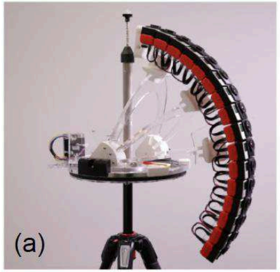
Mei, Christopher, and Patrick Rives. "**Single view point omnidirectional camera calibration from planar grids**." *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007.

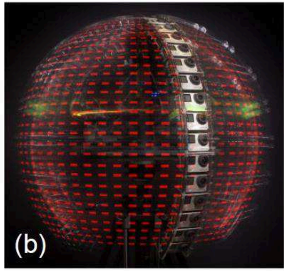
Maugey, Thomas, Laurent Guillo, and Cedric Le Cam. "**FTV360: a multiview 360° video dataset with calibration parameters**." *Proceedings of the 10th ACM Multimedia Systems Conference*. 2019.

Multi-Camera Systems

Imaging with multiple cameras

- Light fields
  - Planar setup
  - Panoramic setup
- Omnidirectional imaging
  - Radial setups
  - Spherical setups
- Volumetric video (3D video)
  - Studio setups

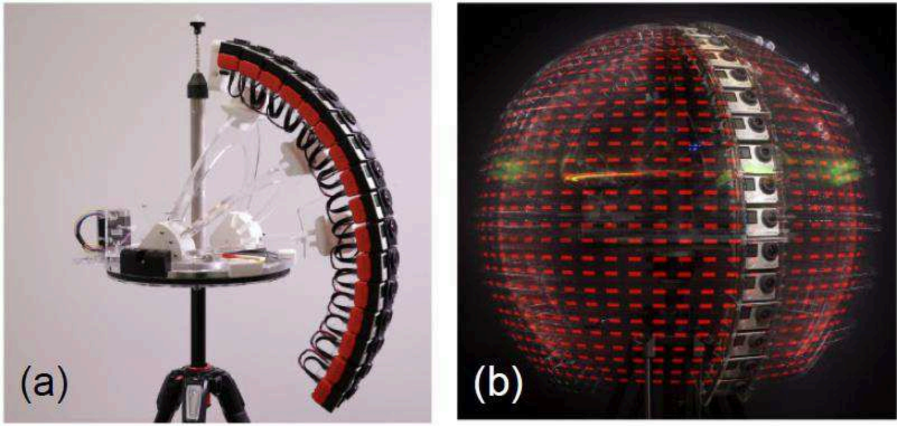




Multi-Camera Systems

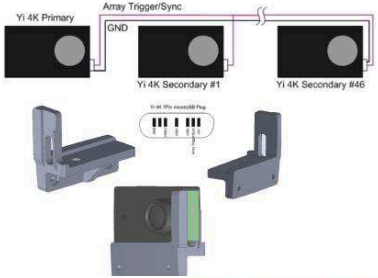
Towards spherical light field capture

- Cameras positioned on a rotating arc
- Full spherical light field is captured, but limited to static scenes



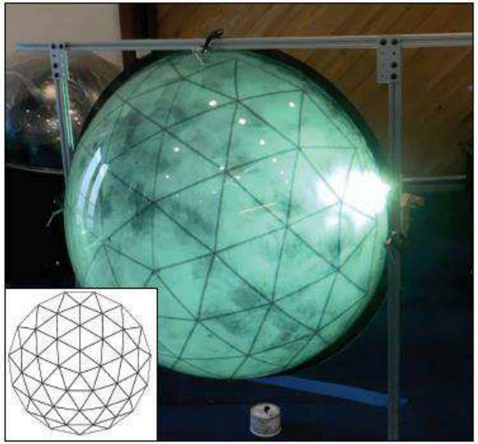
Overbeck, Ryan S., et al. "**A system for acquiring, processing, and rendering panoramic light field stills for virtual reality**." *ACM Transactions on Graphics (TOG)* 37.6 (2018): 1-15.

Multi-Camera Systems



Towards spherical light field capture

- Cameras positioned on a portion of a sphere



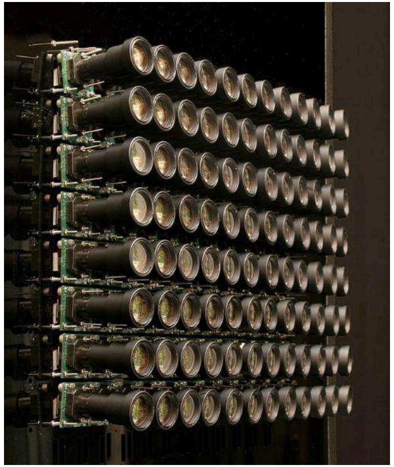
Broxton, Michael, et al. "**A Low Cost Multi-Camera Array for Panoramic Light Field Video Capture.**" *SIGGRAPH Asia 2019 Posters*. 2019. 1-2.

Broxton, Michael, et al. "**Immersive light field video with a layered mesh representation.**" *ACM Transactions on Graphics (TOG)* 39.4 (2020): 86-1.

Multi-Camera Systems

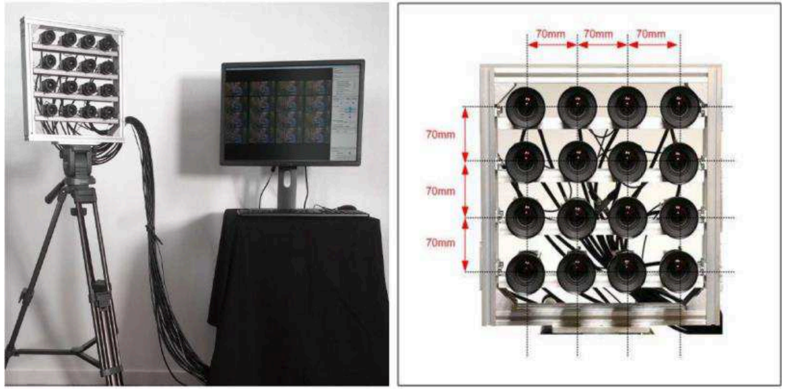
Planar light field capture

- Camera arrays



Stanford, 2004

Wilburn, Bennett. "**High-performance imaging using arrays of inexpensive cameras**." (2005)



Technicolor, 2017

Sabater, Neus, et al. "**Dataset and Pipeline for Multi-view Light-Field Video**." *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017.

Saarland university, 2018

Attribution: www.sauceproject.eu/

Multi-Camera Systems

Omnidirectional imaging capture

- Cameras are placed to cover the whole visual angle of the unit sphere
- Enables video capture of the whole sphere

Google jump

Facebook Surround 360

Multi-Camera Systems

3D models & Volumetric video

- Studio setups
  - RGB cameras
  - + IR cameras

Schreer, Oliver, et al. "**Capture and 3D Video Processing of Volumetric Video**." *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., ... & Sullivan, S. (2015). **High-quality streamable free-viewpoint video**. *ACM Transactions on Graphics (ToG), 34(4)*, 69.

Pagés, Rafael, et al. "**Affordable content creation for free-viewpoint video and VR/AR applications**." *Journal of Visual Communication and Image Representation* 53 (2018): 192-201.

Multi-Camera Systems

3D models & Volumetric video

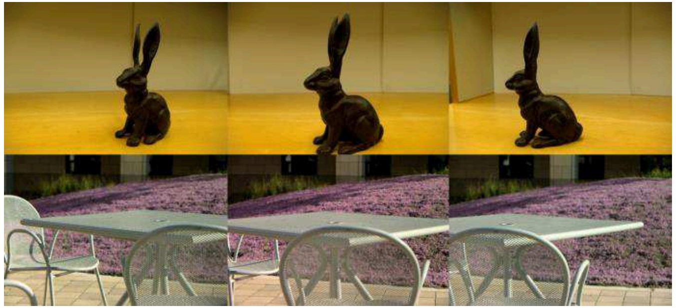
- Studio setups
  - RGB cameras
  - + IR cameras
- Mobile solution

Pagés, Rafael, et al. "**Affordable content creation for free-viewpoint video and VR/AR applications**." *Journal of Visual Communication and Image Representation* 53 (2018): 192-201.

Multi-Camera Systems

Combining advanced cameras

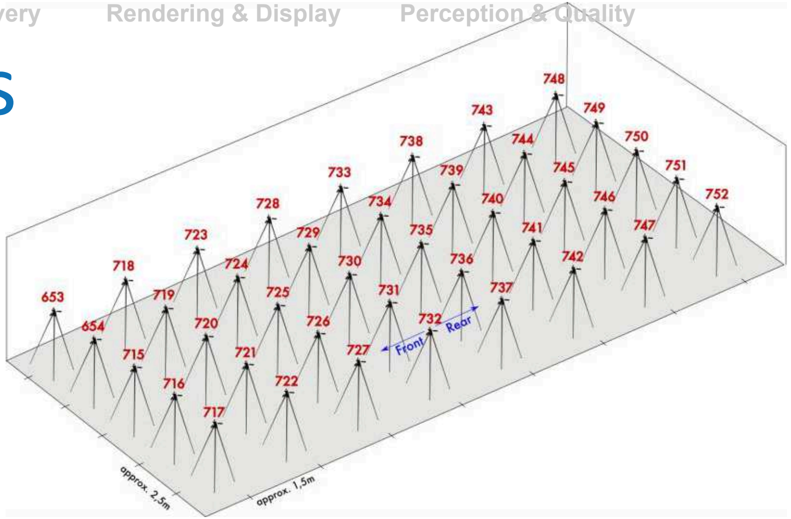
- Plenoptic multiview
- Plenoptic panorama



Johannsen, Ole, Antonin Sulc, and Bastian Goldluecke. "**On linear structure from motion for light field cameras**." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

Dansereau, Donald G., Bernd Girod, and Gordon Wetzstein. "**LiFF: Light Field Features in Scale and Depth**." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019

Multi-Camera Systems



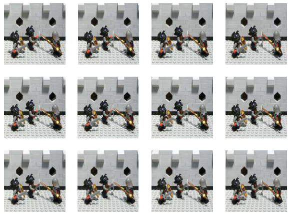
Combining advanced cameras

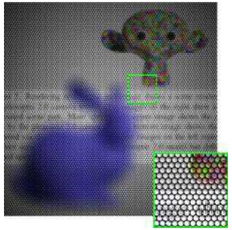
- Grid of omnidirectional cameras
- Allows 6DoF

Maugey, Thomas, Laurent Guillo, and Cedric Le Cam. "**FTV360: a multiview 360° video dataset with calibration parameters**." *Proceedings of the 10th ACM Multimedia Systems Conference*. 2019.

Attribution: project.inria.fr/ftv360/

Data Formats and Representations





Light fields

- Collection of images

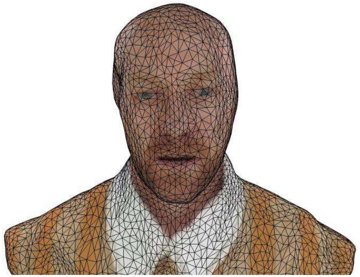
Image-based

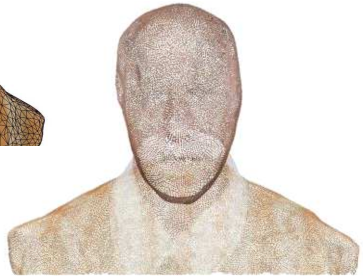
Omnidirectional imaging

- Projections from sphere to plane

3D models & Volumetric video

- Textured meshes
- Point clouds



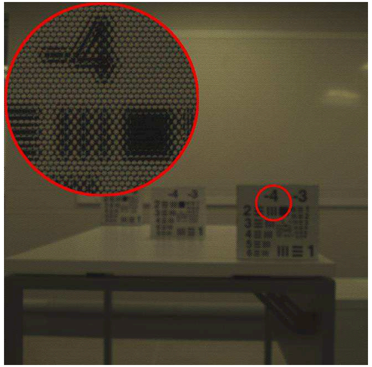


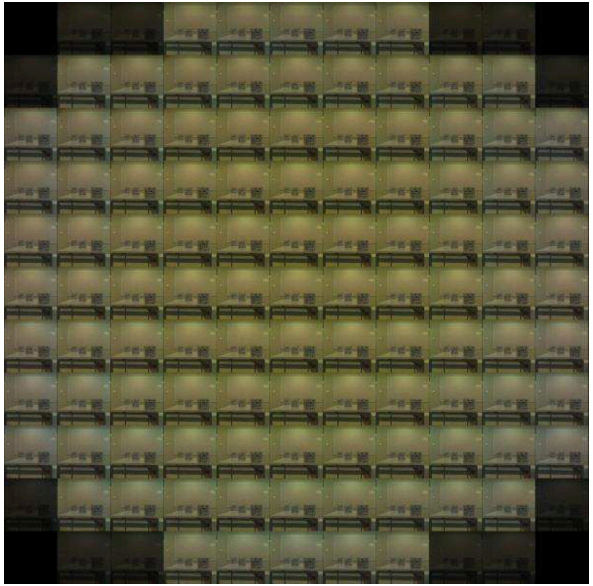
Geometry-based

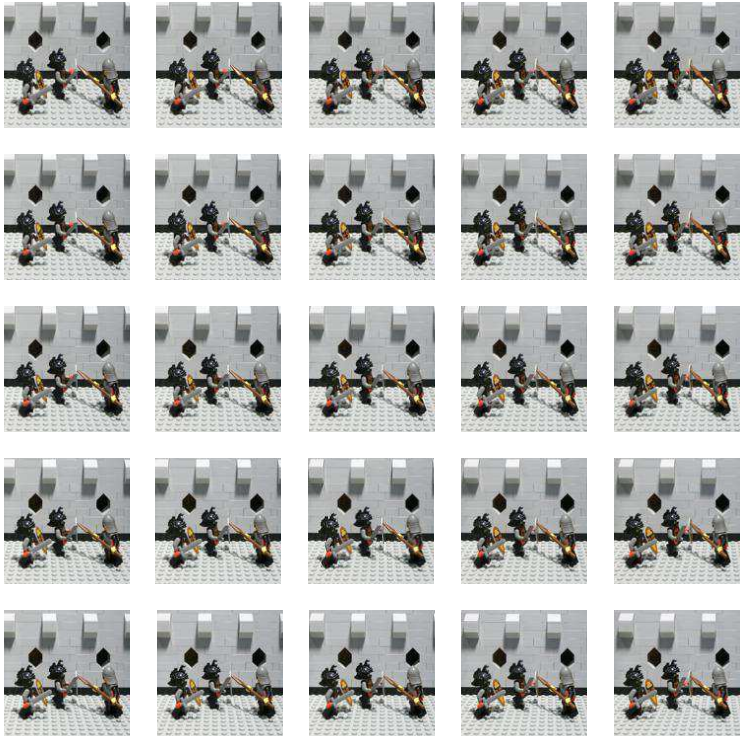
Data Formats and Representations

Light fields

- Collection of images
- Raw image from plenoptic cameras
- + metadata



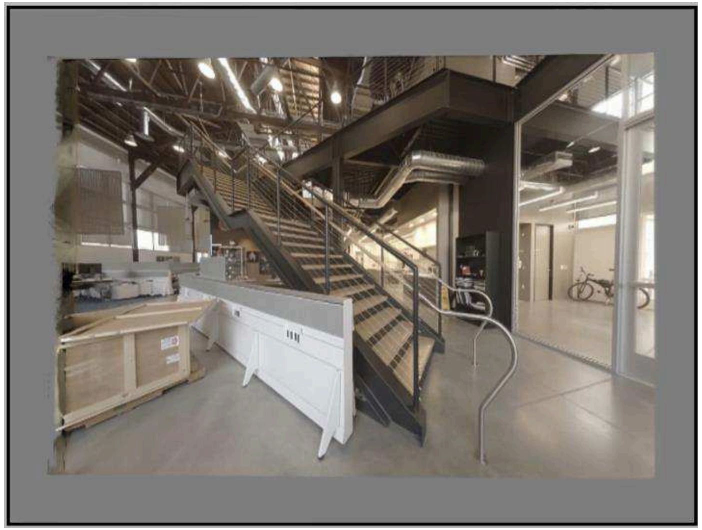




Data Formats and Representations

Light fields

- Multi-Plane Images (MPIs)

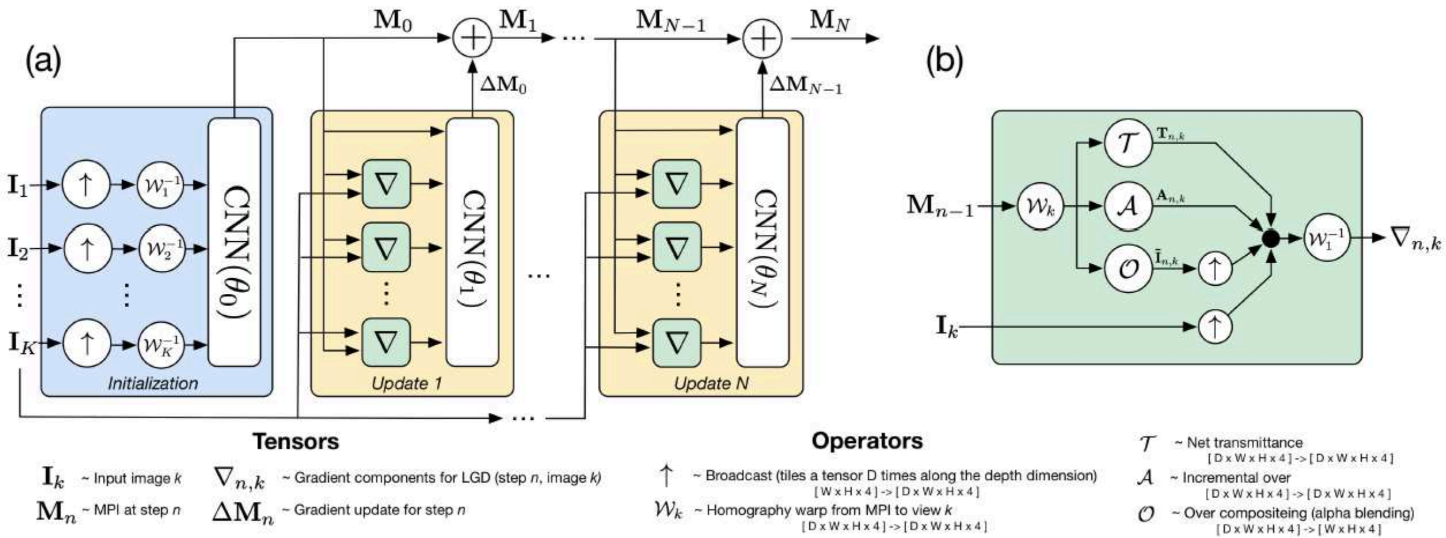



Zhou, Tinghui, et al. "**Stereo magnification: learning view synthesis using multiplane images.**" *ACM Transactions on Graphics (TOG)* 37.4 (2018): 1-12.
Flynn, John, et al. "**Deepview: View synthesis with learned gradient descent.**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
Srinivasan, Pratul P., et al. "**Pushing the boundaries of view extrapolation with multiplane images.**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
Mildenhall, Ben, et al. "**Local light field fusion: Practical view synthesis with prescriptive sampling guidelines**." *ACM Transactions on Graphics (TOG)* 38.4 (2019): 1-14.

Data Formats and Representations

Light fields

- Multi-Plane Images (MPIs)

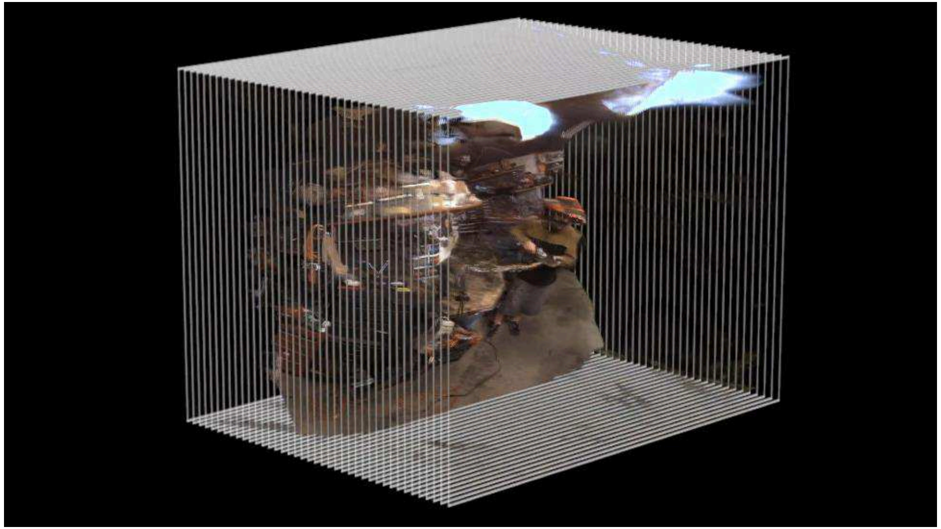


Flynn, John, et al. "**Deepview: View synthesis with learned gradient descent.**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

Data Formats and Representations

Light fields

- Multi-Plane Images (MPIs)
- Multi-Sphere Images (MSIs)

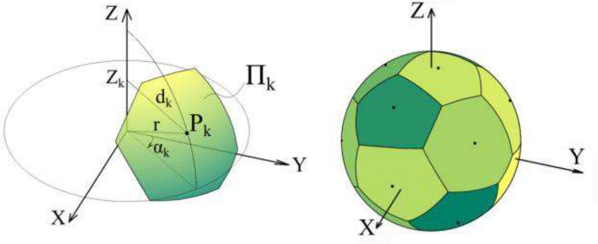


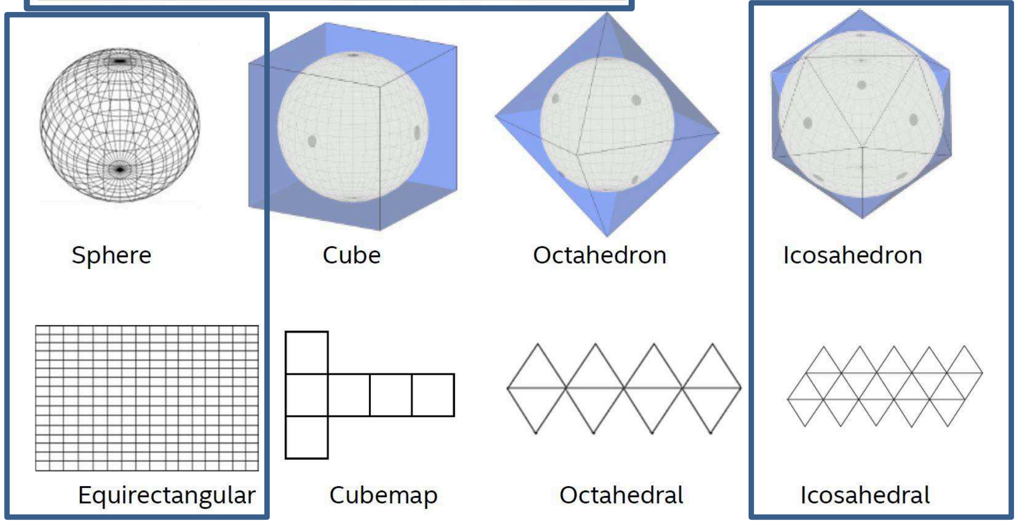
Broxton, Michael, et al. "**Immersive light field video with a layered mesh representation.**" *ACM Transactions on Graphics (TOG)* 39.4 (2020): 86-1.

Data Formats and Representations

Omnidirectional imaging

- Projections from sphere to plane
- Different projection methods
    - Equirectangular projection
    - Cubemap projection
    - Octahedron projection
    - Isocahedron projection
    - Custom projections
        - Voronoi cells

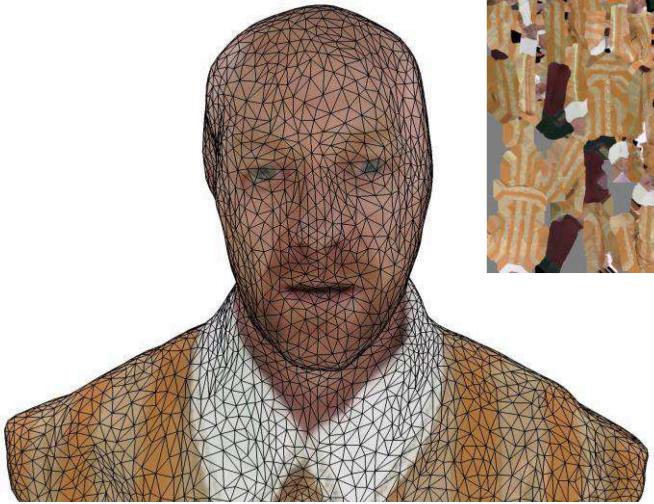




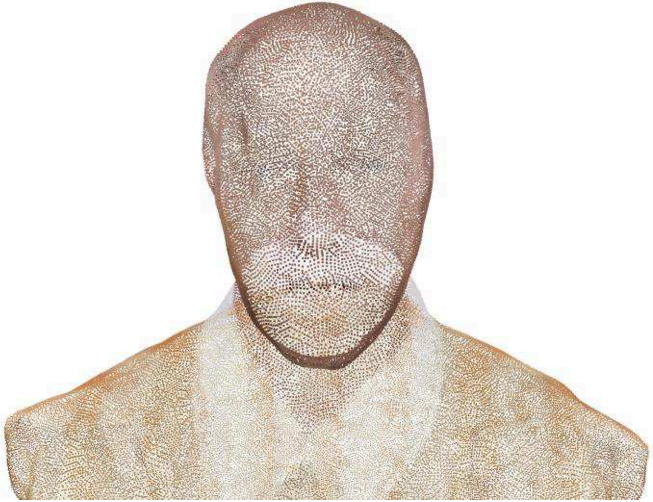
Boyce, Jill, "**Video codec standardization update for 360 degree video**", *IEEE Signal Processing Society Santa Clara Valley Chapter 2017*

Data Formats and Representations

Textured polygonal meshes
- Vertices and Faces
- Texture atlas

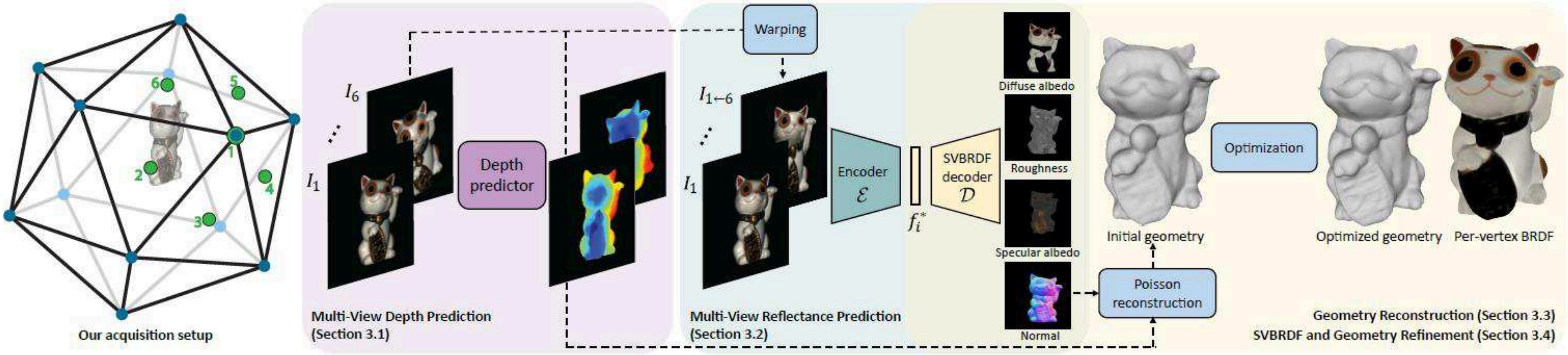


Colour point clouds
- Points (3D coordinates)
- Attributes (e.g., colour, normal, etc.)



Learning-based methods

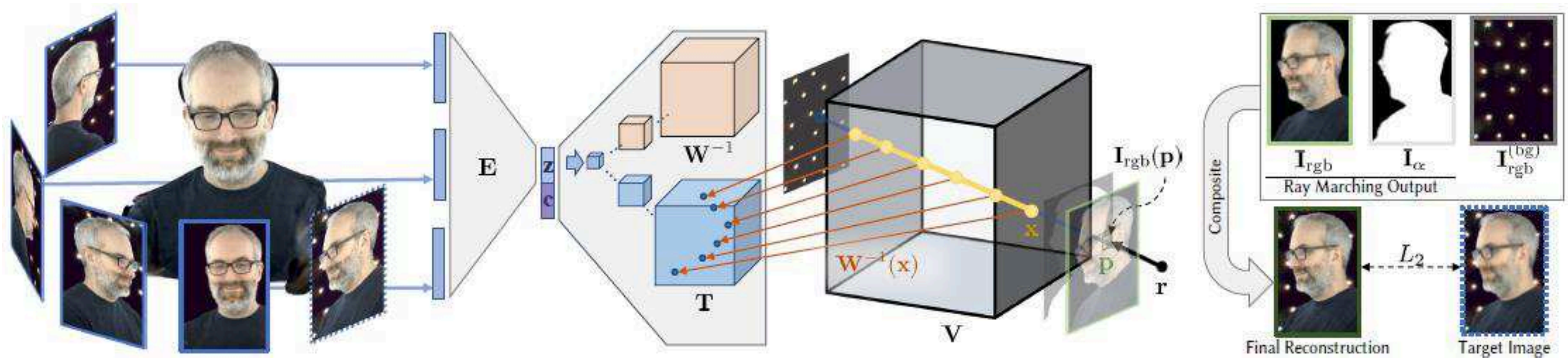
Deep 3D capture



Bi, Sai, et al. "**Deep 3D Capture: Geometry and Reflectance from Sparse Multi-View Images**." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

Learning-based methods

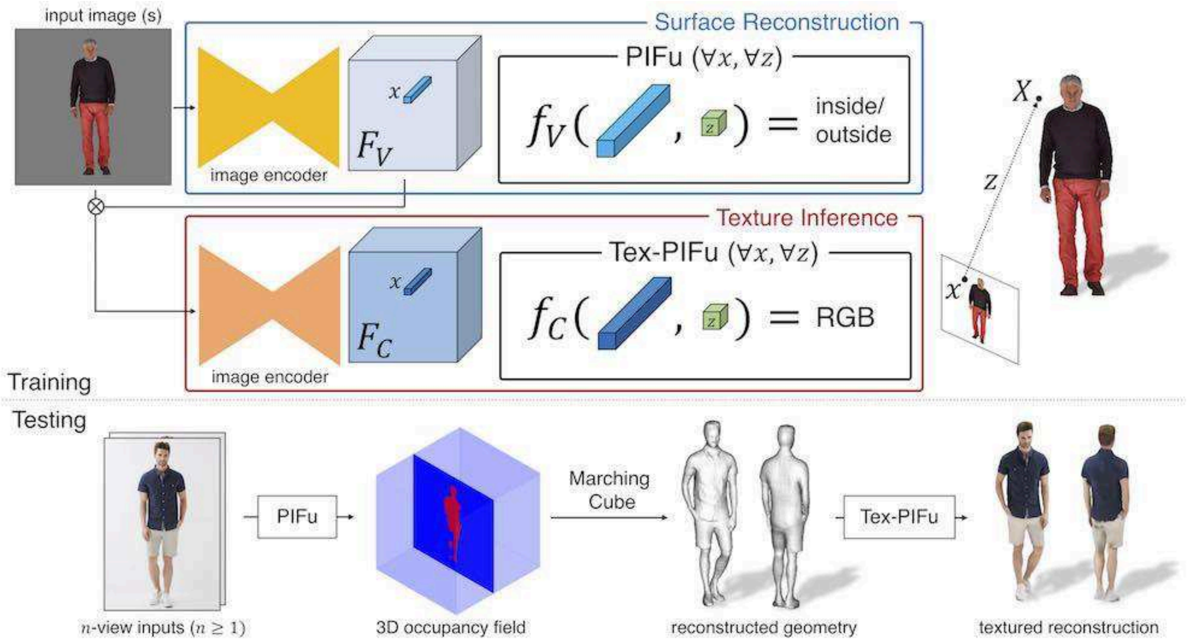
Neural volumes



Lombardi, Stephen, et al. "**Neural volumes: Learning dynamic renderable volumes from images.**" *ACM Transactions on Graphics (TOG)* 38.4 (2019): 65.

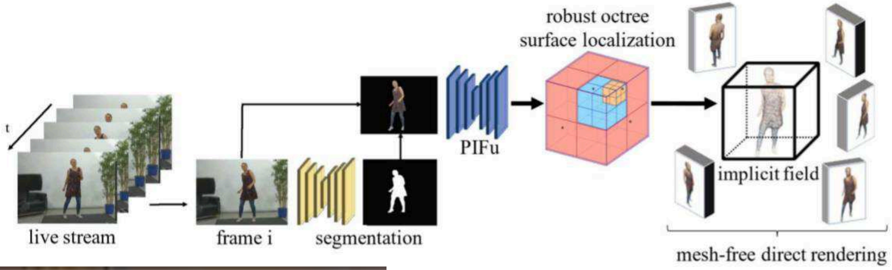
Learning-based methods

Pixel-Aligned Implicit Function (PiFU)



Saito Shunsuke, Huang Zeng, Natsume Ryota, et al. "**PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization**." In *International Conference on Computer Vision* 2019.

Learning-based methods



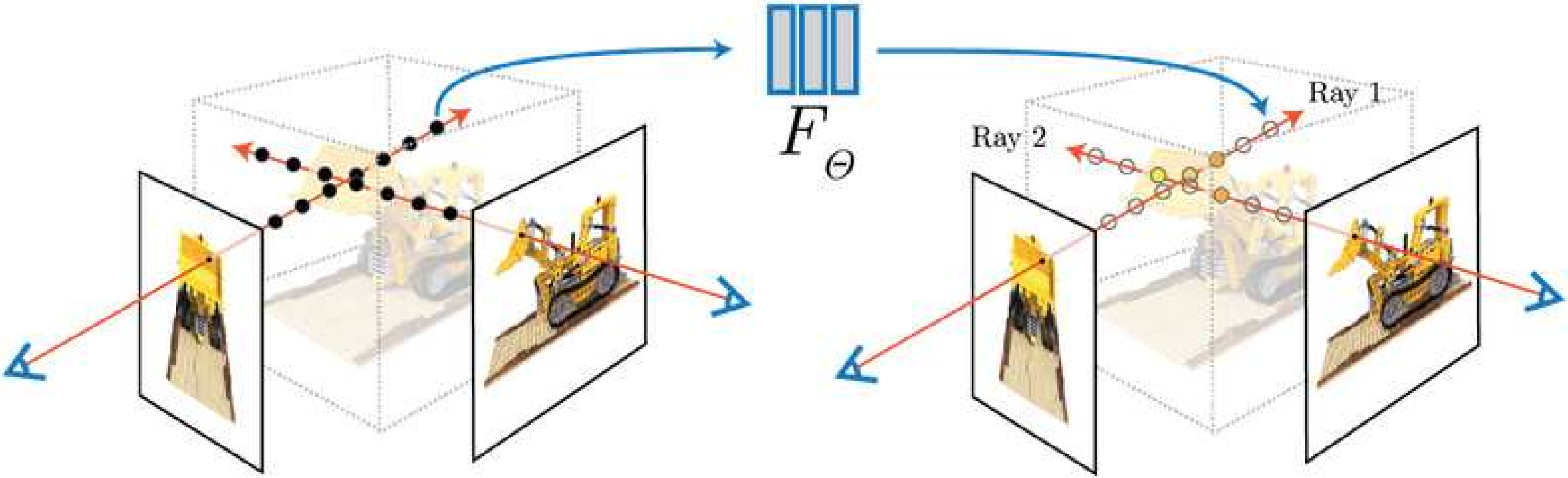
Monocular Real-Time Volumetric Capture

- PiFU-based system
- Texture atlas

R. Li, Y. Xiu, S. Saito, Z. Huang, K. Olszewski, & H. Li. "**Monocular Real-Time Volumetric Performance Capture**". In *European Conference on Computer Vision*, pp. 49-67, August 2020

Learning-based methods

Neural Radiance Fields (NeRF)



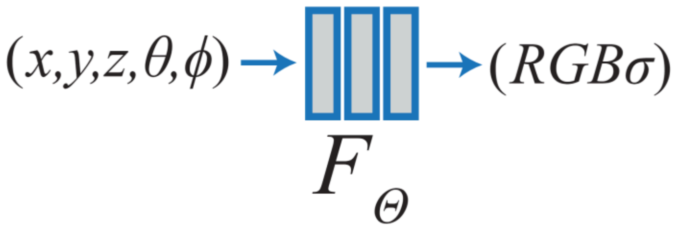
Mildenhall, Ben, et al. "**NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis**." In *European Conference on Computer Vision,* August 2020.

Learning-based methods

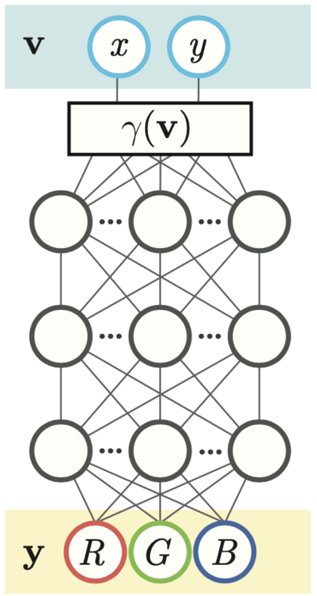
Neural Radiance Fields (NeRF)

Mildenhall, Ben, et al. "**NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis**." In *European Conference on Computer Vision,* August 2020.

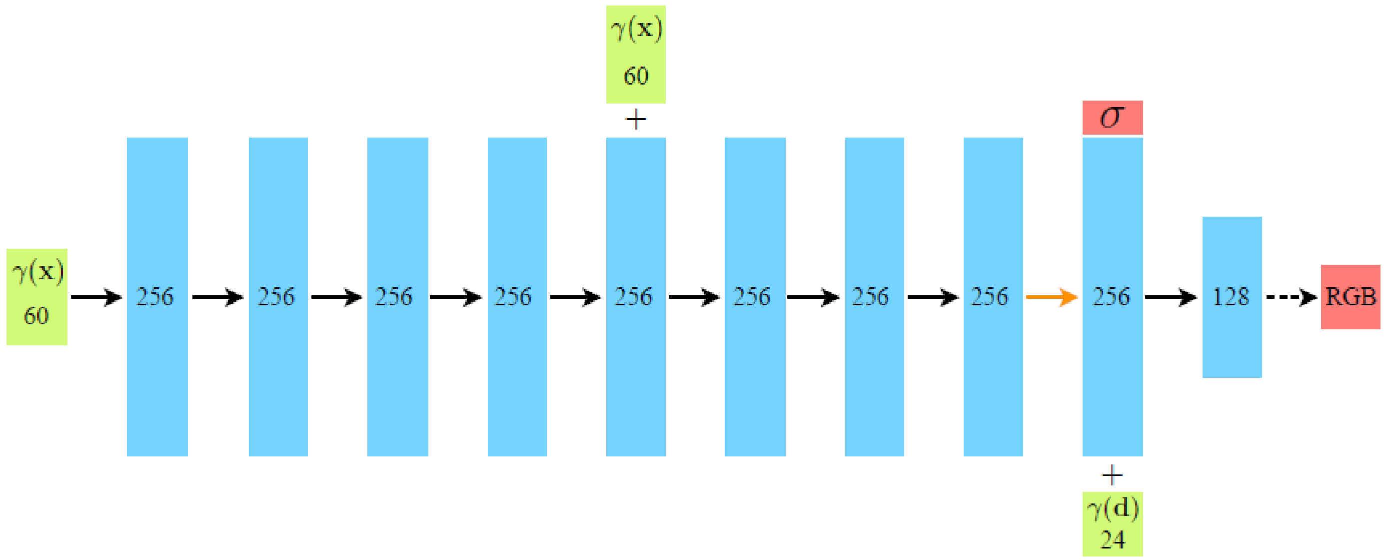
Learning-based methods



Neural Radiance Fields (NeRF)



(a) Coordinate-based MLP



Mildenhall, Ben, et al. "**NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis**." In *European Conference on Computer Vision,* August 2020.

Summary: Acquisition & Data Format

Key concepts of immersive imaging

- Single-camera systems
  - Using traditional 2D cameras
  - Specialized cameras
- Multi-camera systems

Data Formats

- Light field images
- Omnidirectional images
- Textured meshes and point clouds

Next part:

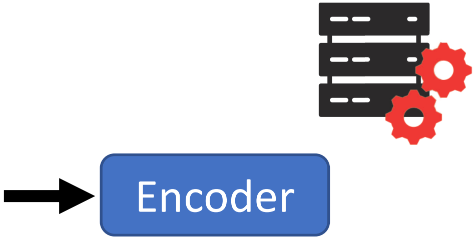
- How to encode and transmit all this data?

Part III: Content Delivery

- How is the acquired immersive content delivered? What are the compression methods? What are the streaming methods?

Part III: Content Delivery

Video streaming

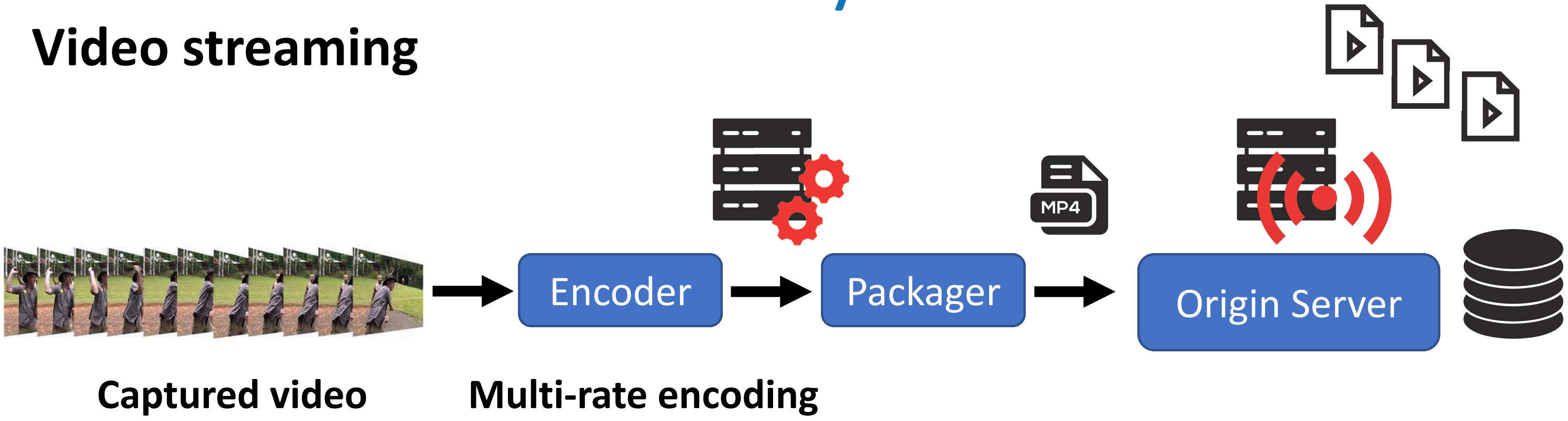


Captured video

Multi-rate encoding

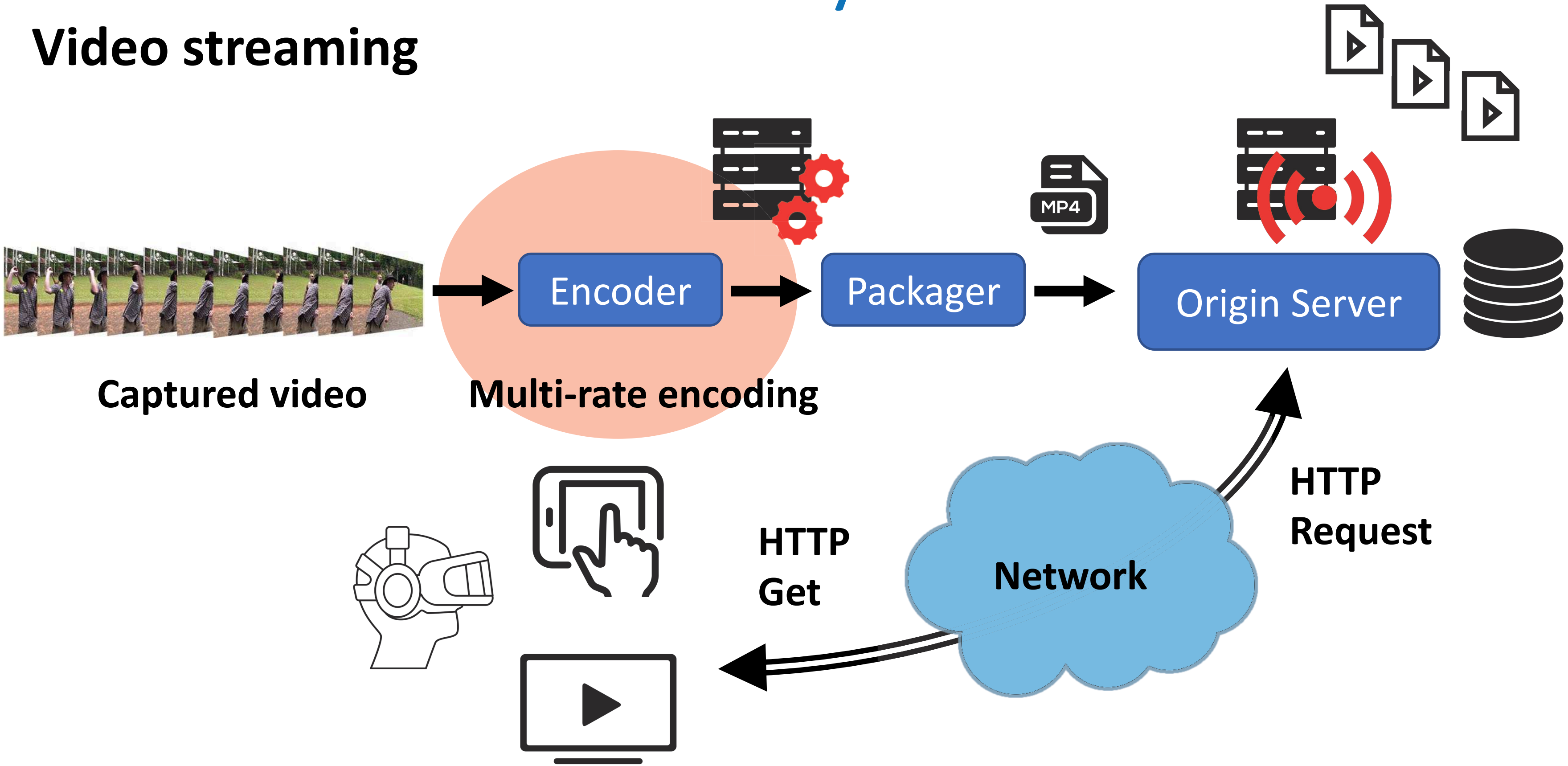
Part III: Content Delivery

Video streaming



Part III: Content Delivery

Video streaming



Part III: Content Delivery

Video coding

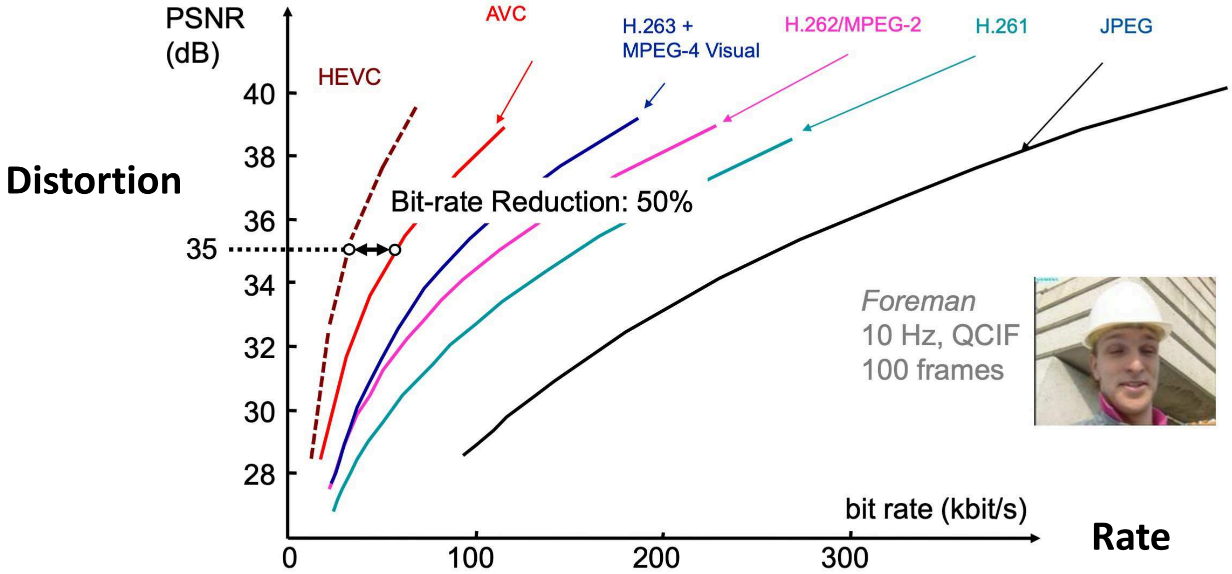
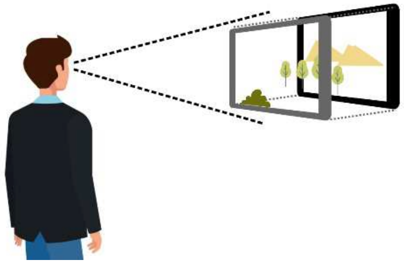


Figure: Rate-distortion curve for various video coding standards

* Figure has been taken from Jens-Rainer Ohm and Mathias Wien' s tutorial slides at ICME 2018: Trends and Recent Developments in Video Coding Standardization

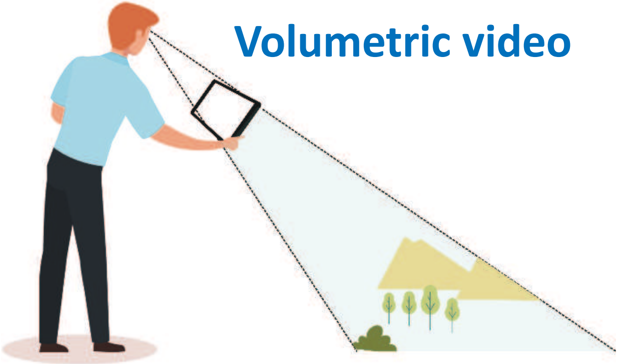
Part III: Content Delivery

Light Fields

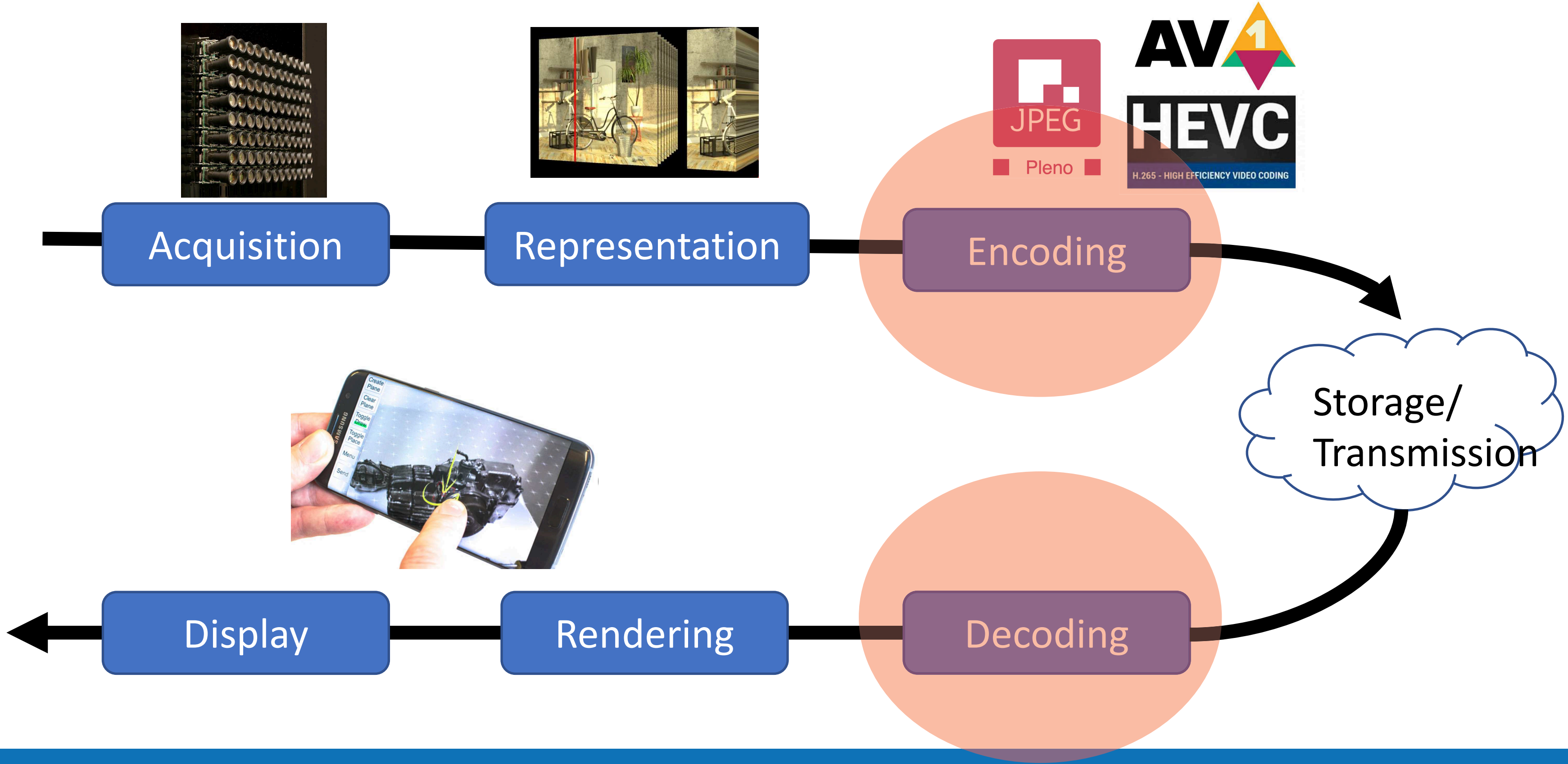


360-degree video

Volumetric video



Delivery of light fields



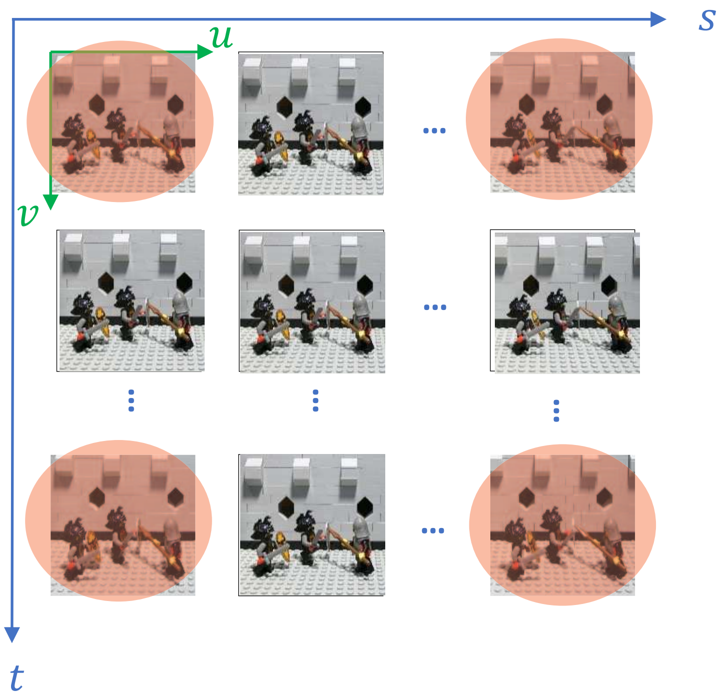
Light field coding

- Sparse subset
- Standard video coding
- Multiview video coding
- JPEG Pleno
- Learning-based coding

Light field coding

Based on a sparse subset

- Encode a sparse subaperture images (SAIs)
- Reconstruct missing SAIs at decoder

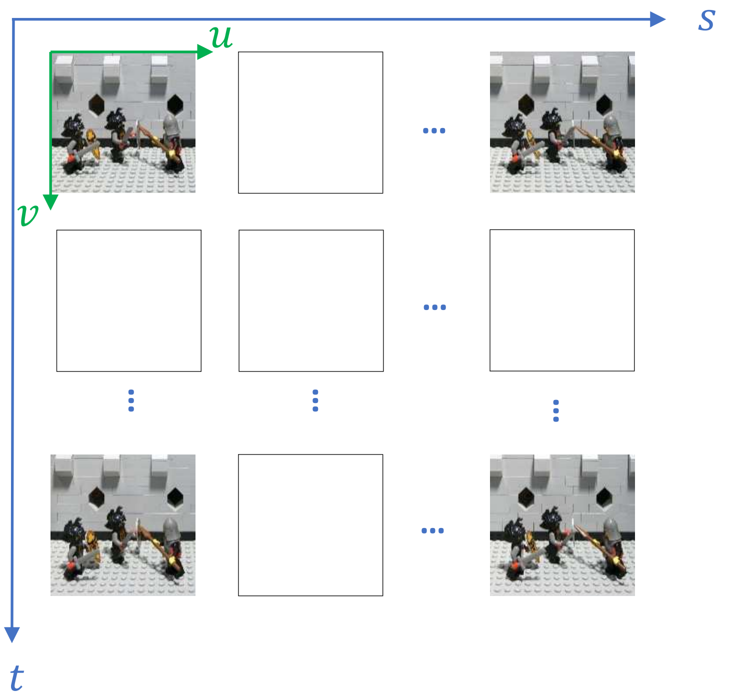


Magnor, Marcus, and Bernd Girod. "**Hierarchical coding of light fields with disparity maps**." *Proceedings 1999 International Conference on Image Processing,* IEEE, 1999.
Dib, Elian, Mikaël Le Pendu, and Christine Guillemot. "**Light field compression using Fourier disparity layers**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.
Jiang, Xiaoran, Mikael Le Pendu, and Christine Guillemot. "**Light field compression using depth image based view synthesis**." *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2017.

Light field coding

Based on a sparse subset

- Encode a sparse subaperture images (SAIs)
- Reconstruct missing SAIs at decoder

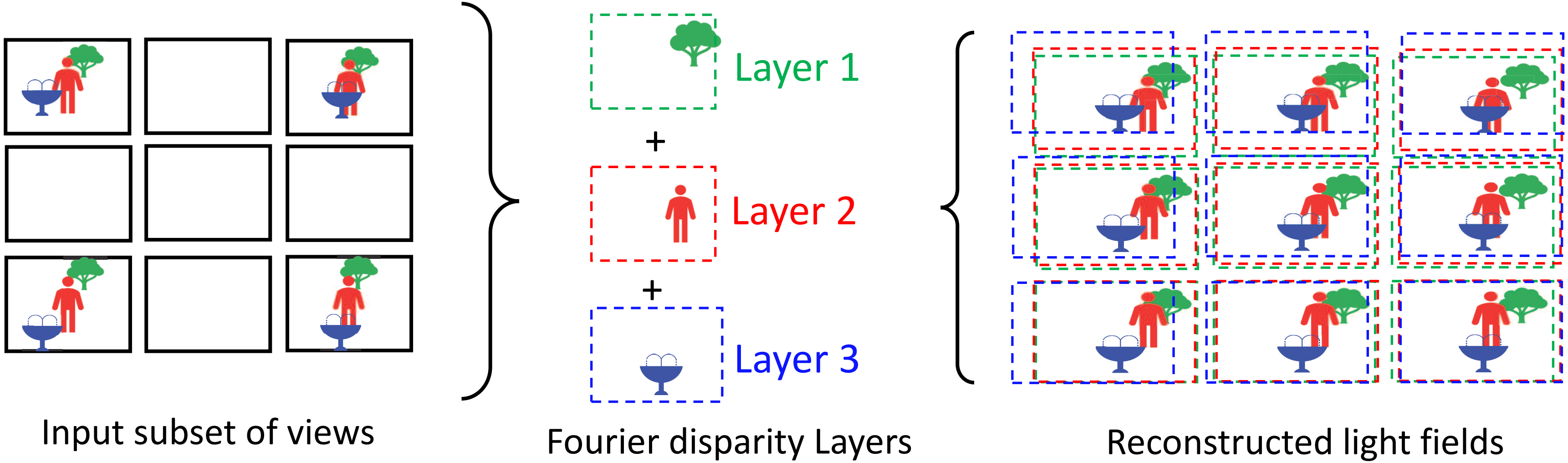


Magnor, Marcus, and Bernd Girod. "**Hierarchical coding of light fields with disparity maps**." *Proceedings 1999 International Conference on Image Processing,* IEEE, 1999.
Dib, Elian, Mikaël Le Pendu, and Christine Guillemot. "**Light field compression using Fourier disparity layers**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.
Jiang, Xiaoran, Mikael Le Pendu, and Christine Guillemot. "**Light field compression using depth image based view synthesis**." *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2017.

Light field coding

Based on a sparse subset

- Light field reconstruction from a sparse set of views using Fourier disparity layers (FDL)



Dib, Elian, Mikaël Le Pendu, and Christine Guillemot. "**Light field compression using Fourier disparity layers**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.
Le Pendu, Mikael, Christine Guillemot, and Aljosa Smolic. "**A fourier disparity layer representation for light fields**." *IEEE Transactions on Image Processing* 28.11 (2019): 5740-5753.

Light field coding

Based on a sparse subset

- Graph-based representation is used to exploit correlation between SAI
- The main idea is to estimate how similar each perspective image by representing connection weighs between each sub-aperture image

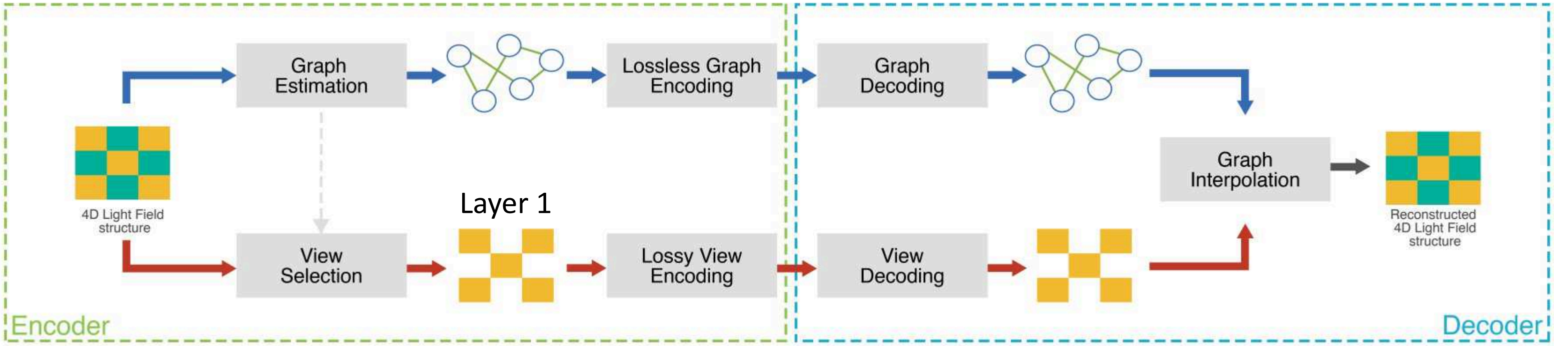


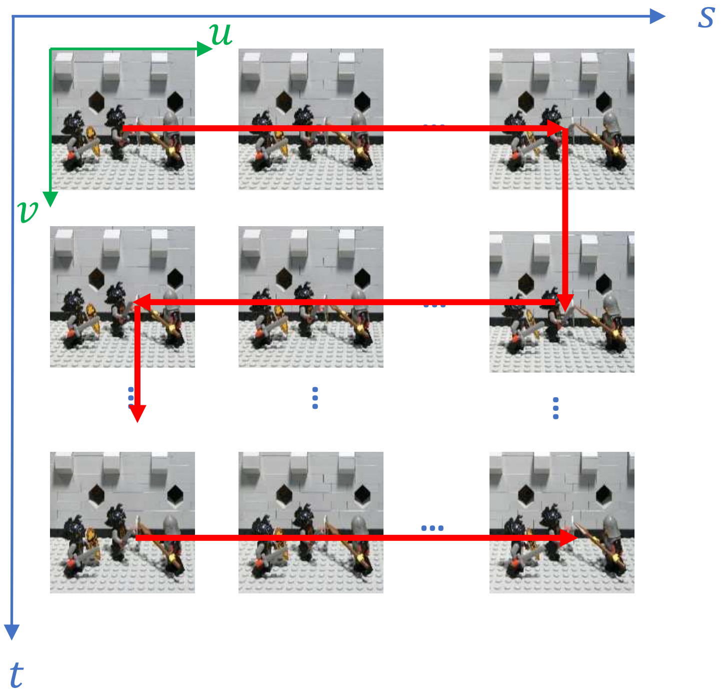
Figure. Overview of the graph-based learning approach

Viola, Irene, et al. "**A graph learning approach for light field image compression**." *Applications of Digital Image Processing XLI*. Vol. 10752. International Society for Optics and Photonics, 2018.

Light field coding

Based on standard video coding

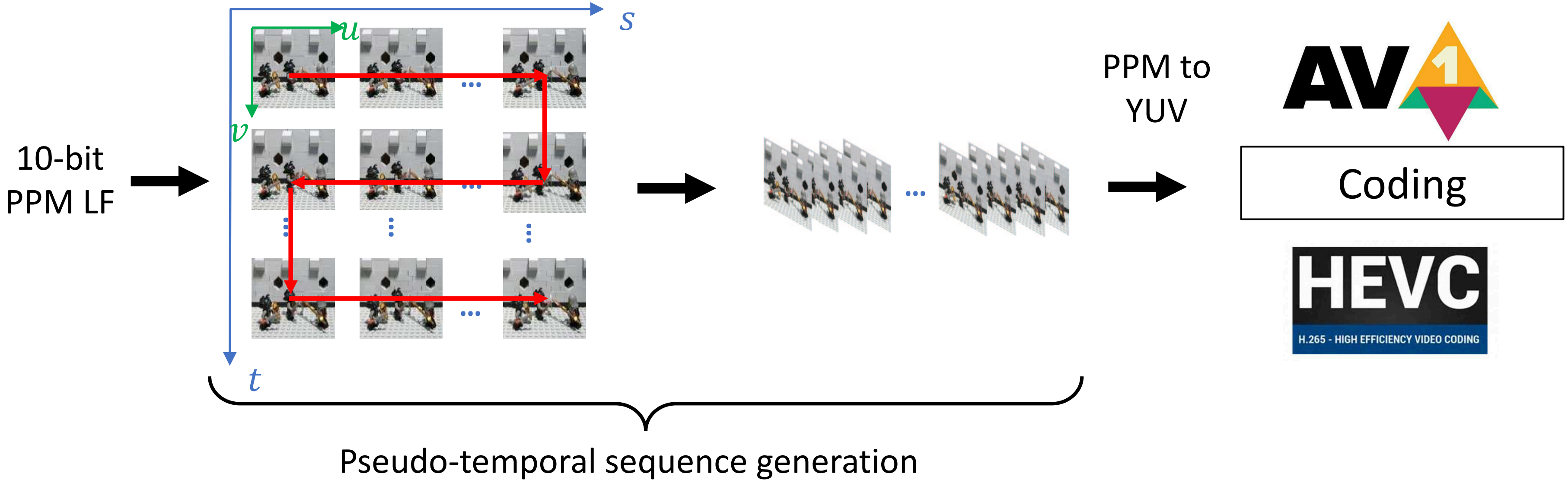
- Encode all SAIs as a pseudo-video
- Coding performance depends on the SAIs order



Li, Li, et al. "**Pseudo-sequence-based 2-D hierarchical coding structure for light-field image compression**." *IEEE Journal of Selected Topics in Signal Processing* 11.7 (2017): 1107-1119.
Conti, Caroline, Luís Ducla Soares, and Paulo Nunes. "**Dense Light Field Coding: A Survey**." *IEEE Access* 8 (2020): 49244-49284.

Light field coding

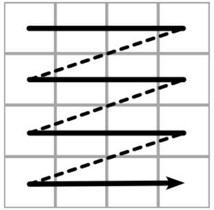
Based on standard video coding



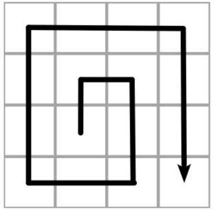
Li, Li, et al. "**Pseudo-sequence-based 2-D hierarchical coding structure for light-field image compression**." *IEEE Journal of Selected Topics in Signal Processing* 11.7 (2017): 1107-1119.
Conti, Caroline, Luís Ducla Soares, and Paulo Nunes. "**Dense Light Field Coding: A Survey**." *IEEE Access* 8 (2020): 49244-49284.

Light field coding

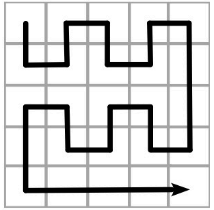
Based on standard video coding



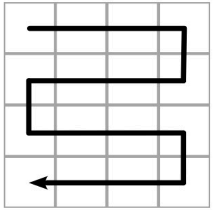
Raster



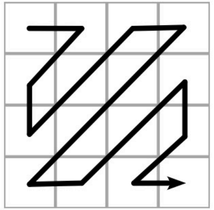
Spiral



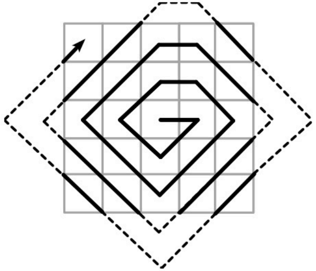
U-shape



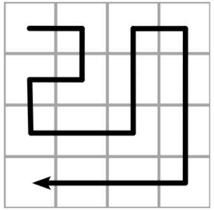
Serpentine



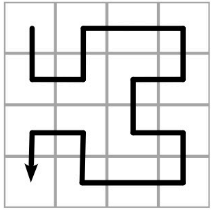
Zig-zag



Lozenge



Perpendicular



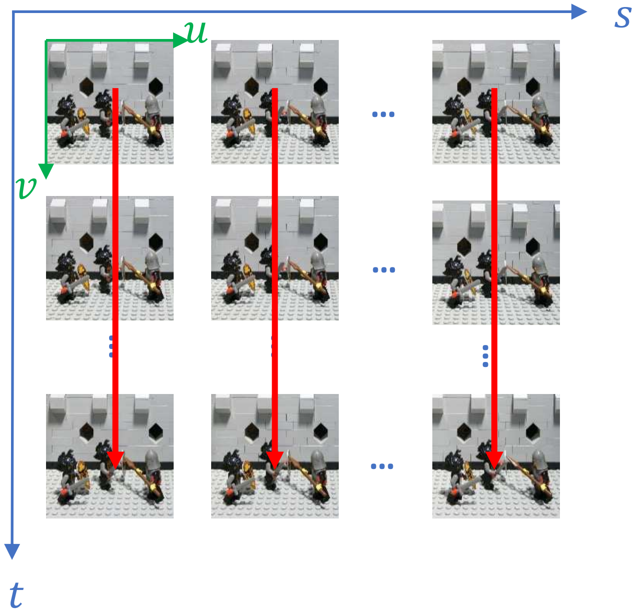
Hilbert

Perra, Cristian, and Pedro Assuncao. "**High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement**." *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2016.
Conti, Caroline, Luís Ducla Soares, and Paulo Nunes. "**Dense Light Field Coding: A Survey**." *IEEE Access* 8 (2020): 49244-49284.

Light field coding

Based on multi-view video coding

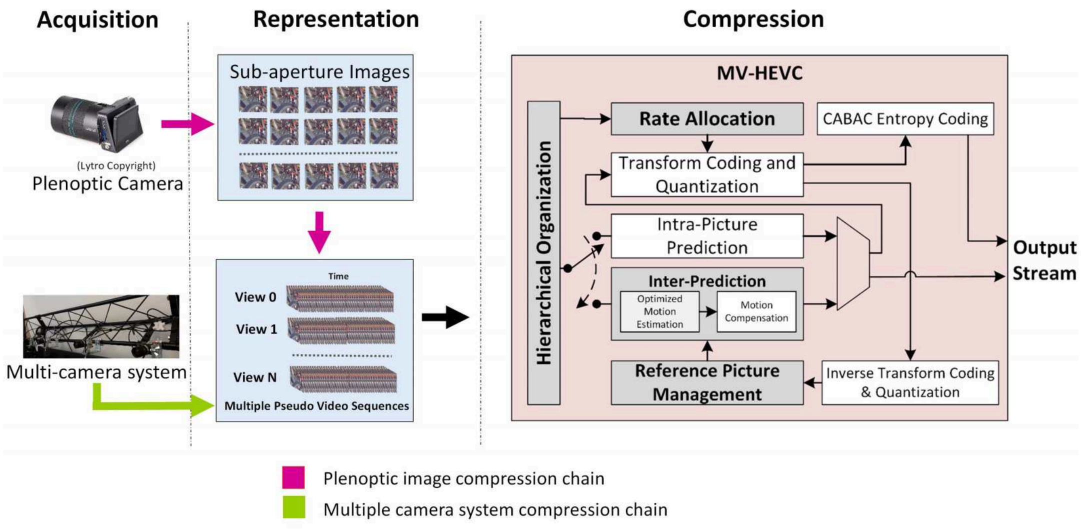
- Encode all SAIs with time using multi-view video coding standards, e.g., MV-HEVC or 3D-HEVC
- 2D prediction structure is designed to better adapt to SAIs
- Exploiting spatial-angular redundancies



Li, Li, et al. "**Pseudo-sequence-based 2-D hierarchical coding structure for light-field image compression**." *IEEE Journal of Selected Topics in Signal Processing* 11.7 (2017): 1107-1119.
Conti, Caroline, Luís Ducla Soares, and Paulo Nunes. "**Dense Light Field Coding: A Survey**." *IEEE Access* 8 (2020): 49244-49284.

Light field coding

Based on multi-view video coding



Ahmad, Waqas, Roger Olsson, and Marten Sjostrom. "**Towards a generic compression solution for densely and sparsely sampled light field data**." *25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018.
Ahmad, Waqas, et al. "**Computationally Efficient Light Field Image Compression Using a Multiview HEVC Framework**." *IEEE Access* 7 (2019).

Light field coding

JPEG Pleno

- Multidimensional Light field Encoder using 4D Transforms and Hexadeca-trees (MuLE-TH)

  Exploiting the 4D redundancy of light fields by using a 4D transform and hexadeca-trees

de Carvalho, Murilo B., et al. "**A 4D DCT-based lenslet light field codec**." *25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018.

- Warping and Sparse Prediction (WaSP)

  The method uses depth, disparity and sparse prediction information to reconstruct the final set of LFs

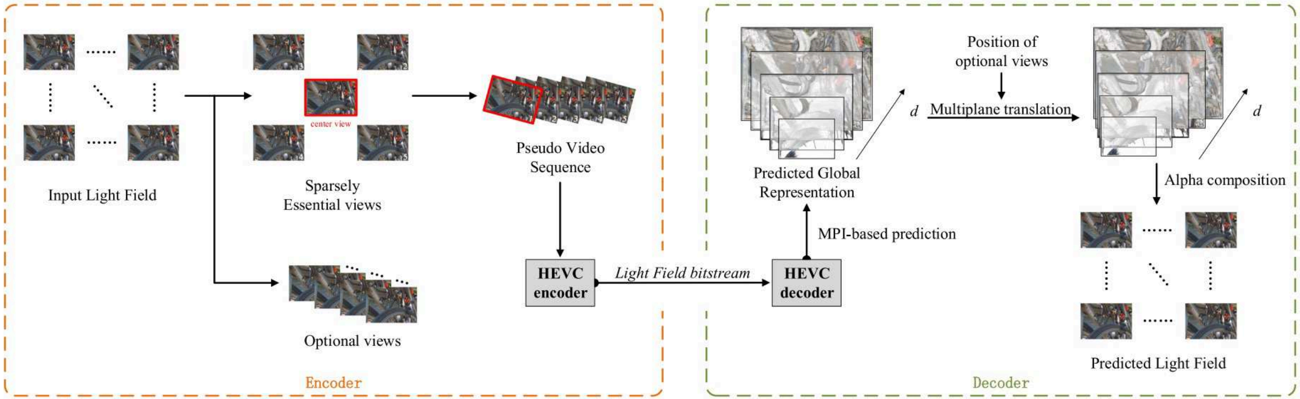
Astola, Pekka, and Ioan Tabus. "**WaSP: Hierarchical warping, merging, and sparse prediction for light field image compression**." *7th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2018.

Source code: https://github.com/EduardoAmaro/vm-2.1-RDO

Light field coding

Learning-based coding

- Utilizing learning based multiplane images (MPI)



Chen, Y., An, P., Huang, X., Yang, C., Liu, D., & Wu, Q. (2020). **Light field compression using global multiplane representation and two-step prediction**. IEEE Signal Processing Letters, 27, 1135-1139.

Light field streaming

- Only a few views are transmitted for free viewpoint rendering

Ramanathan, Prashant, Mark Kalman, and Bernd Girod. "**Rate-distortion optimized interactive light field streaming**." IEEE Transactions on Multimedia 9.4 (2007): 813-825.

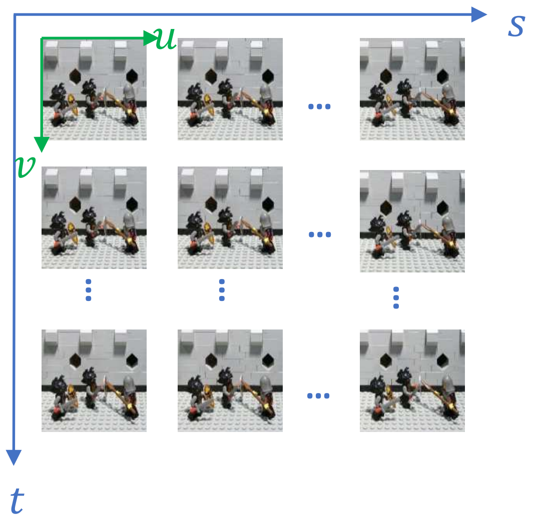
- Only subset of SAIs are transmitted

Peixoto, Eduardo, et al. "**Progressive communication for interactive light field image data streaming**." 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017.

Peixoto, Eduardo, et al. "**Progressive sub-aperture image recovery for interactive light field data streaming**." 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

de Souza, Wallace Bruno S., et al. "**A sub-aperture image selection refinement method for progressive light field transmission**." 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 2018.

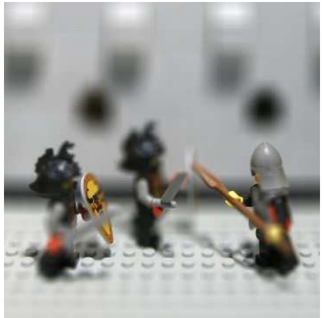
Light field streaming



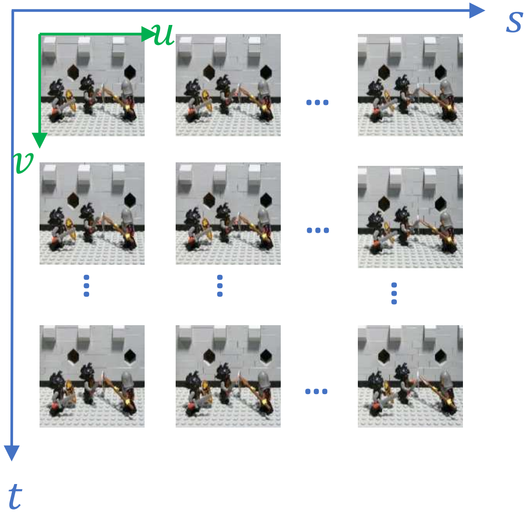
Cloud – Light field storage

Alain, Martin, Cagri Ozcinar, and Aljosa Smolic. "**A Study of Light Field Streaming for An Interactive Refocusing Application**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

Light field streaming



User side – Light field visualisation

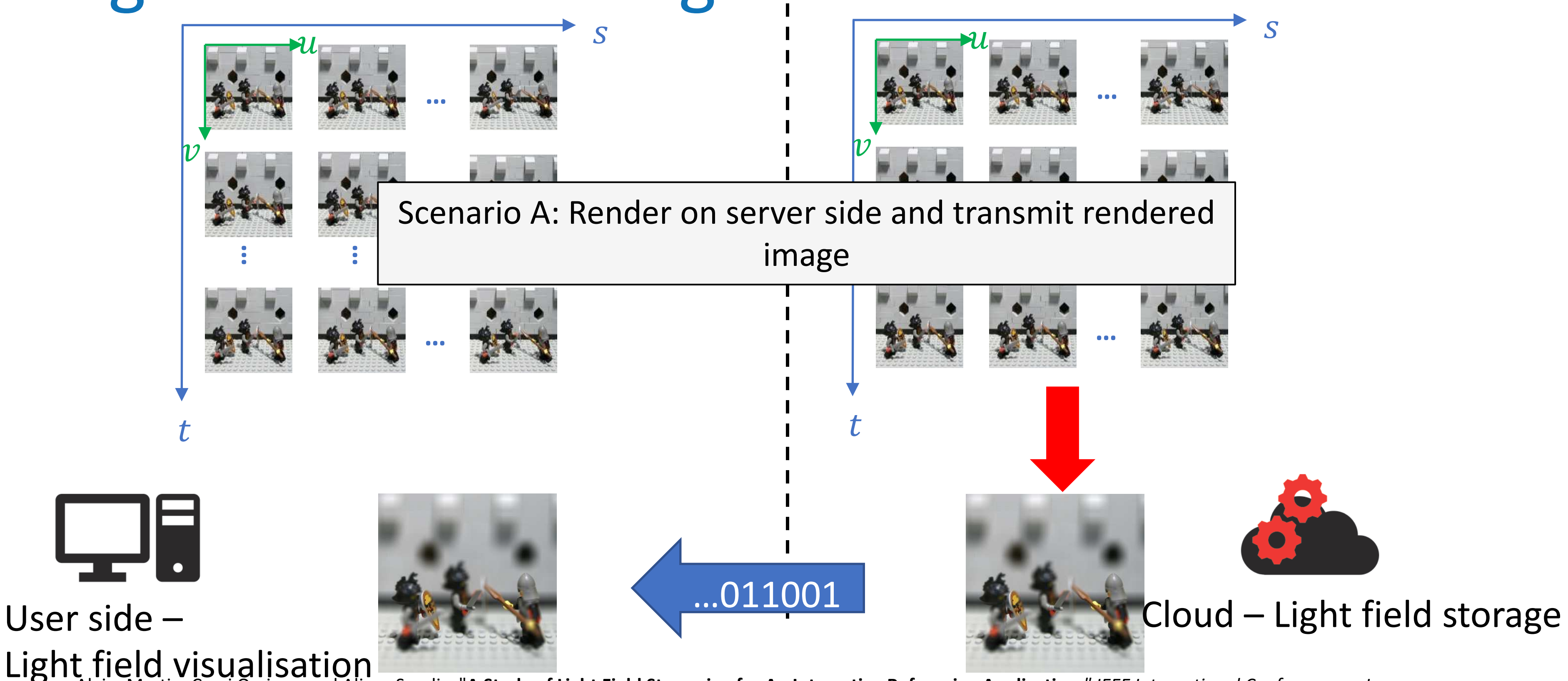


User side –
Light field visualisation

Cloud – Light field storage

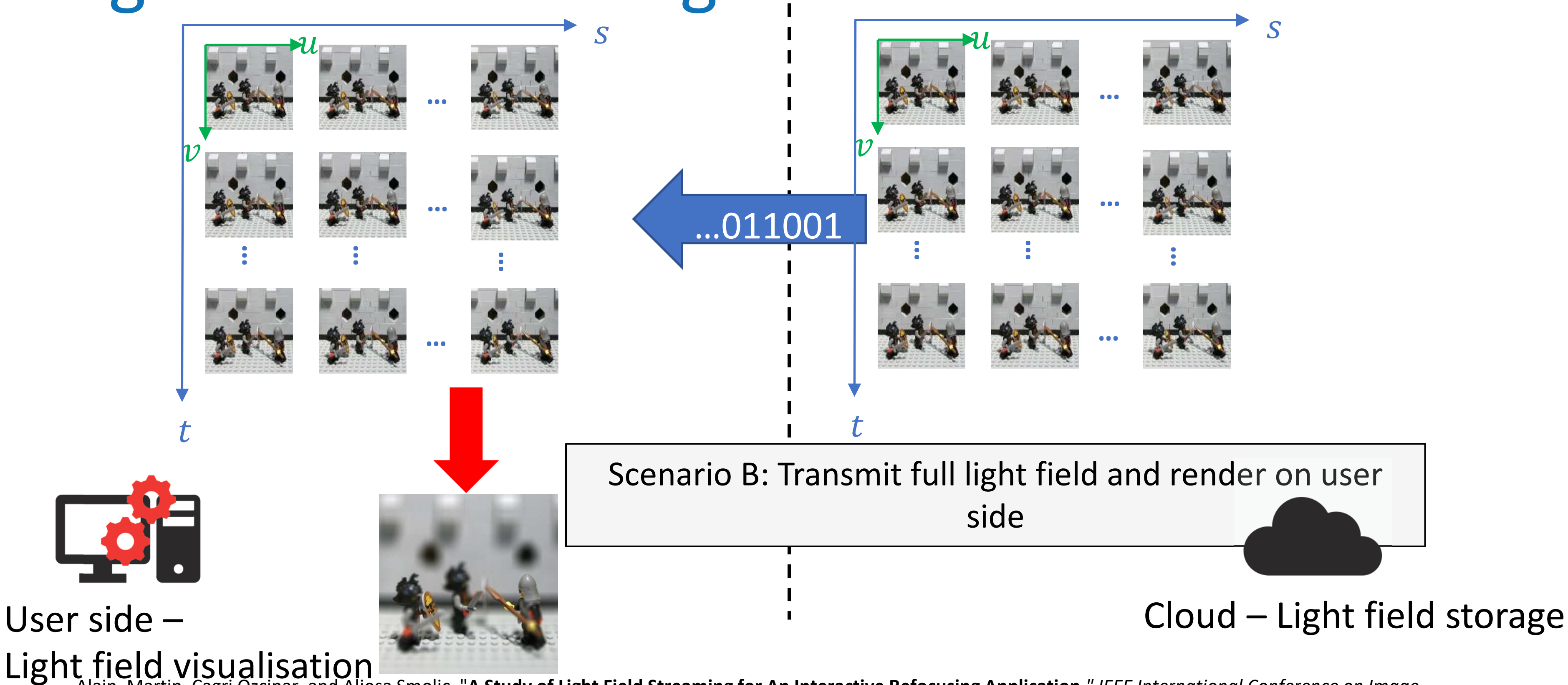
Alain, Martin, Cagri Ozcinar, and Aljosa Smolic. "**A Study of Light Field Streaming for An Interactive Refocusing Application**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

Light field streaming



Alain, Martin, Cagri Ozcinar, and Aljosa Smolic. "**A Study of Light Field Streaming for An Interactive Refocusing Application**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

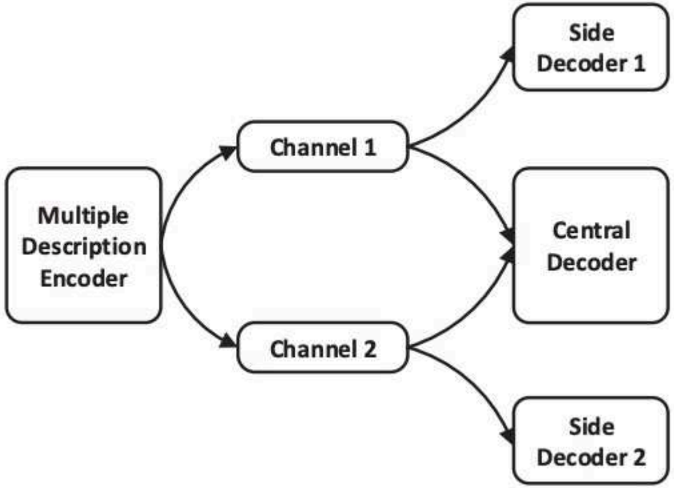
Light field streaming



Alain, Martin, Cagri Ozcinar, and Aljosa Smolic. "**A Study of Light Field Streaming for An Interactive Refocusing Application**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

Light field streaming

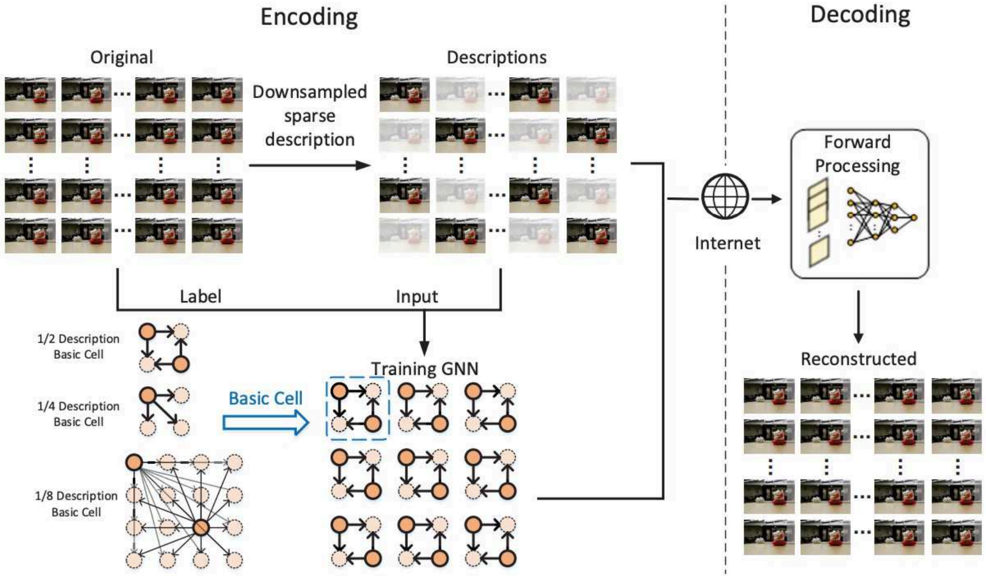
- Best-effort delivery of LF videos over a wireless network
- Encode LFs with multiple description coding (MDC), where each description can be transmitted through a different network path



X. Hu, Y. Pan, Y. Wang, L. Zhang and S. Shirmohammadi, "**Multiple Description Coding for Best-Effort Delivery of Light Field Video using GNN-based Compression**," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2021.3129918.

Light field streaming

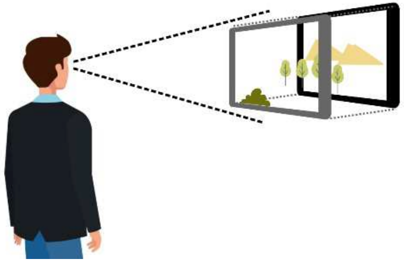
- Graph Neural Network (GNN) based LF compression and MDC



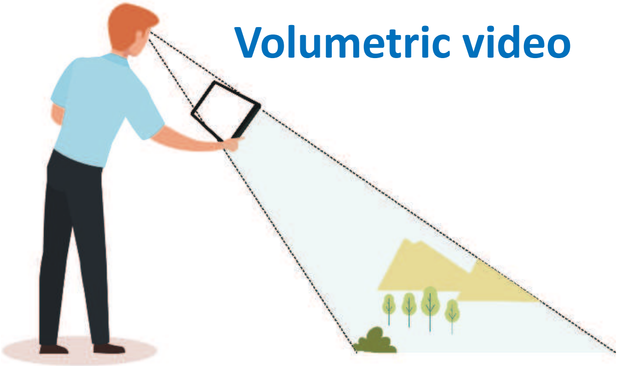
X. Hu, Y. Pan, Y. Wang, L. Zhang and S. Shirmohammadi, "**Multiple Description Coding for Best-Effort Delivery of Light Field Video using GNN-based Compression**," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2021.3129918.

Part III: Content Delivery

Light Fields

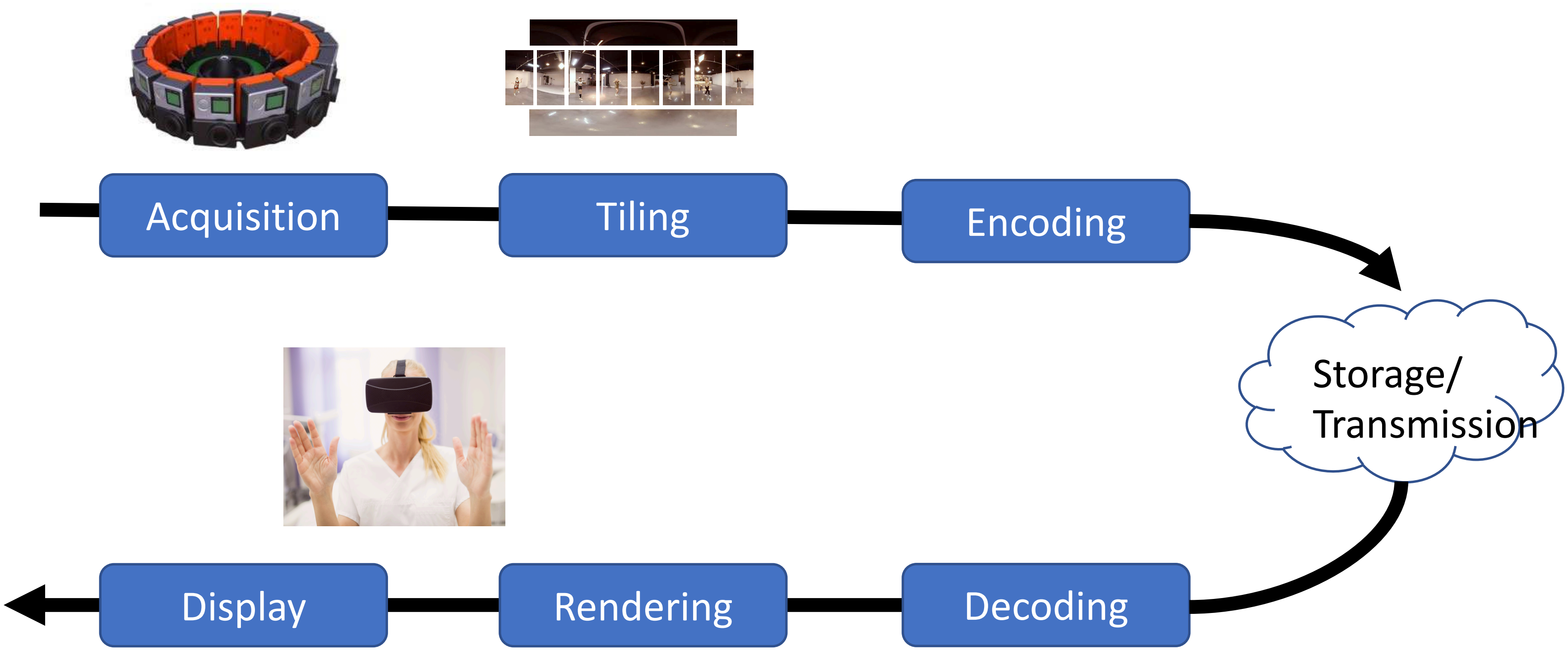


360-degree video



Volumetric video

Delivery of 360-degree video



360-degree video coding

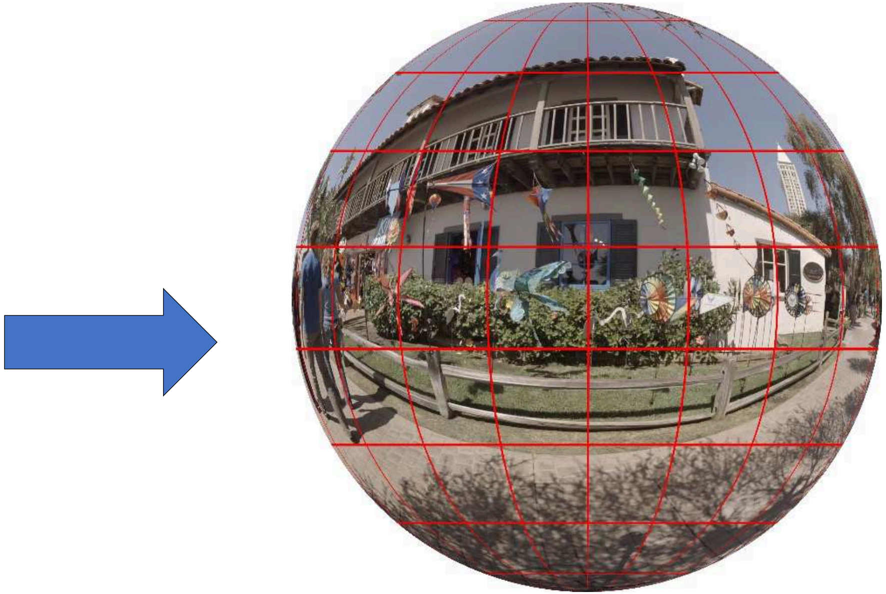
360-degree

180-degree

Equirectangular representation

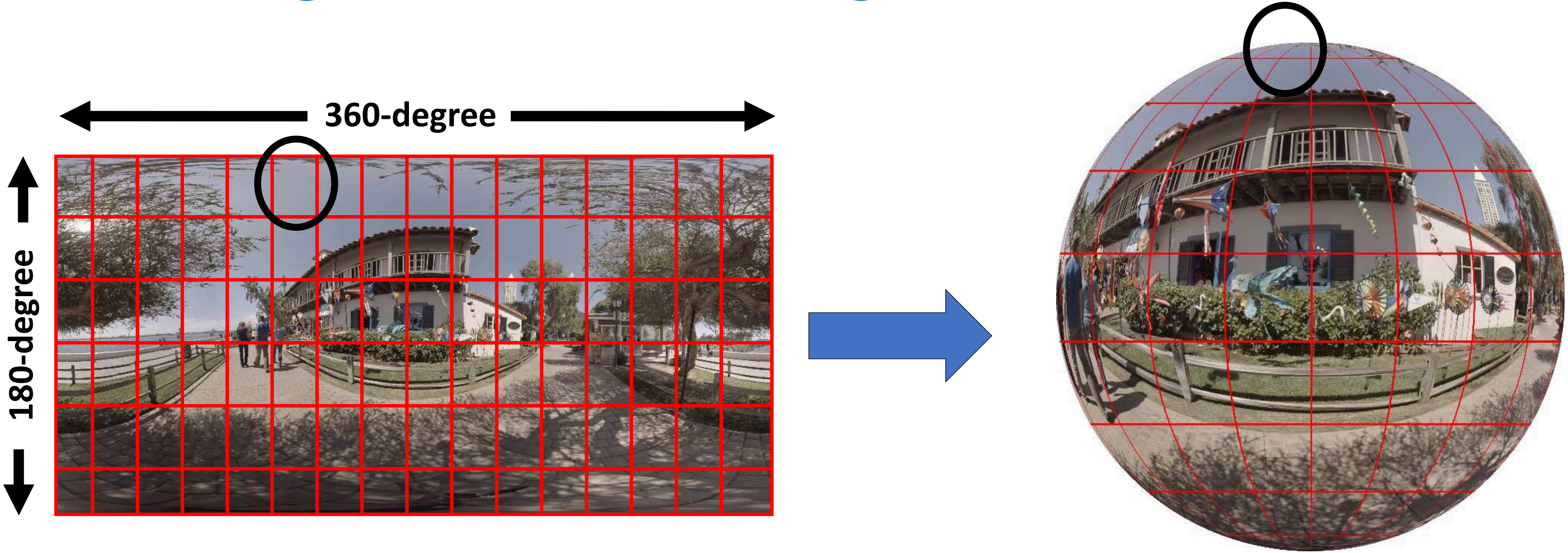
360-degree video coding

Equirectangular representation



Spherical representation

360-degree video coding



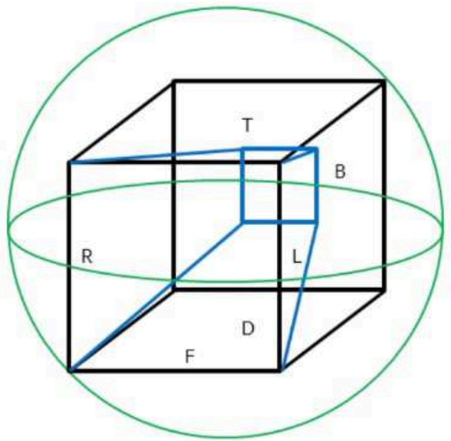
Equirectangular representation

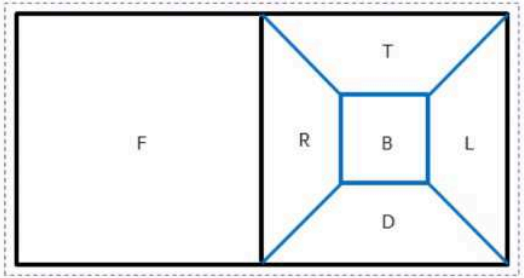
Spherical representation

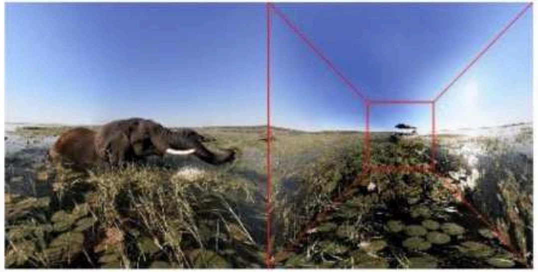
360-degree video coding

360-degree video coding

Projections







Source code:
https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/trunk

ISO/IEC JTC1/SC29/WG11/M. 2016. **VR/360 Video Truncated Square Pyramid Geometry for OMAF**.

360-degree video streaming

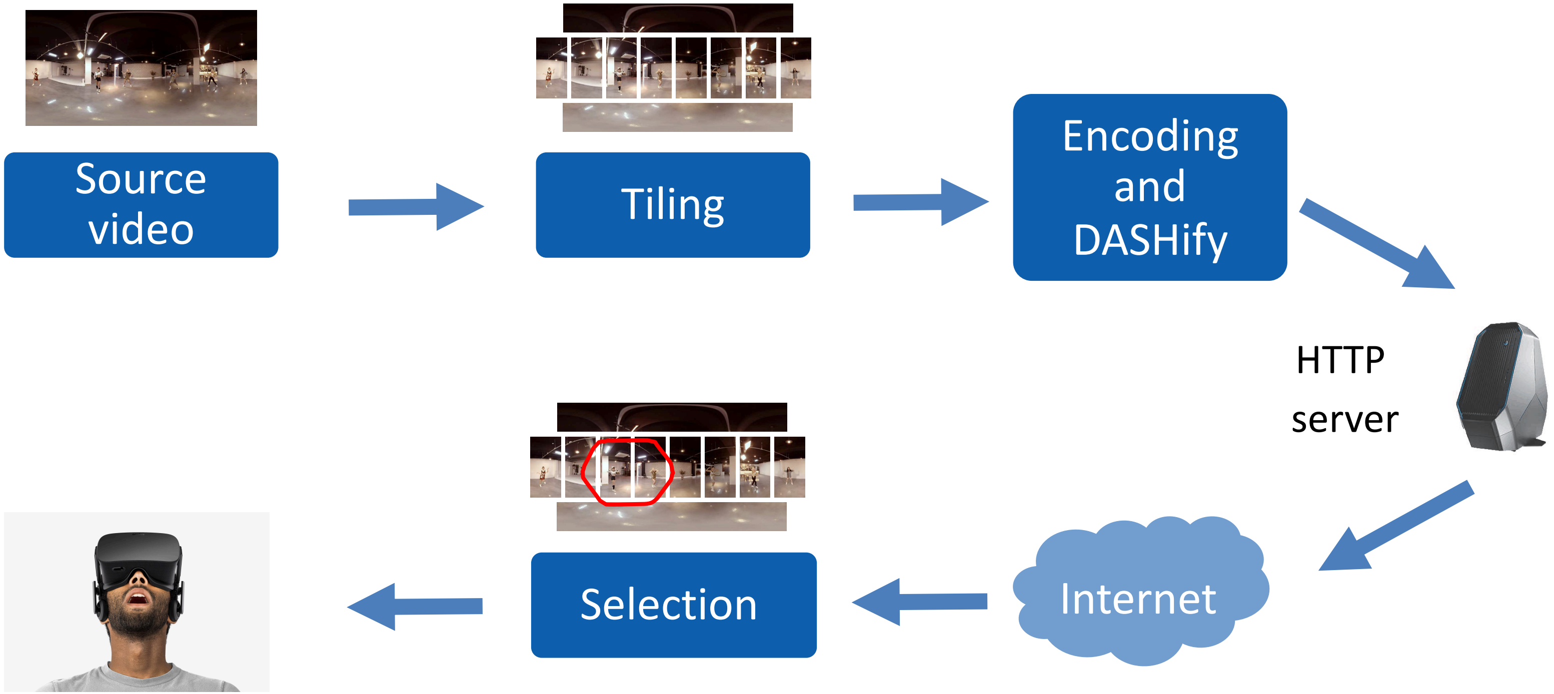


Figure: Viewport-aware adaptive streaming using tiles method

Ozcinar, Cagri, Ana De Abreu, and Aljosa Smolic. "**Viewport-aware adaptive 360 video streaming using tiles for virtual reality**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.

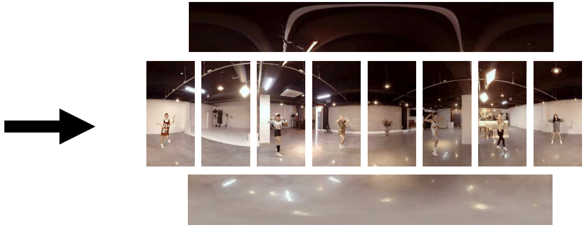
360-degree video streaming

Open source tool

- GPAC project (https://github.com/gpac/)
- Each frame can be divided into tiles and Spatial Representation Description (SRD) can be used to stream 360-degree videos using MPEG-DASH standard

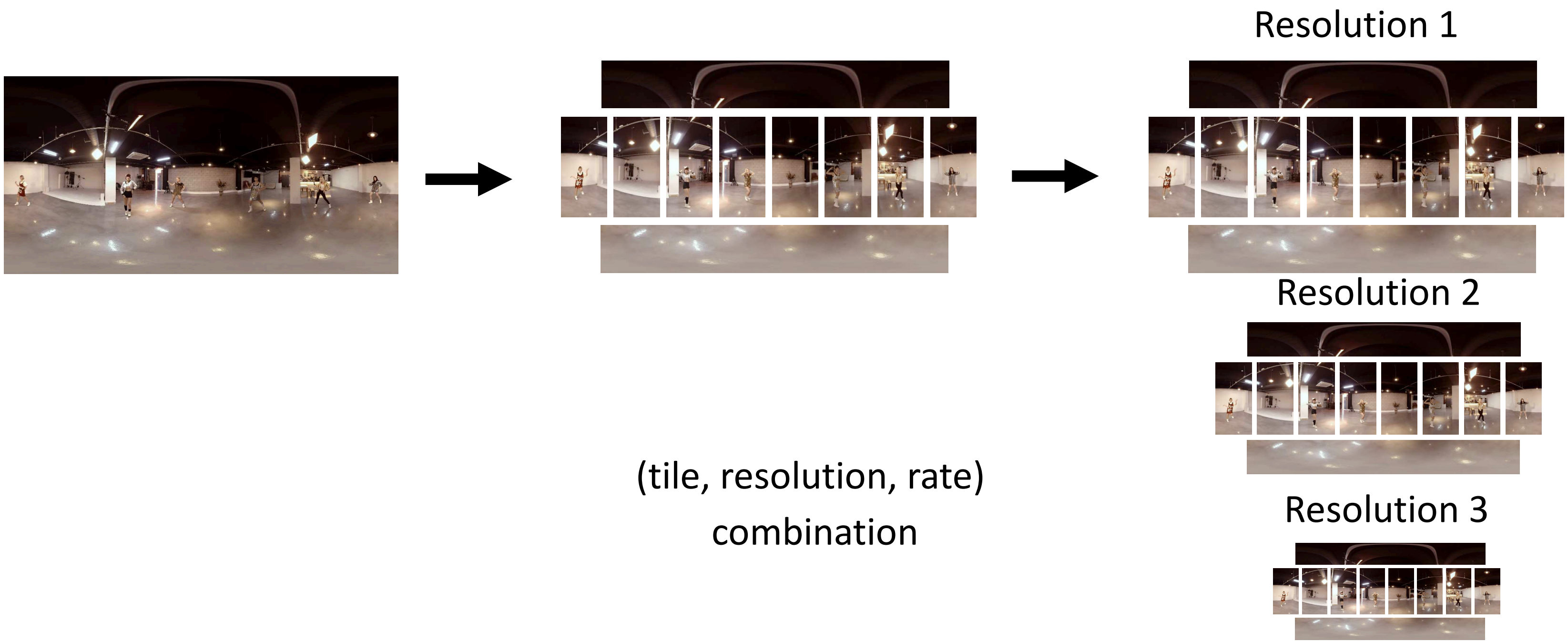
360-degree video streaming

Server-side optimization



360-degree video streaming

Server-side optimization



360-degree video streaming

Server-side optimization

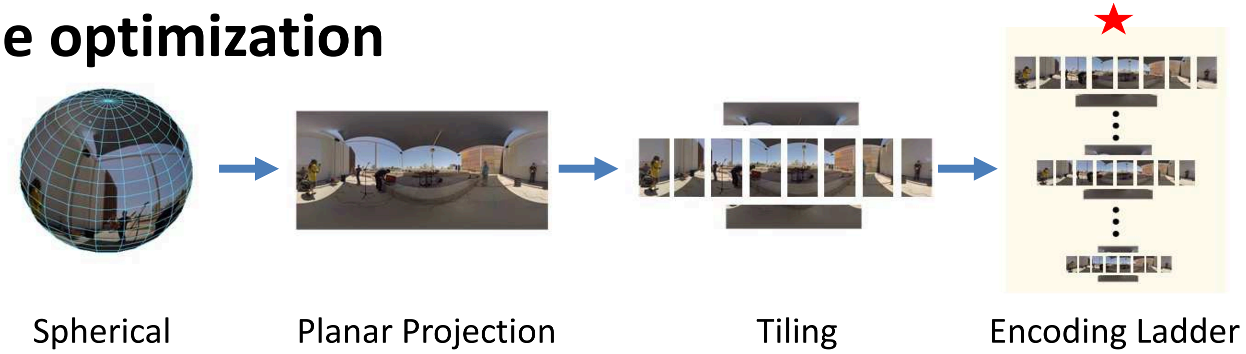


Figure: Overview of the different formats and representations

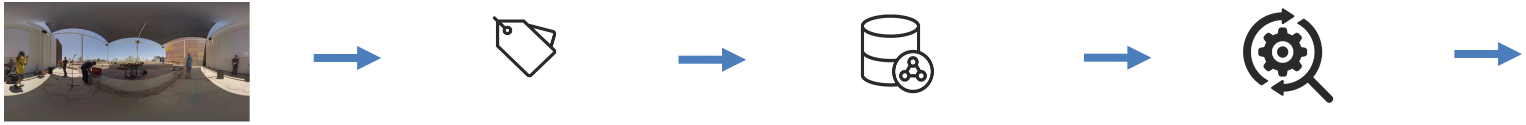
Ozcinar, Cagri, et al. "**Estimation of optimal encoding ladders for tiled 360 VR video in adaptive streaming systems**." *IEEE International Symposium on Multimedia (ISM)*. IEEE, 2017.

360-degree video streaming

Server-side optimization

Spherical | Planar Projection | Tiling | Encoding Ladder

Figure: Overview of the different formats and representations



360 VR Video → Content Category → Encoding and Storage Costs Models → Linear Optimization Algorithm → Cost-Optimal Encoding Ladder

Figure: Proposed encoding ladder estimation method

Ozcinar, Cagri, et al. "**Estimation of optimal encoding ladders for tiled 360 VR video in adaptive streaming systems**." *IEEE International Symposium on Multimedia (ISM)*. IEEE, 2017.

360-degree video streaming

Server-side optimization

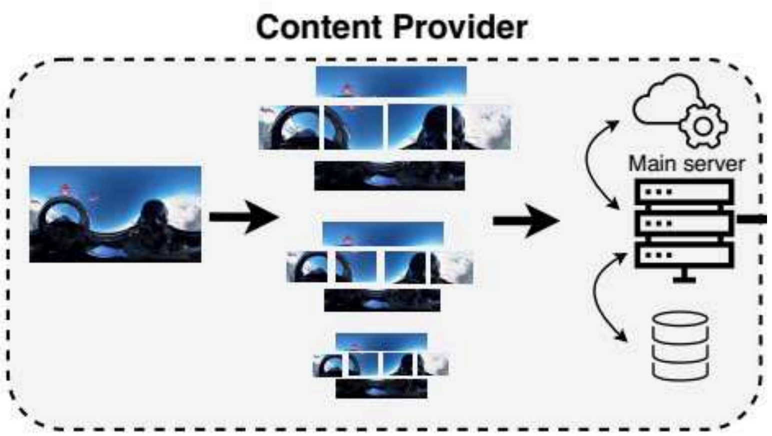


Figure: Schematic diagram of the adaptive 360 VR video streaming system

Rossi, Silvia, et al. "**Do Users Behave Similarly in VR? Investigation of the User Influence on the System Design**." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16.2 (2020): 1-26.

360-degree video streaming

Server-side optimization

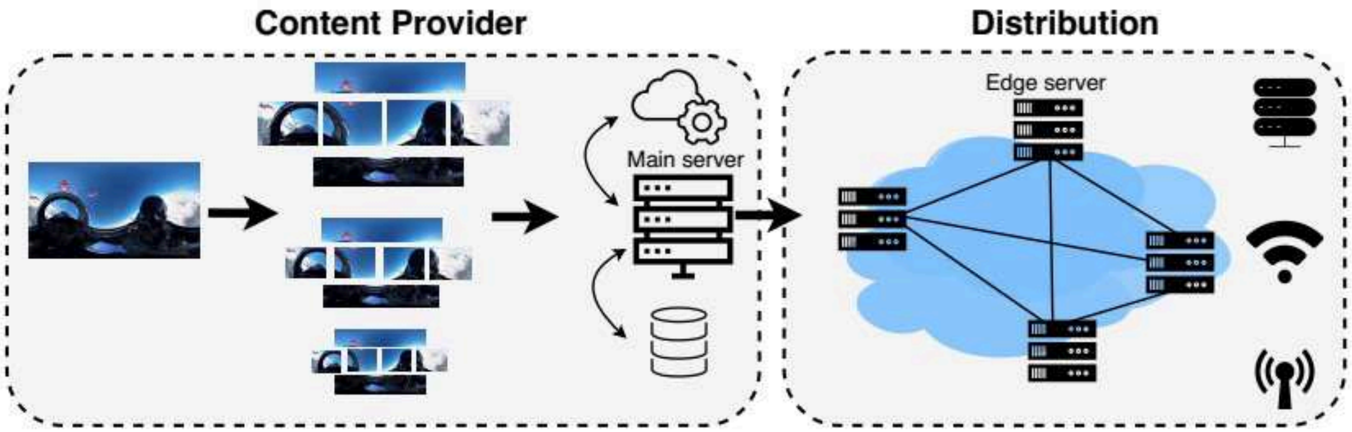


Figure: Schematic diagram of the adaptive 360 VR video streaming system

Rossi, Silvia, et al. "**Do Users Behave Similarly in VR? Investigation of the User Influence on the System Design**." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16.2 (2020): 1-26.

360-degree video streaming

Server-side optimization

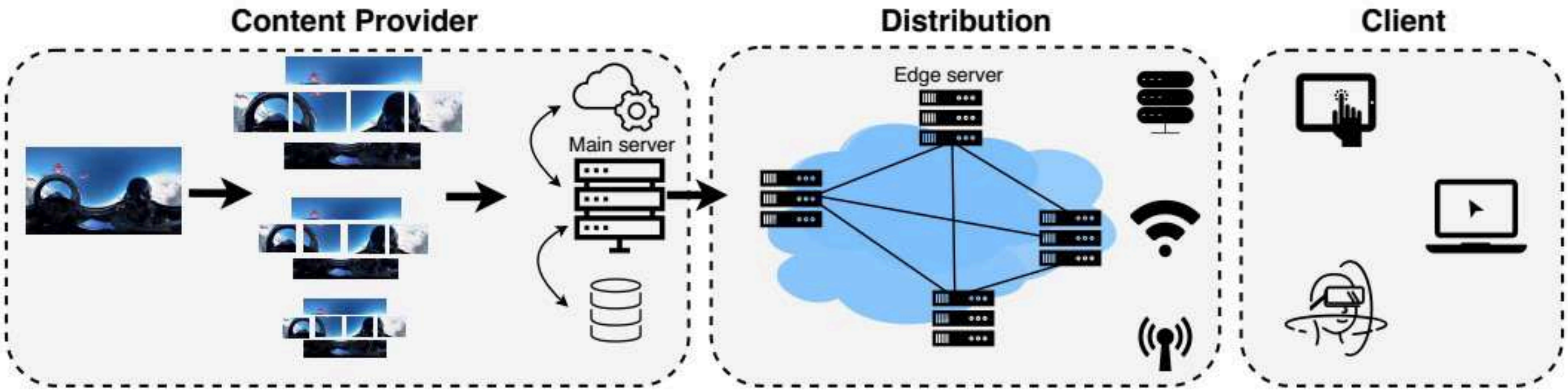
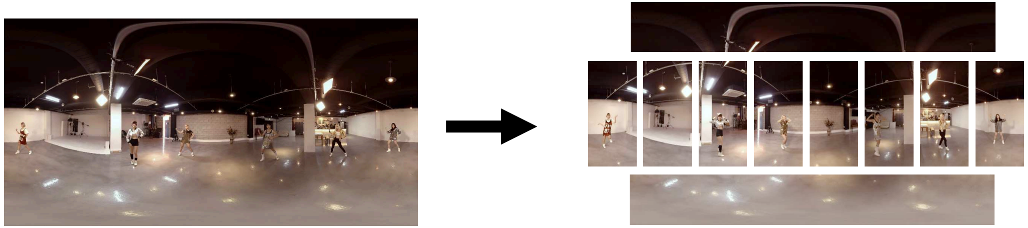


Figure: Schematic diagram of the adaptive 360 VR video streaming system

Rossi, Silvia, et al. "**Do Users Behave Similarly in VR? Investigation of the User Influence on the System Design**." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16.2 (2020): 1-26.

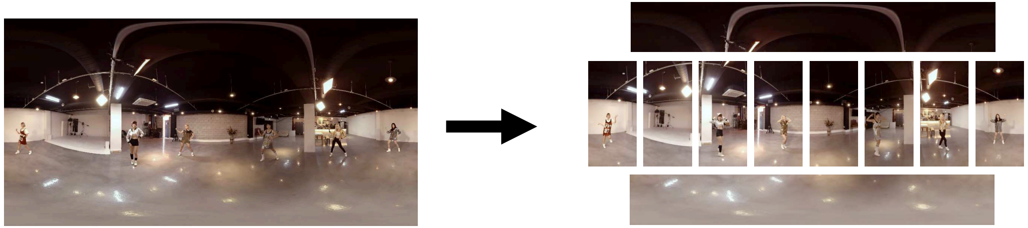
360-degree video streaming

Dynamic tiling



360-degree video streaming

Dynamic tiling



merge

merge

360-degree video streaming

Dynamic tiling

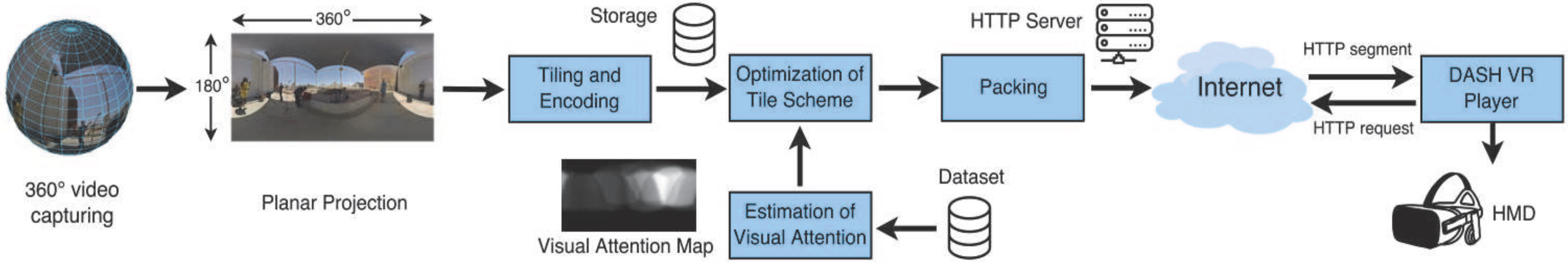


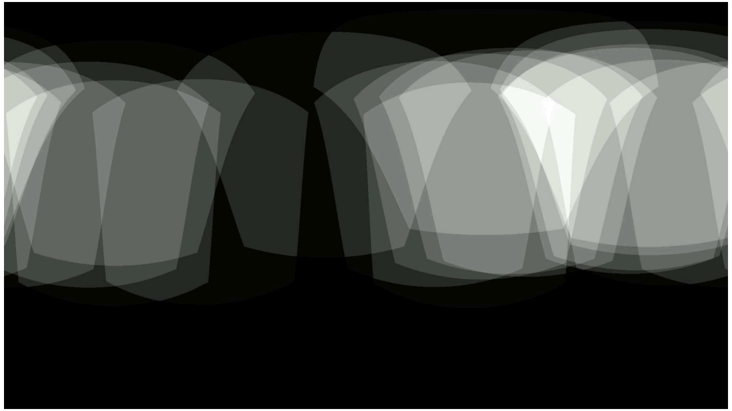
Figure: Schematic diagram of the adaptive 360 VR video streaming system

Ozcinar, Cagri, Julián Cabrera, and Aljosa Smolic. "**Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality**." IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9.1 (2019): 217-230.

360-degree video streaming

Visual attention

One user

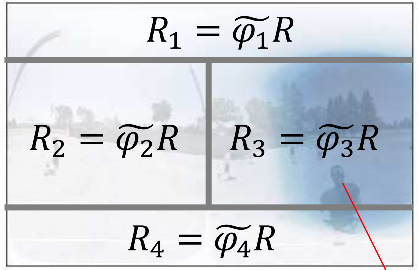


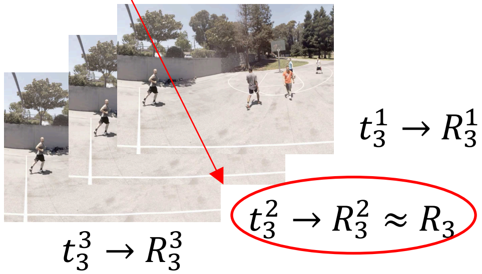
Several users

360-degree video streaming

Bitrate-allocation

Target bitrate of DASH representation -> *R* Bits/s





Computation of $\varphi_i = \frac{\sum_i^{M_t} \sum_j^{N_t} v_a^k(i,j)}{M_t N_t}$

$$\sum_i R_i^j \leq R$$

360-degree video streaming

Visual-attention based metric

Considers the relevance of each pixel according to visual attention data.

$$\text{MSE} = \frac{\sum_i^M \sum_j^N e(i,j)^2}{M \quad N}$$

$$\text{VA-MSE} = \frac{\sum_i^M \sum_j^N e(i,j)^2 v_a(i,j)}{S_{Vp}}$$

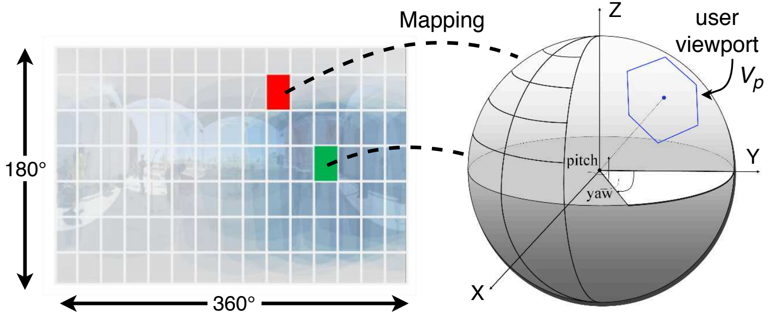
Visual attention map

360-degree video streaming

Spherical weighted metric

Accounts for the geometrical distortion of the ERP

$$\text{VASW-MSE} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} e(i,j)^2 w(i,j) v_a(i,j)}{S_{V_p} = \sum_{i,j\in V_p} w(i,j)},$$



Y Sun, A Lu, and L Yu, "**Weighted-to-Spherically-Uniform quality evaluation for omnidirectional vid**eo," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1408–1412, 2017.

360-degree video streaming

Optimal tiling scheme selection

Target bitrate of DASH representation -> *R* Bits/s

$T_s^* = \max_{T_s} Q(T_s, V_A, B(R, V_A))$ -> Evaluation of each tiling scheme, T_s

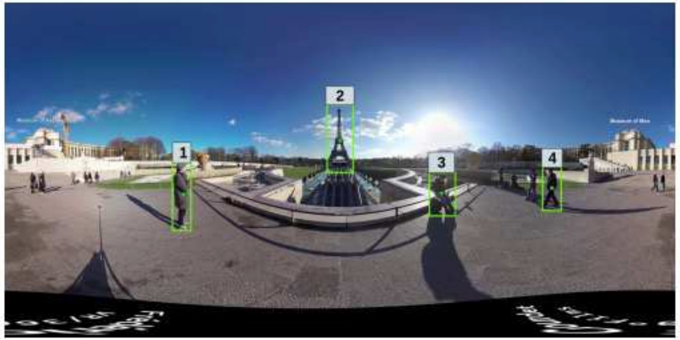
VASW-PSNR (T_1)

VASW-PSNR (T_2)

VASW-PSNR (T_3)

VASW-PSNR (T_4)

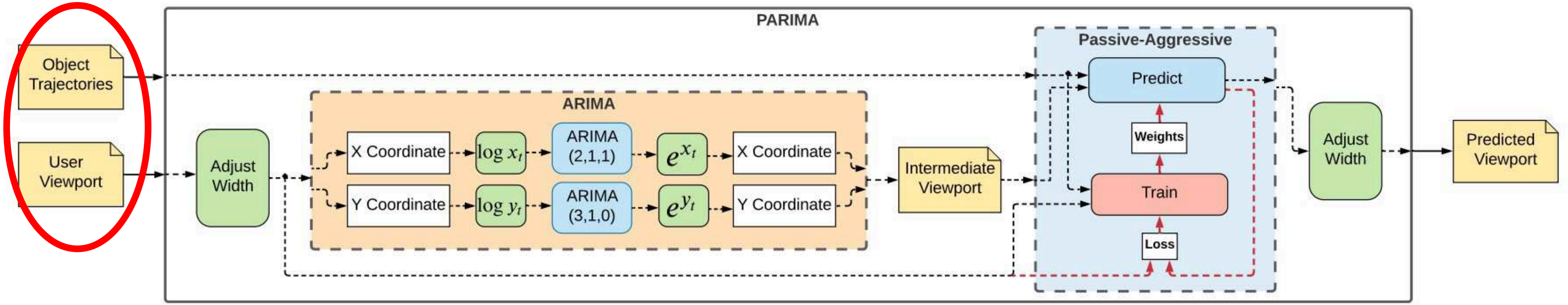
360-degree video streaming



Spherical Object tracking for 360° videos

Chopra, Lovish, Sarthak Chakraborty, Abhijit Mondal, and Sandip Chakraborty. **"PARIMA: Viewport Adaptive 360-Degree Video Streaming."** In Proceedings of the Web Conference 2021, pp. 2379-2391. 2021.

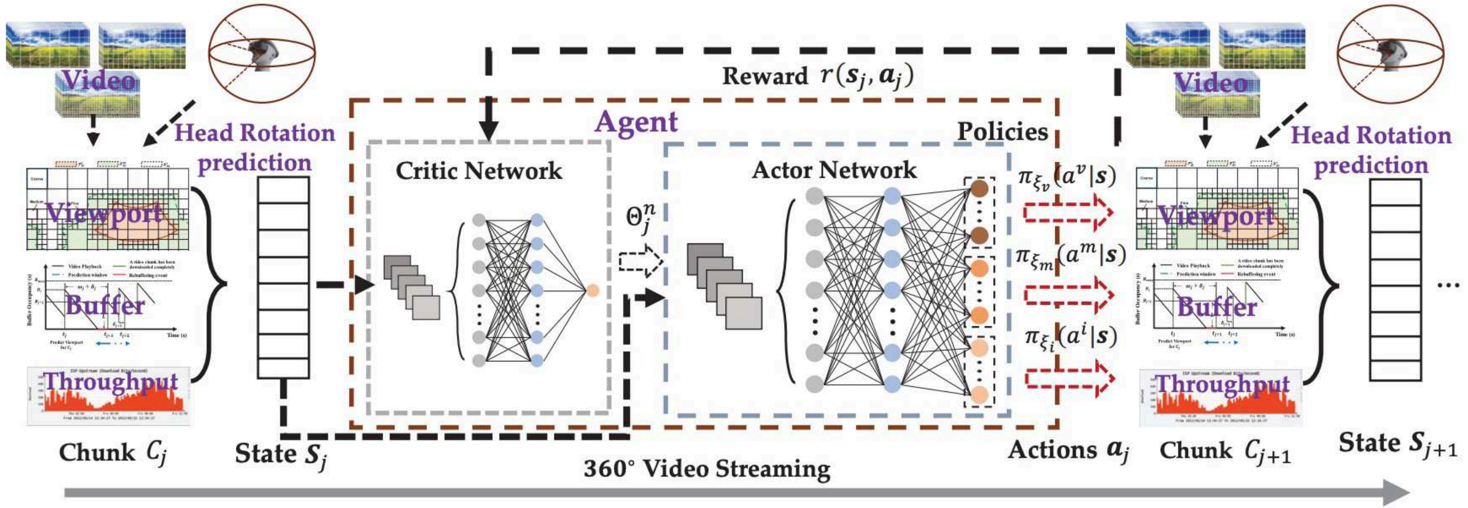
360-degree video streaming



Time-series model with Passive Aggressive Regression

Chopra, Lovish, Sarthak Chakraborty, Abhijit Mondal, and Sandip Chakraborty. **"PARIMA: Viewport Adaptive 360-Degree Video Streaming."** In Proceedings of the Web Conference 2021, pp. 2379-2391. 2021.

360-degree video streaming

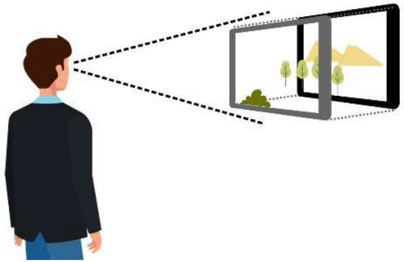


Main framework of the proposed DRL-based rate adaptation algorithm for adaptive 360-degree video streaming

N. Kan, J. Zou, C. Li, W. Dai and H. Xiong, "**RAPT360: Reinforcement Learning-Based Rate Adaptation for 360-degree Video Streaming with Adaptive Prediction and Tiling**," in IEEE Transactions on Circuits and Systems for Video Technology, doi: 10.1109/TCSVT.2021.3076585.

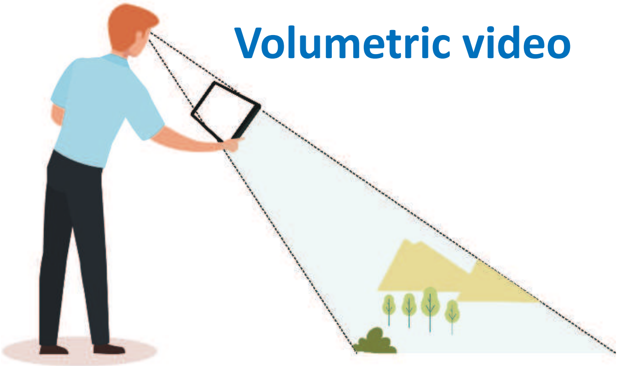
Part III: Content Delivery

Light Fields

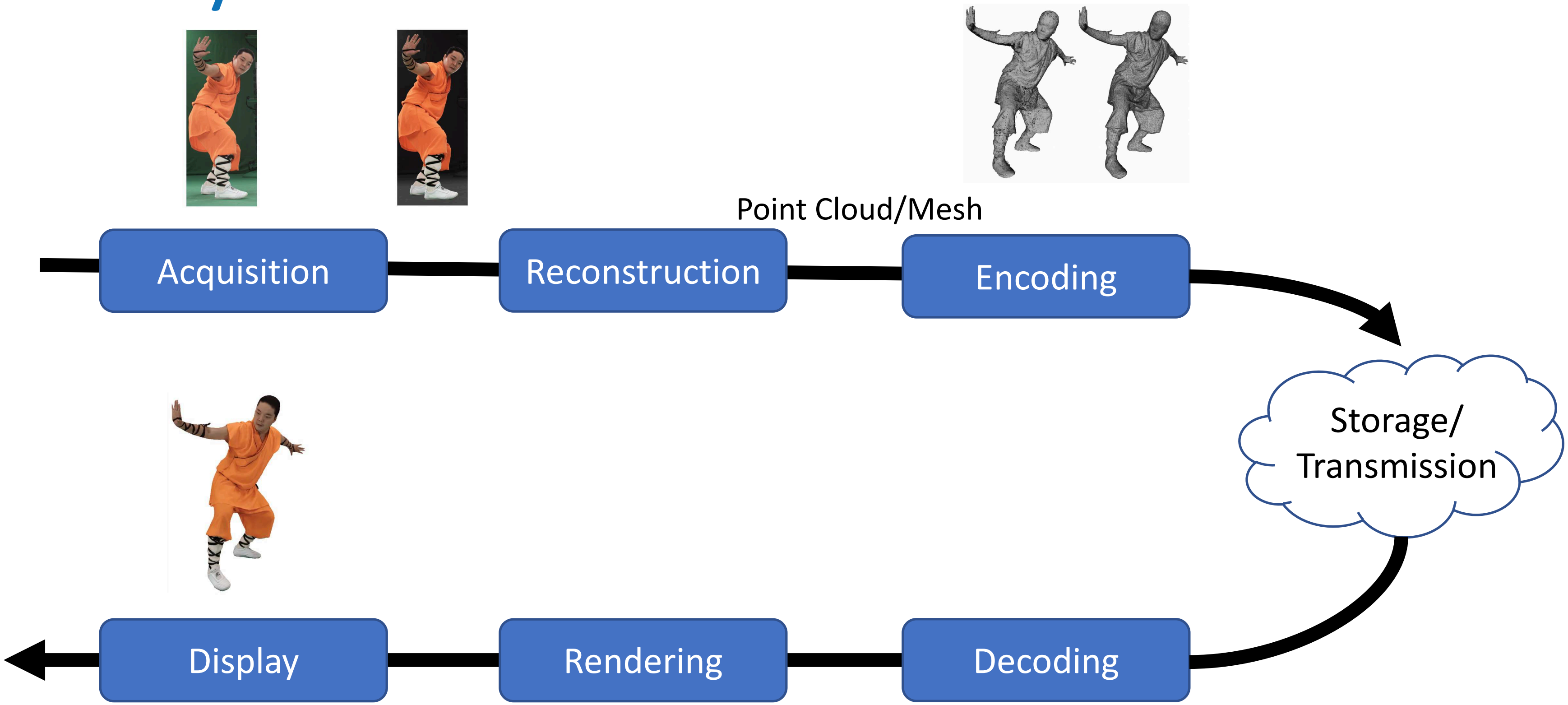


360-degree video

Volumetric video



Delivery of volumetric video



Volumetric video coding

MPEG issued a call for proposals on point cloud coding (PCC)

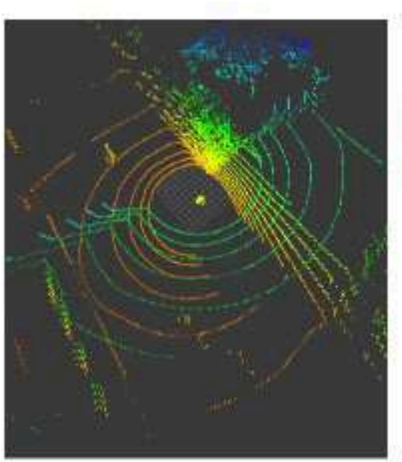
Two distinct coding technologies were selected for point cloud compression (PCC) standardization activities

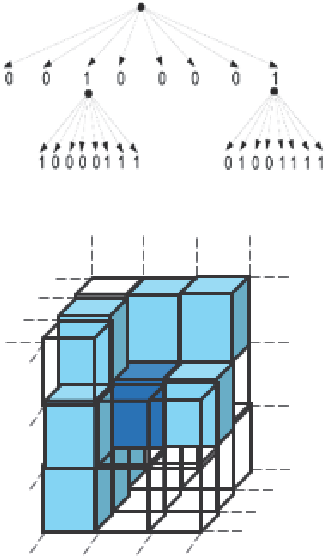
Geometry-based PCC (G-PCC) and video-based PCC (V-PCC)

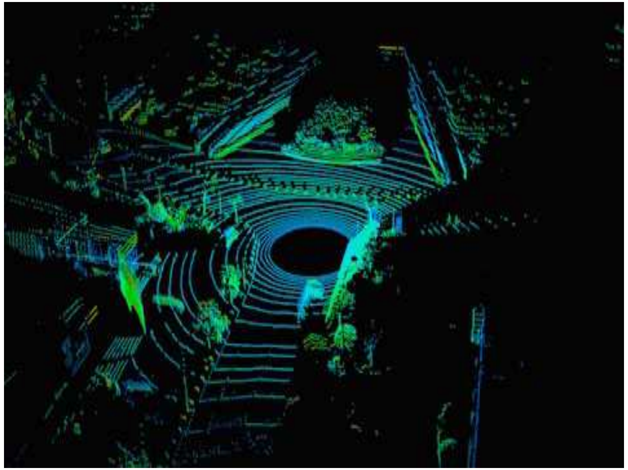
One for categories 1 and 3 (TMC13) and other one for category 2 (TMC2)

Volumetric video coding

Geometry-based PCC





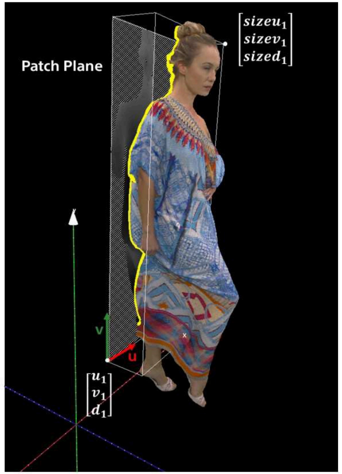


Source code: https://github.com/MPEGGroup/mpeg-pcc-tmc13

Graziosi, D., et al. "**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 (2020).

Volumetric video coding

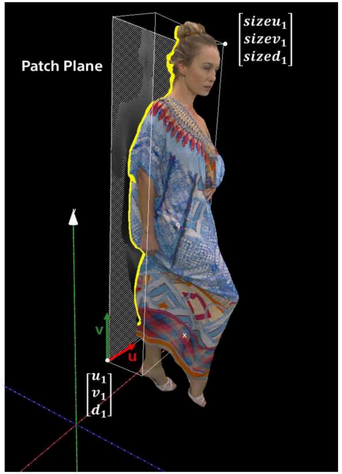
Video-based PCC



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

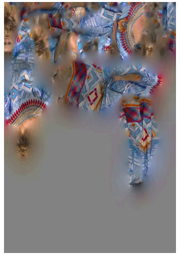
Volumetric video coding

Video-based PCC



Occupancy map

Geometry



Attribute images

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

Volumetric video coding

Based on polygon mesh

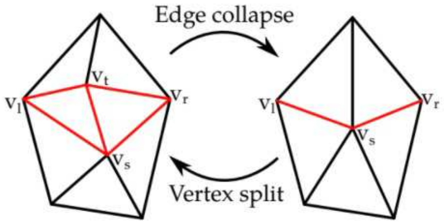
Google Draco mesh compression

Google has released Draco encoders and decoders in open source

Source code: https://github.com/google/draco

MPEG Activities

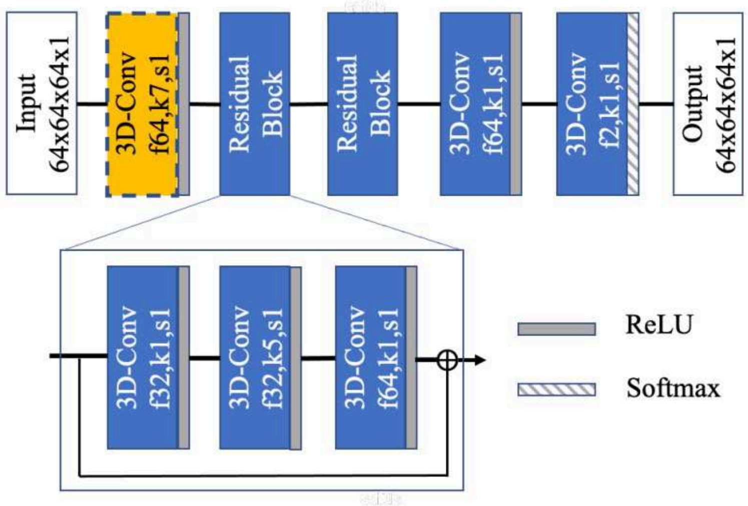
Mesh coding with V-PCC



Rossignac, Jarek. "**Edgebreaker: Connectivity compression for triangle meshes**." IEEE transactions on visualization and computer graphics 5.1 (1999): 47-61.
Maglo, Adrien, et al. "**3d mesh compression: Survey, comparisons, and emerging trends.**" ACM Computing Surveys (CSUR) 47.3 (2015): 1-41.

Volumetric video coding

Based on deep learning



VoxelDNN network architecture

- Outperforms MPEG G-PCC with average 28% rate savings over all tested datasets

Nguyen, Dat Thanh, Maurice Quach, Giuseppe Valenzise, and Pierre Duhamel. "**Learning-based lossless compression of 3d point cloud geometry**." In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.

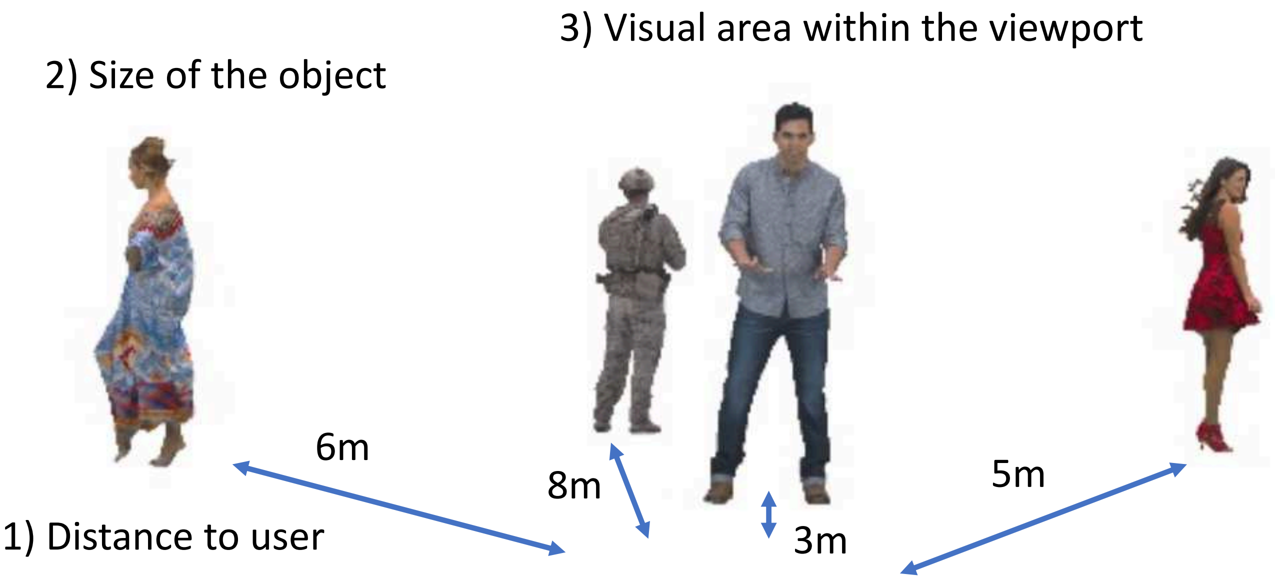
Volumetric video streaming

* Figure has been taken from presentation slides at MM 2019: Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression

van der Hooft, Jeroen, et al. "**Towards 6dof http adaptive streaming through point cloud compression**." Proceedings of the 27th ACM International Conference on Multimedia. 2019.

Volumetric video streaming

Bitrate-allocation

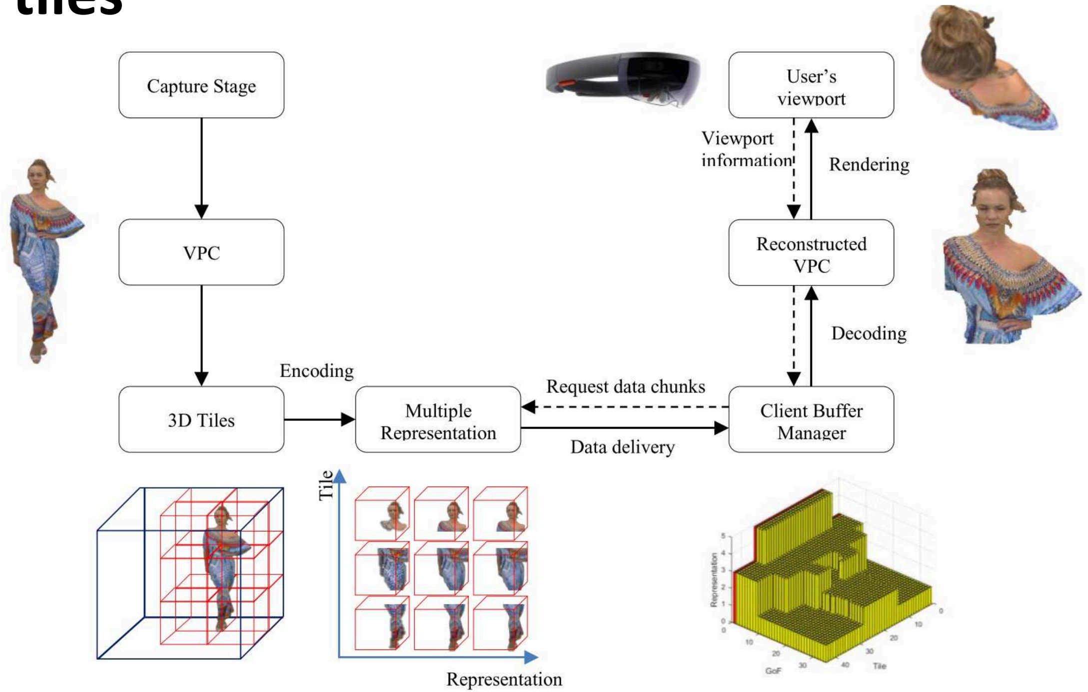


* Figure has been taken from presentation slides at MM 2019: Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression

van der Hooft, Jeroen, et al. "**Towards 6dof http adaptive streaming through point cloud compression**." Proceedings of the 27th ACM International Conference on Multimedia. 2019.

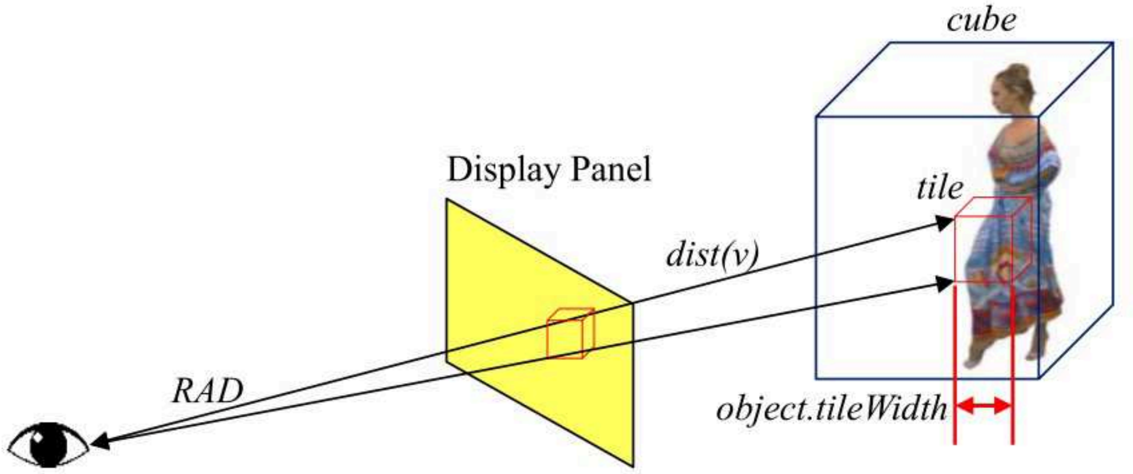
Volumetric video streaming

Based on 3D tiles



Park, Jounsup, Philip A. Chou, and Jenq-Neng Hwang. "**Rate-utility optimized streaming of volumetric media for augmented reality**." IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9.1 (2019): 149-162.

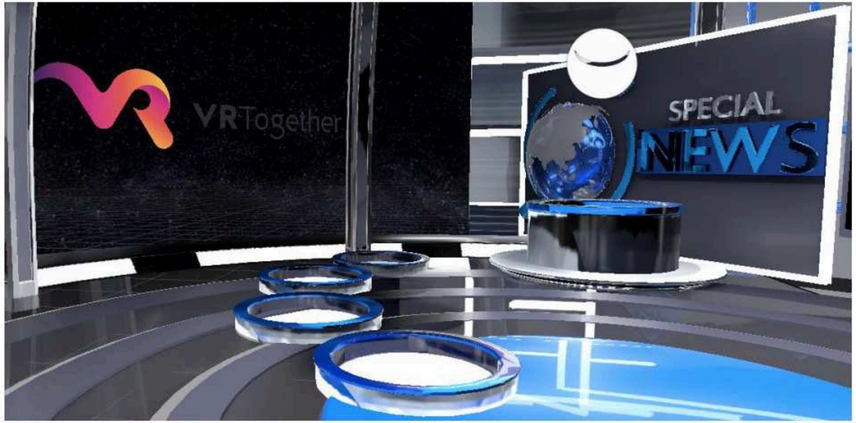
Volumetric video streaming

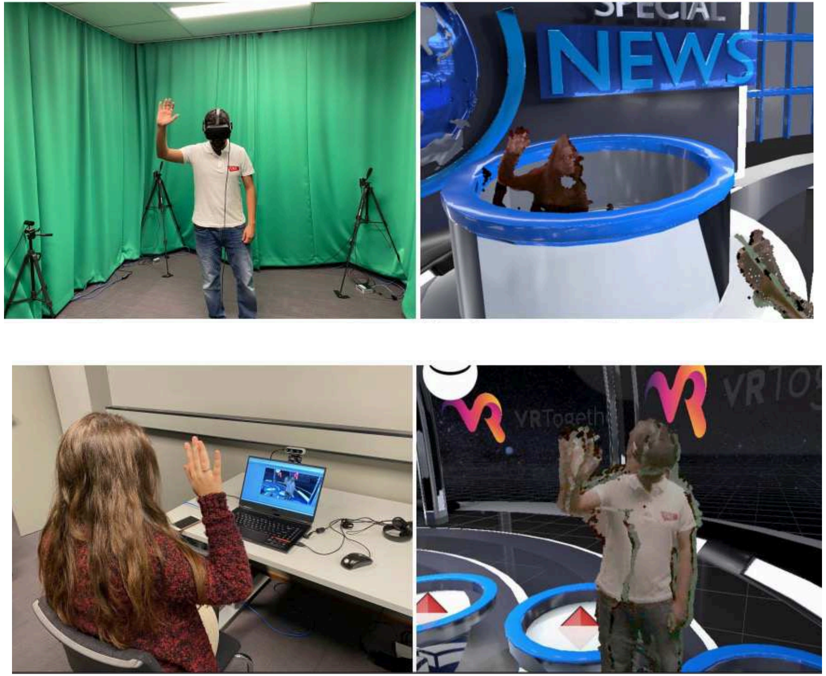


Park, Jounsup, Philip A. Chou, and Jenq-Neng Hwang. "**Rate-utility optimized streaming of volumetric media for augmented reality**." IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9.1 (2019): 149-162.

Volumetric video streaming

Video conferencing scenario





Demo:
https://www.youtube.com/watch?v=noL9pc4OzFY

Jansen, Jack, et al. "**A pipeline for multiparty volumetric video conferencing: transmission of point clouds over low latency DASH.**" *Proceedings of the 11th ACM Multimedia Systems Conference*. 2020.

Volumetric video streaming

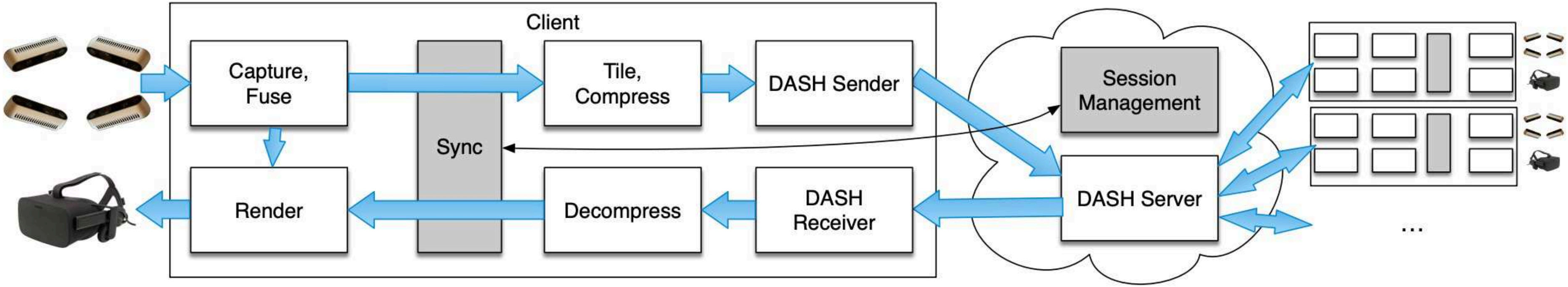


Figure: Architecture of the developed video conferencing system

Jansen, Jack, et al. "**A pipeline for multiparty volumetric video conferencing: transmission of point clouds over low latency DASH.**" *Proceedings of the 11th ACM Multimedia Systems Conference*. 2020.

Volumetric video streaming

A joint communication and computational resources allocation framework for point cloud video streaming.

QoE evaluation metric based on distance between the user and the scene, and quality of each tile.

Optimizing the resources in the system and maximize the QoE.

Li, Jie, et al. "**Joint Communication and Computational Resource Allocation for QoE-driven Point Cloud Video Streaming.**" arXiv preprint arXiv:2001.01403 (2020).

Summary

Brief overview of content delivery systems for traditional digital imaging

Content delivery mechanisms, i.e.:

- Compression
- Streaming

Various immersive imaging modalities:

- Light fields
- Omnidirectional imaging
- Volumetric videos

Next part:

- How are the immersive imaging data rendered and displayed at the receiver side?

Part IV: Rendering and Display Technologies

What are the rendering and display technologies to visualize the captured immersive media? What kind of devices are used?

Part IV: Rendering and Display Technologies

Rendering algorithms

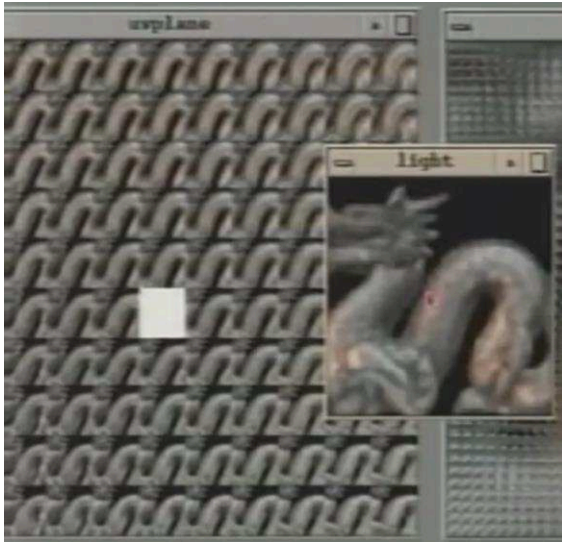
Two main display methods:

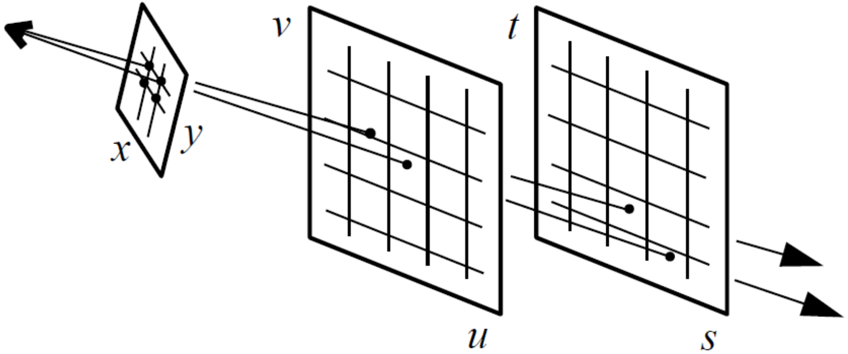
- Traditional planar displays
  - Light field displays
  - AR devices
- Head-mounted displays (HMDs)
  - HMDs for VR
  - HMDs for AR

Rendering algorithms

Viewport rendering from light field

- Quadra-linear interpolation





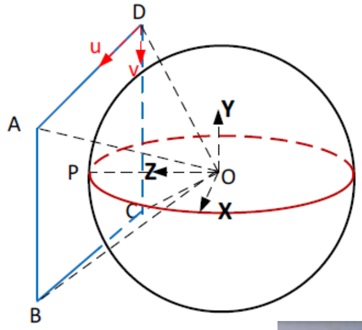
Levoy, Marc, and Pat Hanrahan. "**Light field rendering**." *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996.

Gortler, Steven J., et al. "**The lumigraph**." *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996.

Rendering algorithms

Viewport rendering from omnidirectional images / videos

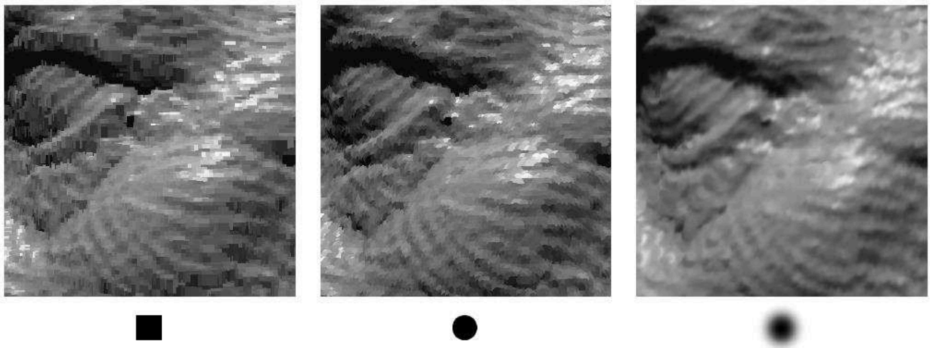
- Project viewport onto sphere
- Use sphere projection to project into image

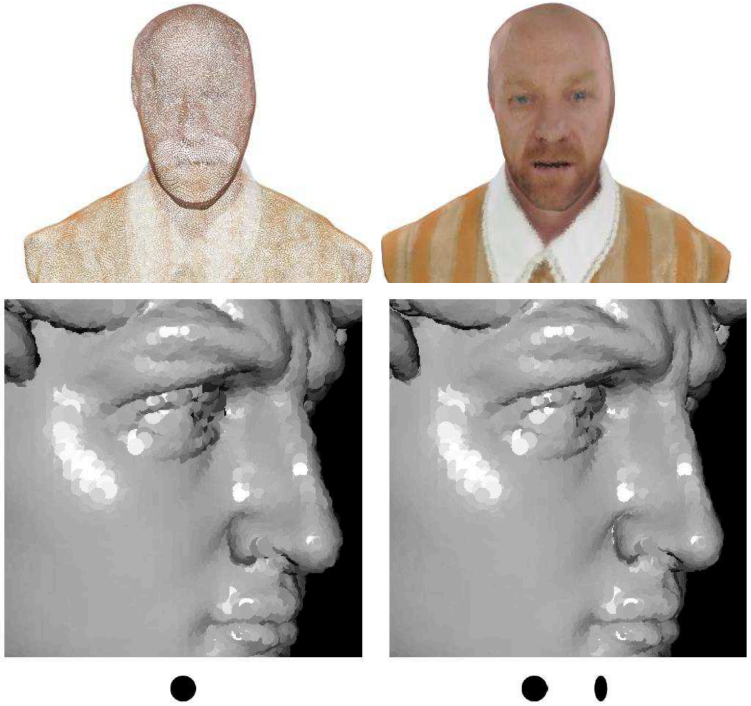


Rendering algorithms

Viewport rendering from point cloud

- Partition the point cloud, e.g. using a kd-tree
- Select points which are visible
- Draw point, e.g. using splatting, optionally using point normal to orient the splat shape

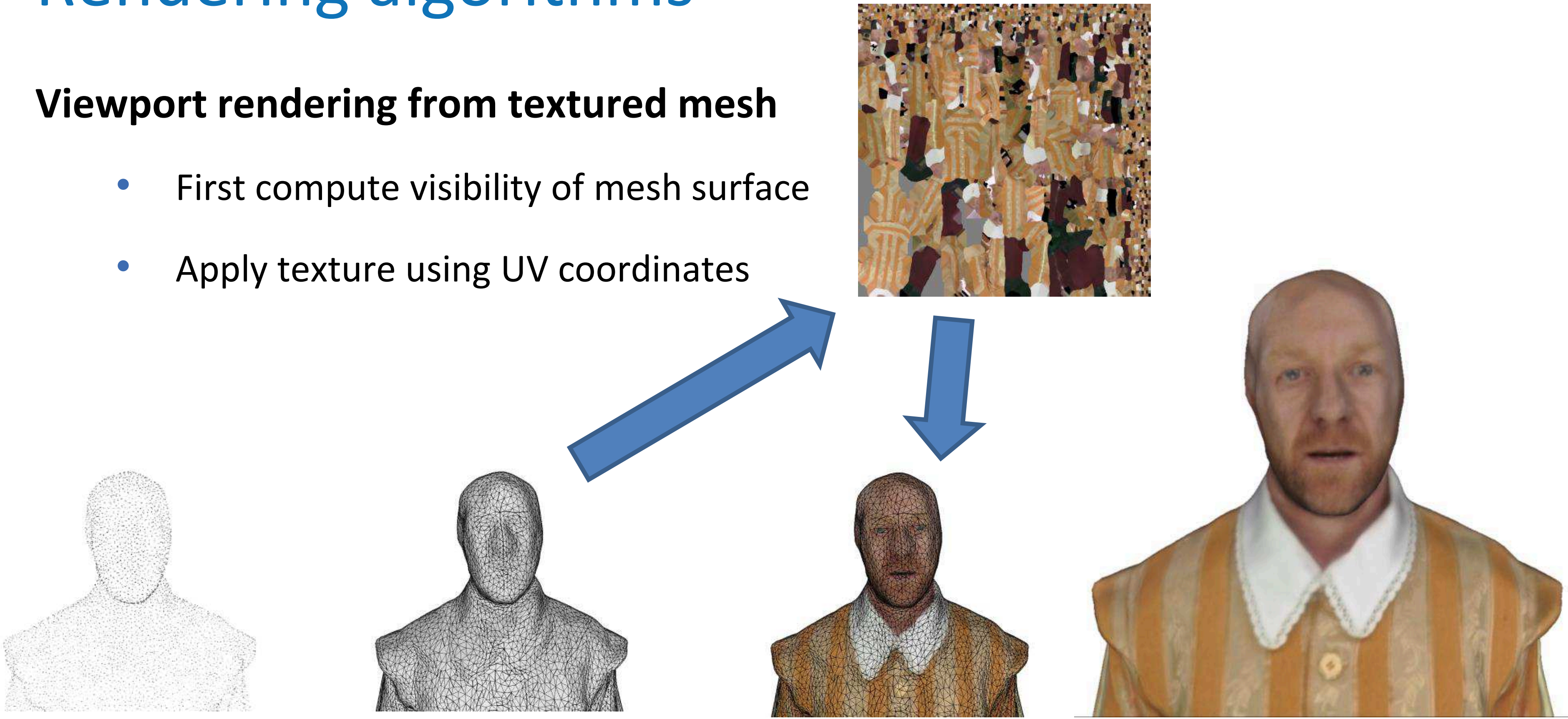




Rendering algorithms

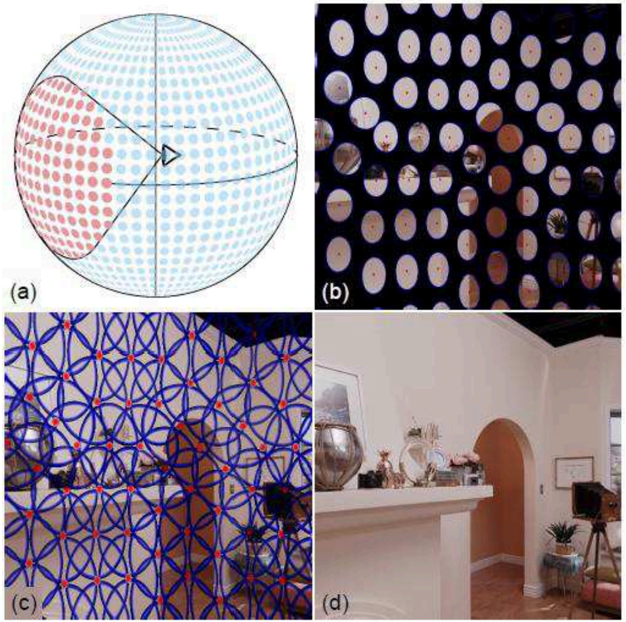
Viewport rendering from textured mesh

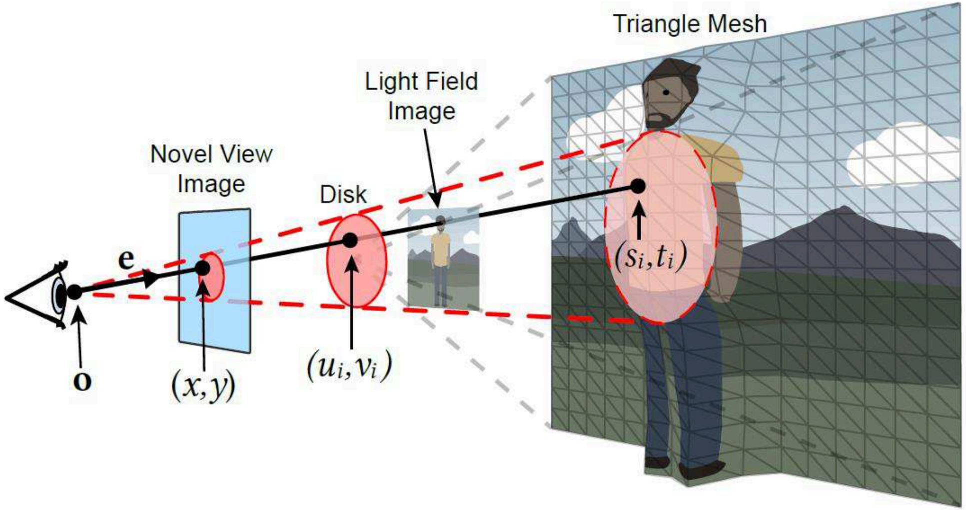
- First compute visibility of mesh surface
- Apply texture using UV coordinates



Rendering algorithms

Viewport rendering from omnidirectional light field using mesh triangulation

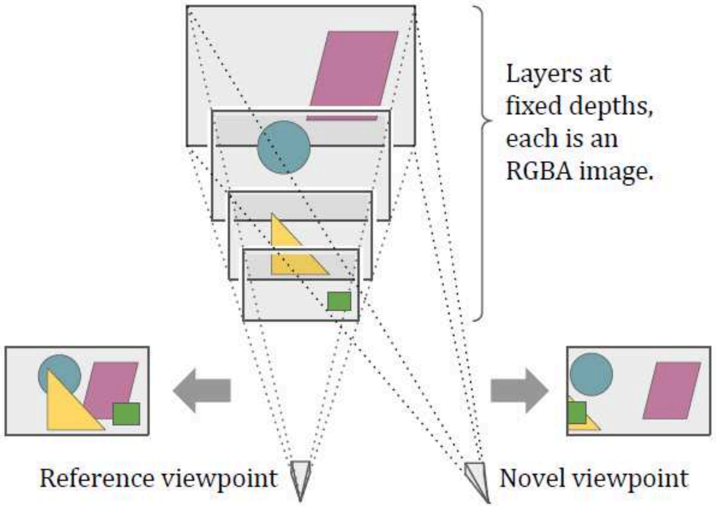


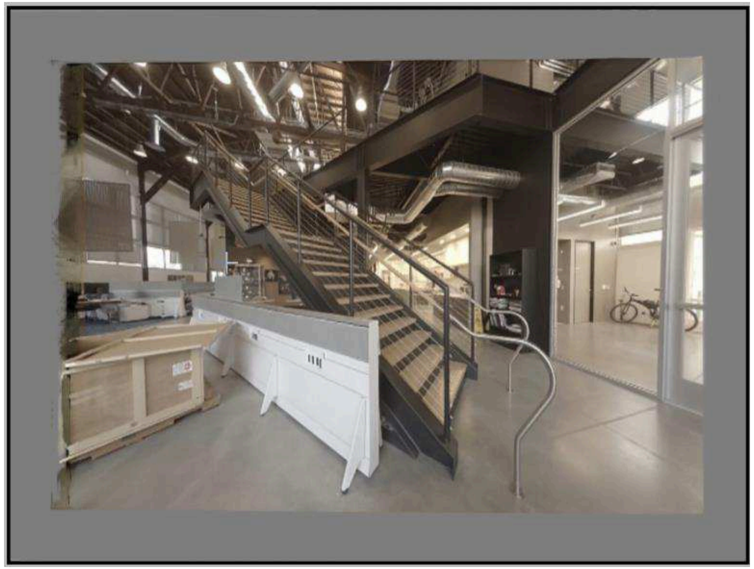


Overbeck, Ryan S., et al. "**A system for acquiring, processing, and rendering panoramic light field stills for virtual reality**." *ACM Transactions on Graphics (TOG)* 37.6 (2018): 1-15.

Rendering algorithms

Viewport rendering from light field using MPIs/MSIs

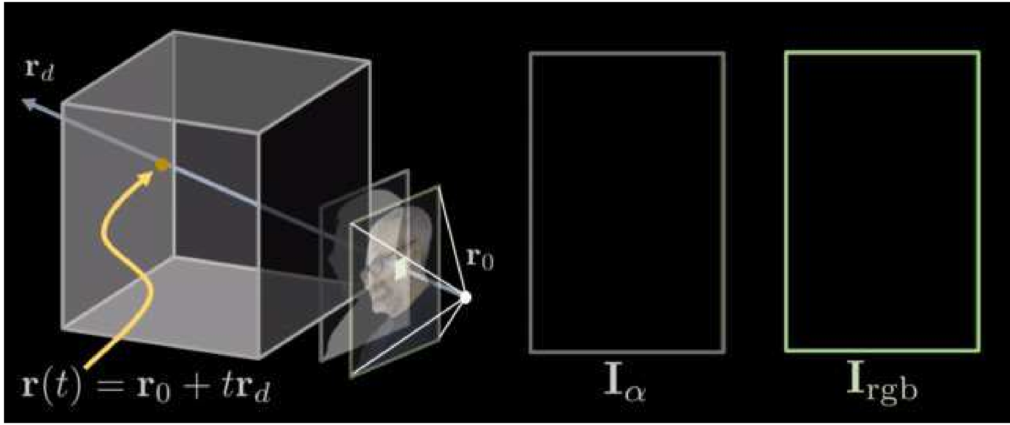


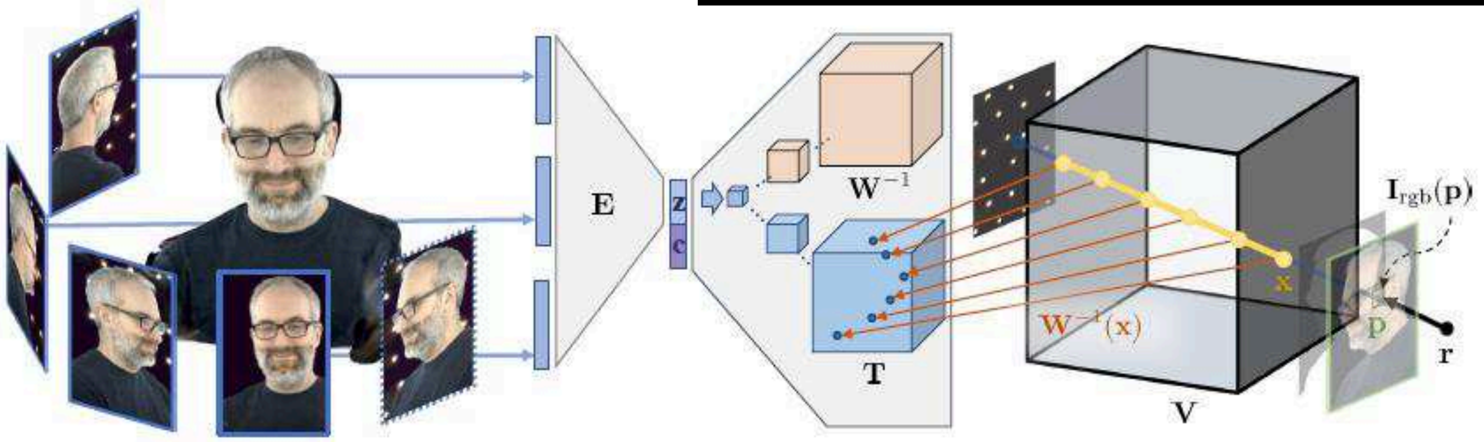


Zhou, Tinghui, et al. "**Stereo magnification: learning view synthesis using multiplane images.**" *ACM Transactions on Graphics (TOG)* 37.4 (2018): 1-12.
Flynn, John, et al. "**Deepview: View synthesis with learned gradient descent.**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
Srinivasan, Pratul P., et al. "**Pushing the boundaries of view extrapolation with multiplane images.**" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
Mildenhall, Ben, et al. "**Local light field fusion: Practical view synthesis with prescriptive sampling guidelines**." *ACM Transactions on Graphics (TOG)* 38.4 (2019): 1-14.
Broxton, Michael, et al. "**Immersive light field video with a layered mesh representation.**" *ACM Transactions on Graphics (TOG)* 39.4 (2020): 86-1.

Rendering algorithms

Deep 3D Capture

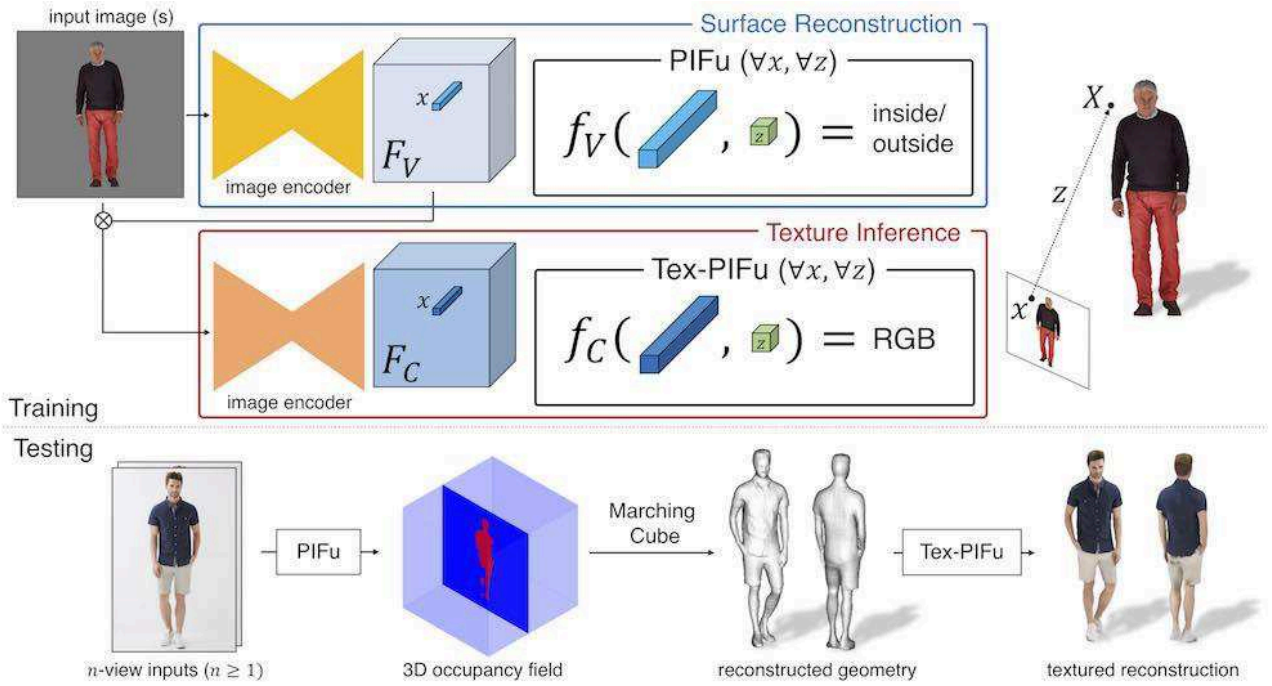




Lombardi, Stephen, et al. "**Neural volumes: Learning dynamic renderable volumes from images.**" *ACM Transactions on Graphics (TOG)* 38.4 (2019): 65.

Rendering algorithms

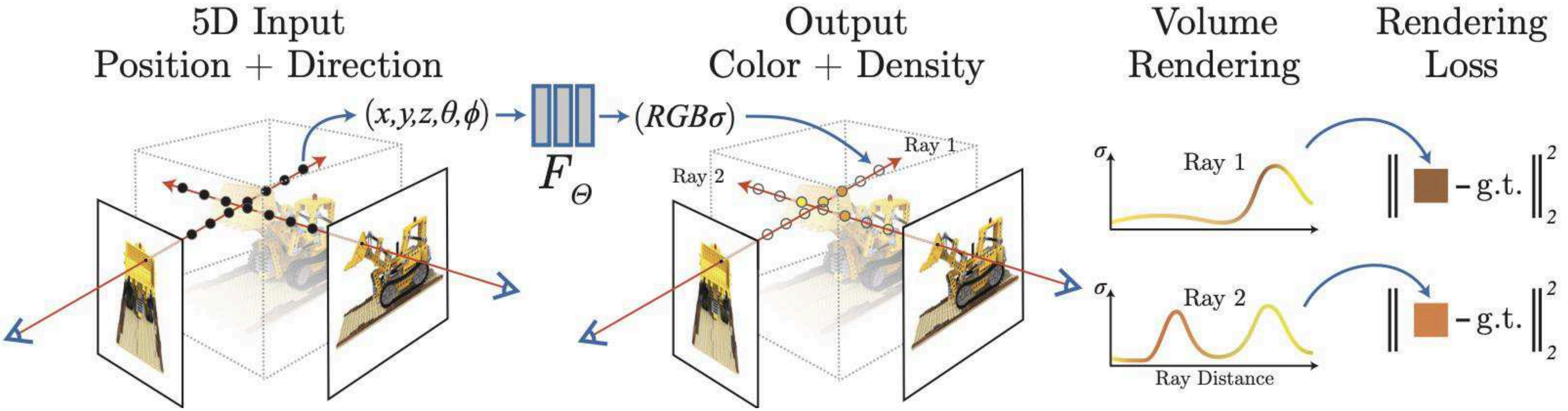
PiFU



Saito Shunsuke, Huang Zeng, Natsume Ryota, et al. "**PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization**." *ICCV* 2019.

Rendering algorithms

NeRF



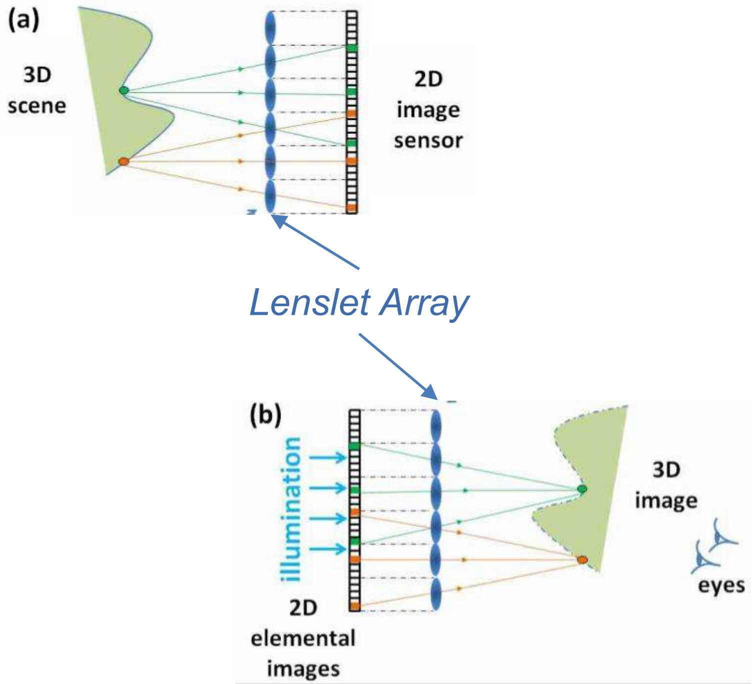
$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t),\mathbf{d})dt\,, \text{ where } T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right). \quad (1)$$

Mildenhall, Ben, et al. "**NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis**." In *European Conference on Computer Vision,* August 2020.

Immersive Imaging on 2D Screens

Light field displays

- Integral imaging
  - The reverse operation of a plenoptic camera such as Lytro



Immersive Imaging on 2D Screens

Light field displays

- Holographic display
- Looking Glass (Portrait, Gen2)

Attribution: lookingglassfactory.com/tech

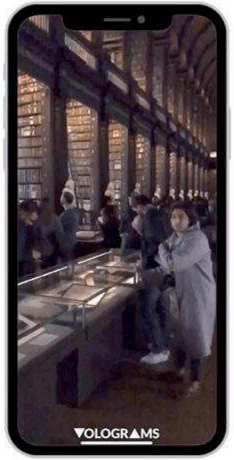
Attribution: Volograms

Immersive Imaging on 2D Screens

Devices to run AR

- Smartphones
- Tablets

- Pros: Affordable, easily accessible
- Cons: Very limited Field-of-View



Head-Mounted Displays

Head Mounted Displays

- Displays that can be worn on the head which are mainly used for virtual reality (VR) and augmented reality (AR) applications.
- VR replaces and covers our reality --- AR adds on top of our reality

Head-Mounted Displays for Virtual Reality

Head Mounted Displays

- Virtual reality headsets block the light coming from outside, to replace our reality
- Mostly rely on remote controllers to interact with the virtual world

Head-Mounted Displays for Virtual Reality

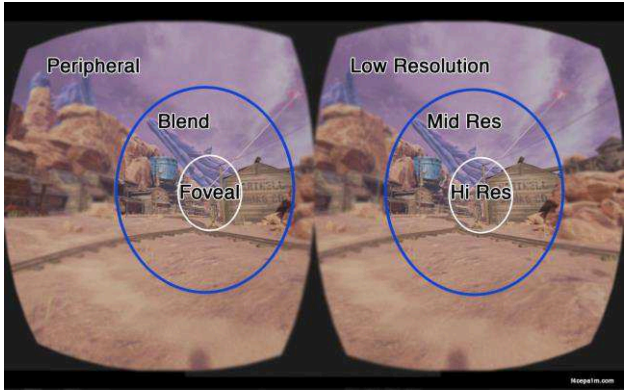
HMDs for VR

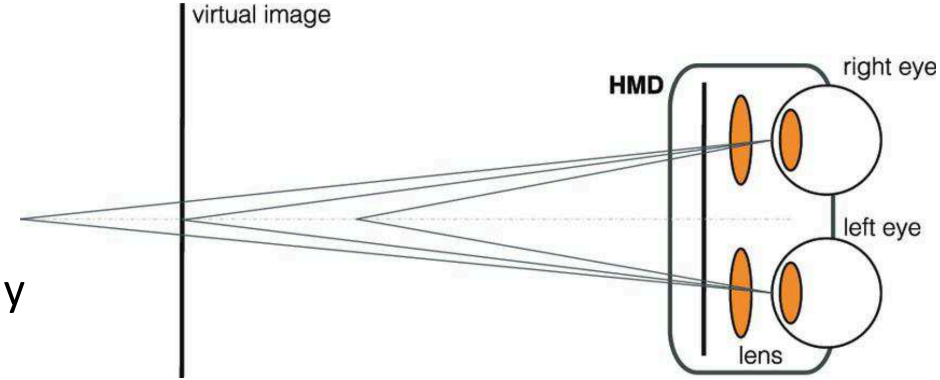
- Oculus Rift S
- Oculus Quest 2
- HTC Vive Pro2,
- HTC Vive Flow
- PlayStation VR
- Valve Index
- Others

Head-Mounted Displays for Virtual Reality

HMDs for VR

- Head tracking w/ sensors or computer vision techniques
- Stereoscopic 3D display
- High frame rate ($\geq$ 90Hz)
- Eye-tracking
  - Foveated rendering
- Lenses map the up-close display to a wide field of view

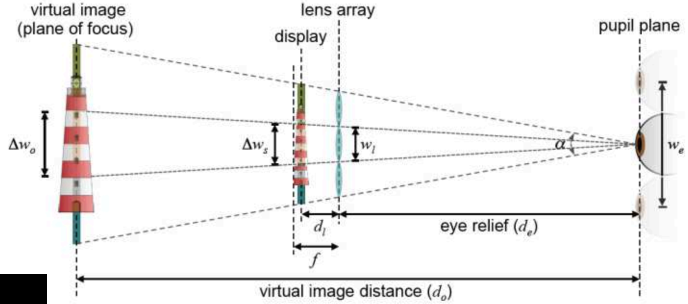


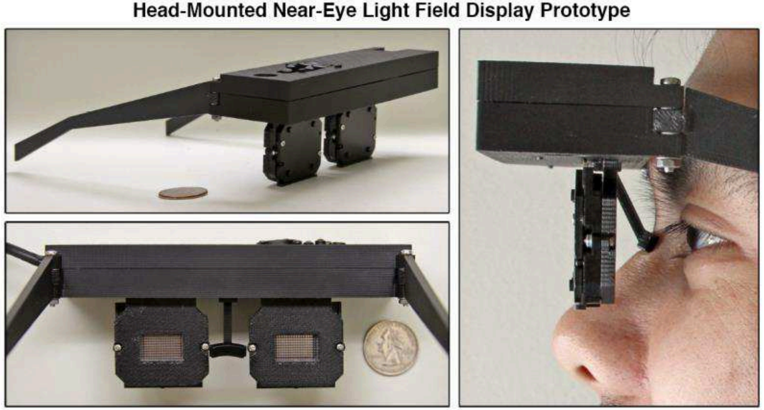


Head-Mounted Displays for Virtual Reality

Near-eye light field displays

- Similar to integral imaging

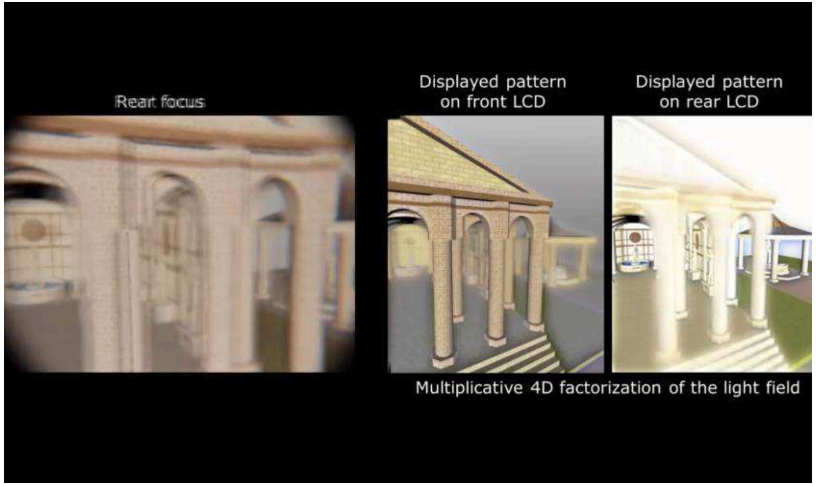


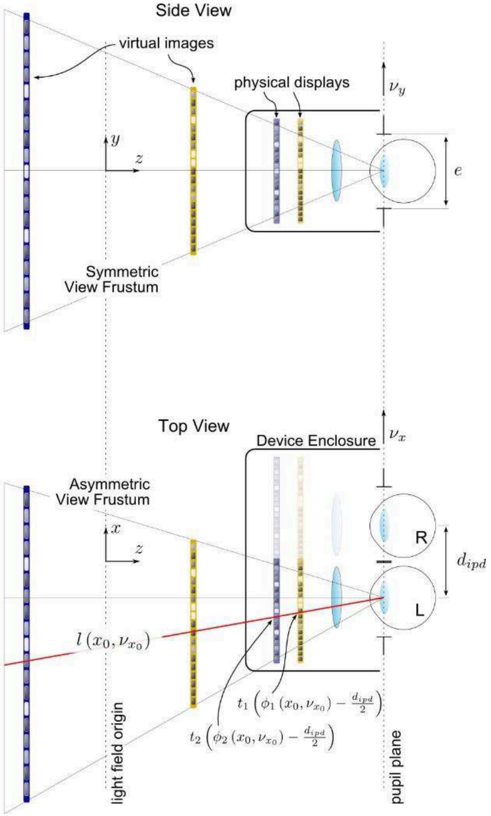


Lanman, Douglas, and David Luebke. "**Near-eye light field displays**." *ACM Transactions on Graphics (TOG)* 32.6 (2013): 1-10.

Head-Mounted Displays for Virtual Reality

Near-eye light field displays





Huang, Fu-Chung, David P. Luebke, and Gordon Wetzstein. "**The light field stereoscope**." *SIGGRAPH Emerging Technologies*. 2015.

Head-Mounted Displays for Augmented Reality

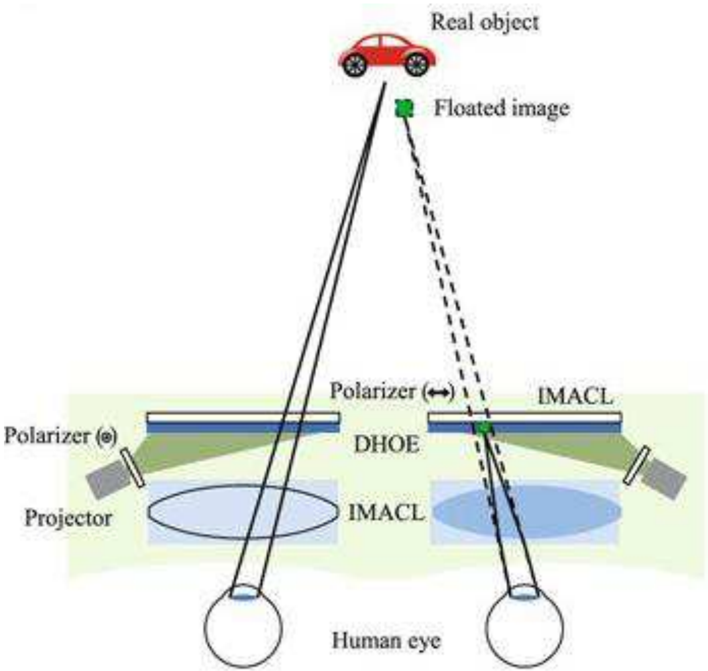
Head Mounted Displays

- Augmented reality headsets do NOT block the light coming from outside, they add on top of our reality
- They use either remote controllers or hand gestures to interact with the system

Head-Mounted Displays for Augmented Reality

HMDs for AR

- Head tracking relative to the environment
- Translucent screen where virtual content is projected



Hong, Jong-Young, et al. "**See-through optical combiner for augmented reality head-mounted display: index-matched anisotropic crystal lens**." *Scientific reports* 7.1 (2017): 1-11.

Head-Mounted Displays for Augmented R

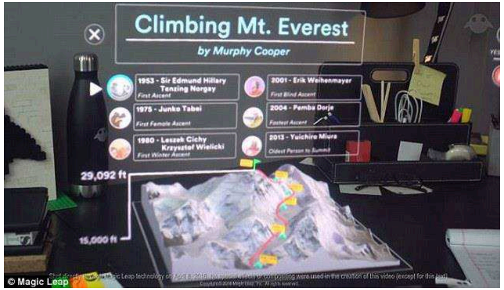
HMDs for AR

- HoloLens 1
- HoloLens 2
- Magic Leap One
- Magic Leap Two
- Epson Moverio
- Vuzix



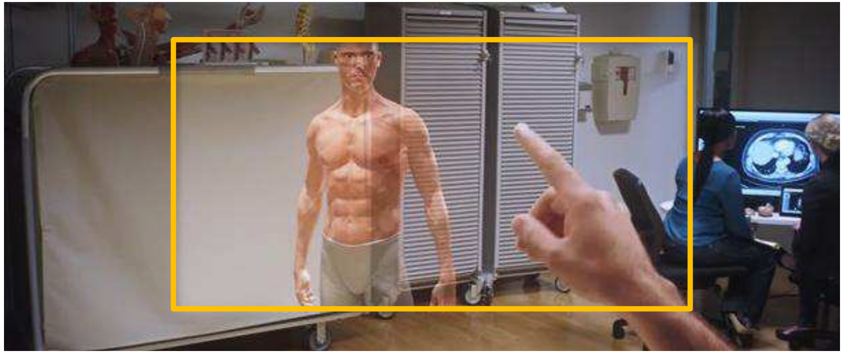



Head-Mounted Displays for Augmented Reality



Attribution: © Magic Leap, 2020

Kim, Kangsoo, Gerd Bruder, and Greg Welch. "**Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality**." *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 2017.



Attribution: Youtube: Microsoft HoloLens: Partner Spotlight with Case Western Reserve University

Head-Mounted Displays for Augmented Reality

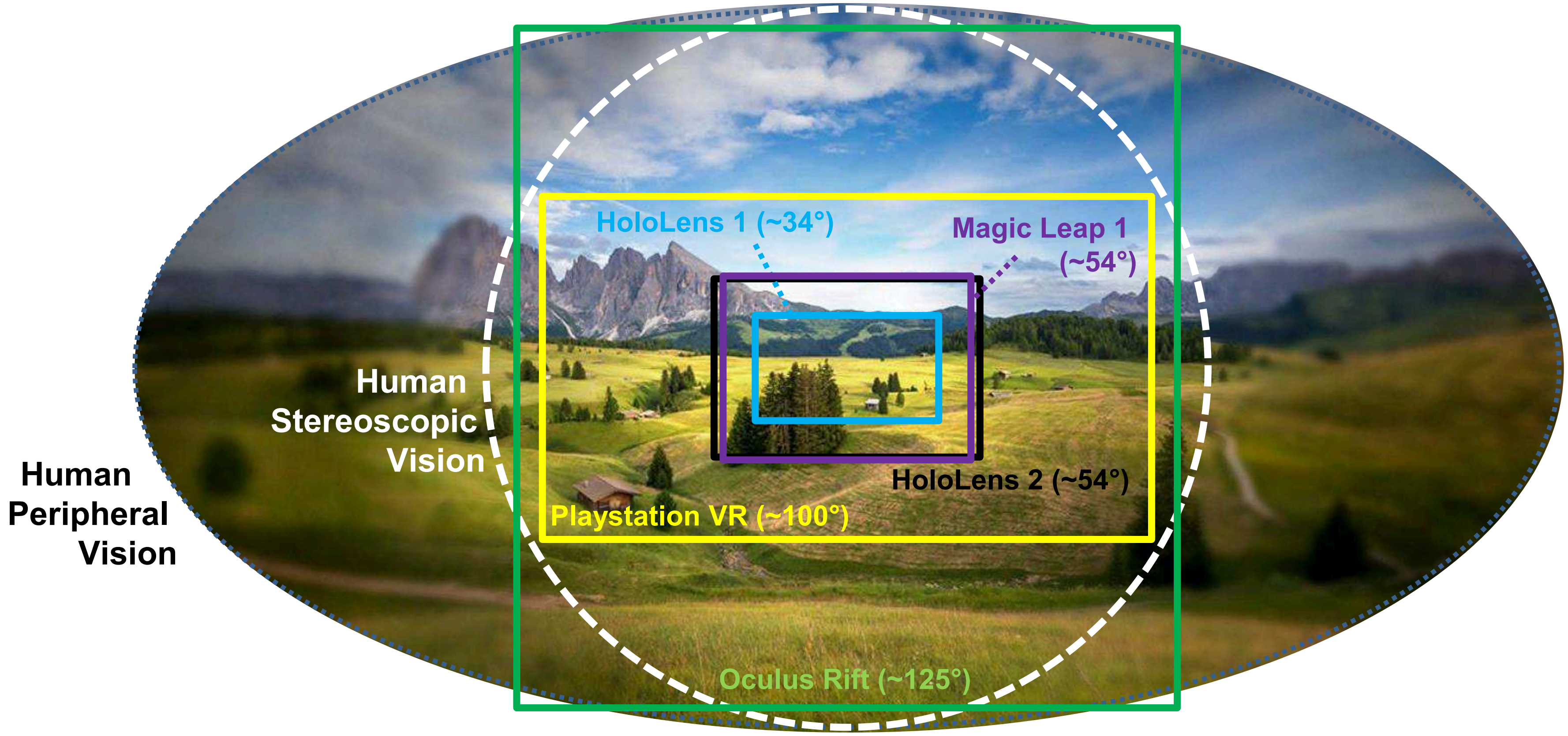
Sleek design.
Narrow field of view (similar to HoloLens 1)

More powerful
Heavier and bulkier

Increased FoV, eye tracking
Expensive

Easy to wear.
Low spatial resolution and FoV.

Head-Mounted Displays Comparison



Head-Mounted Displays Comparison

Head-Mounted Displays for VR/AR

Pass-through AR
"passes" outside view through

VR Mode
Does NOT pass the outside view through

Source: https://techcrunch.com/2017/11/21/apple-acquires-mixed-reality-headset-startup-vrvana-for-30m/

Summary: Rendering & Display

How immersive imaging data are rendered

How immersive imaging outputs are displayed

- 2D screens
- Head-mounted-displays
  - Virtual reality
  - Augmented reality

Next part:

- How are these immersive imaging data visually perceived? How can we assess their quality?

Part V: Perception & Quality Evaluation

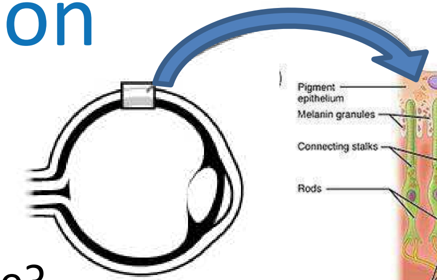
What are the relevant visual perception principles? How do we assess the quality of the processed media? How does visual attention change for immersive imaging?

Part V: Perception & Quality Evaluation

Three main points:

- Visual perception
- Quality assessment
  - Light fields
  - Omnidirectional imaging
  - Volumetric videos
- Visual attention
  - Light fields
  - Omnidirectional imaging
  - Volumetric videos

Visual Perception

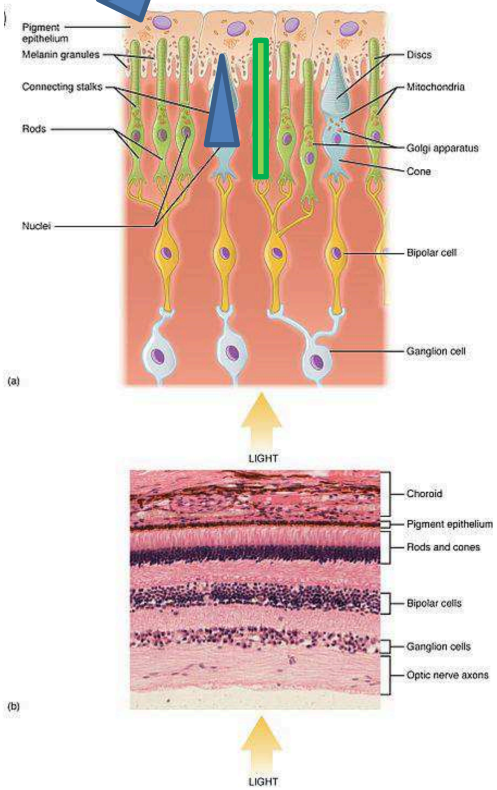


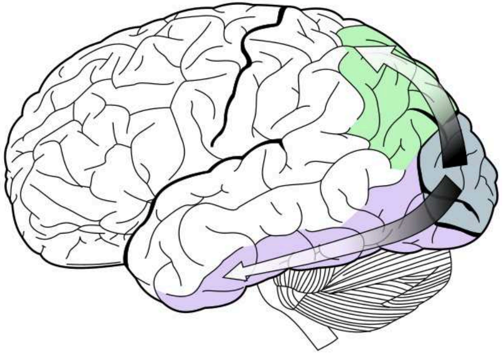
Human Visual System

- How do we see?
- How do we perceive?

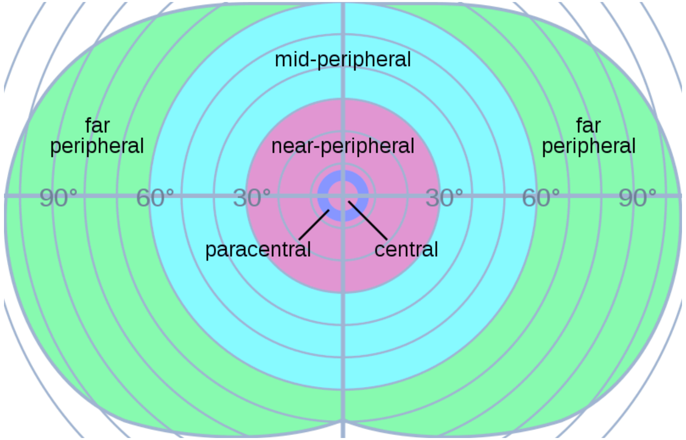
- Eye
  - Ganglion Cells
  - Bipolar Cells
  - Photoreceptors
  - Central Nervous System

- Brain
  - Primary visual cortex
  - Memory





Visual Perception

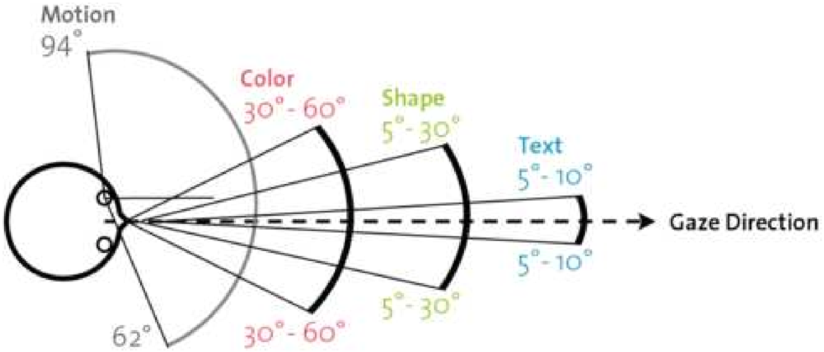


Attribution: "Field of view" by Zyxwv99, licenced under CC-BY-SA 3.0, https://commons.wikimedia.org/wiki/File:Field_of_view.svg

Visual angle

- Field of view
  - monocular vision and binocular vision
  - colour vision and peripheral vision

- Acuity and color perception changes



Attribution: "Visual Perception Human FOV" by D3kc1s, licenced under CC-BY-SA 4.0, https://en.wikipedia.org/wiki/File:Visual_Perception_Human_FOV.png

Visual Perception

Visual angle

- Field of view
  - monocular vision and binocular vision
  - colour vision and peripheral vision

- Acuity and color perception changes

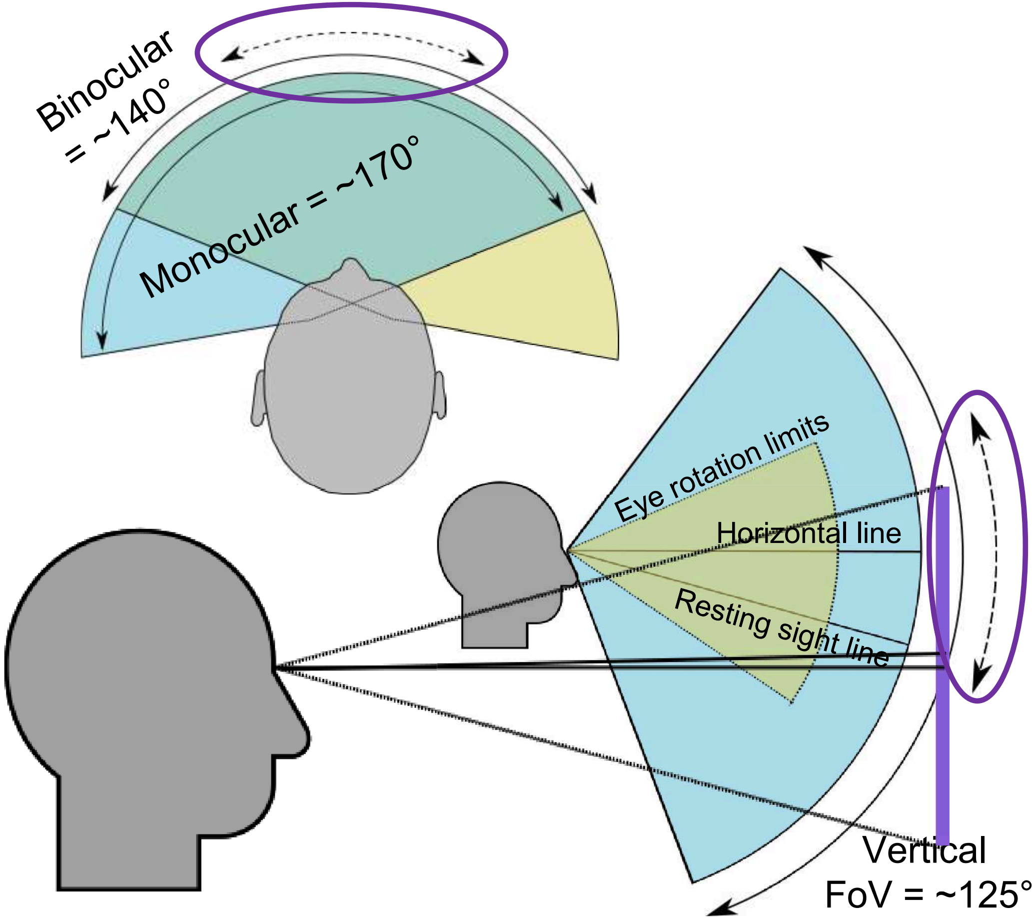
Visual Perception

Visual angle

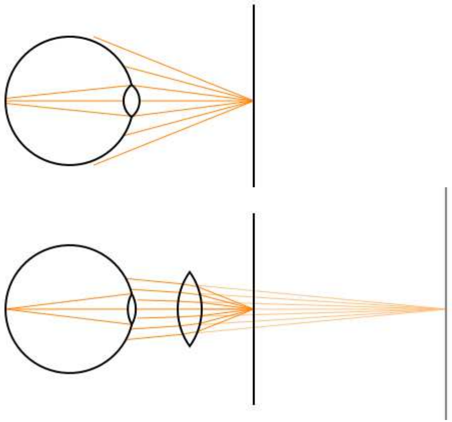
- Field of view
  - monocular vision and binocular vision
  - colour vision and peripheral vision

- Acuity and color perception changes

- Pixel-per-degree



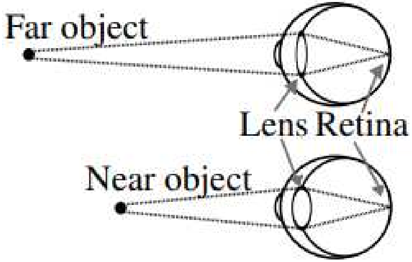
Visual Perception



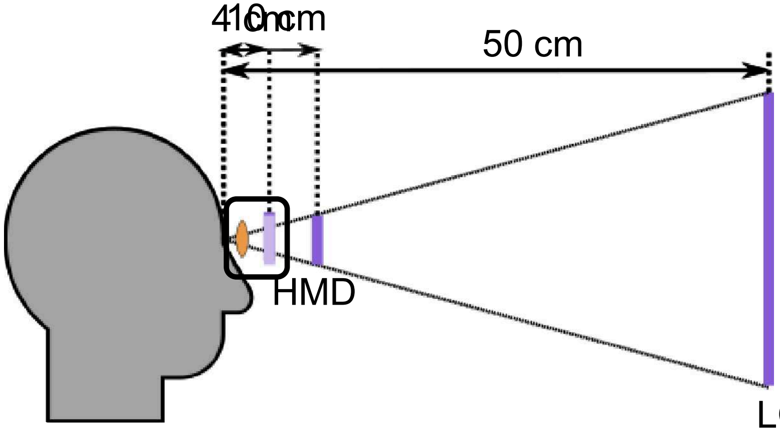
Attribution: "Virtual Screen" by Paulo Pacheco, https://xinreality.com/wiki/File:Virtual_Screen.png

Visual angle

- Pixel-per-degree
- Visual perception for head mounted displays
  - Resolution
  - Placement

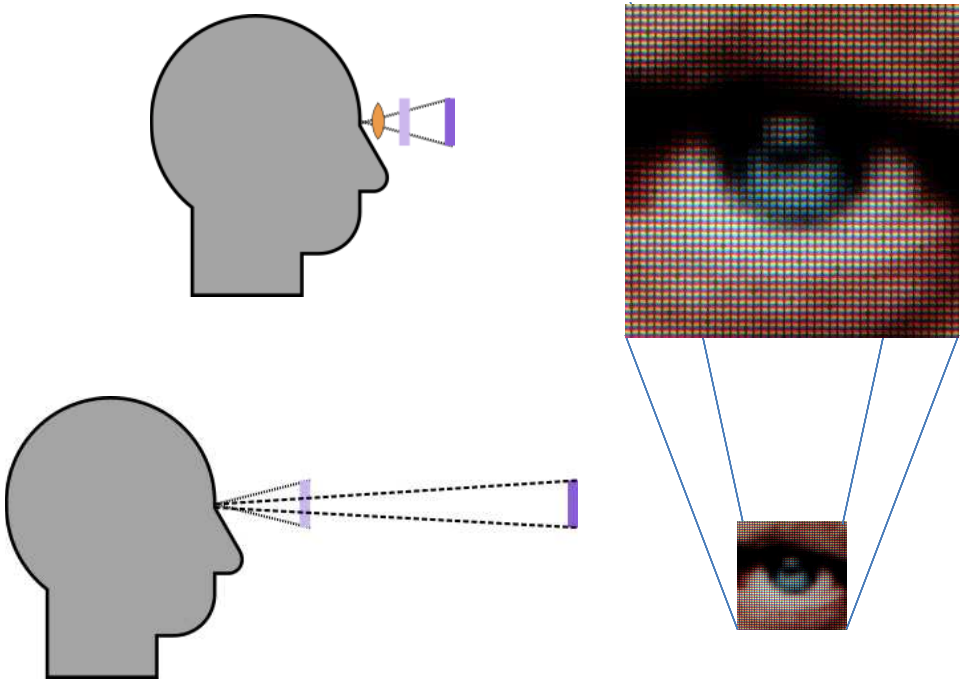


Accommodation

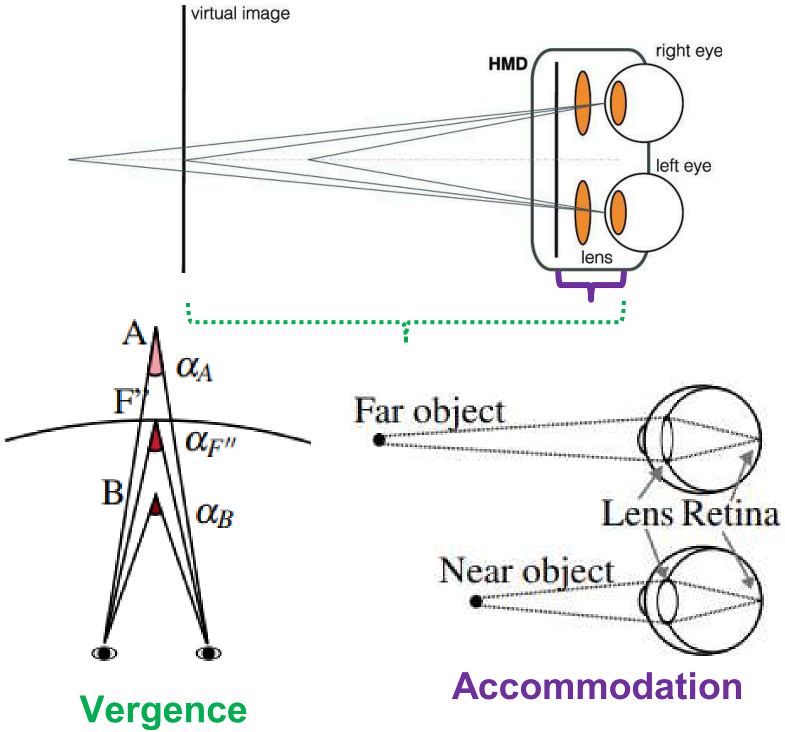


Visual Perception

Low Visual Angle (pixel per degree)



Vergence-Accommodation Conflict



Quality Assessment: General

Quality Assessment

- Necessary to ensure an adequate level of service and user satisfaction

- Definition of "Quality"
  - "set of inherent characteristics, we consider quality in terms of the evaluated excellence or goodness"

- Definition of "Quality of Experience"
  - "the degree of delight or annoyance of the user of an application or service"

Brunnström, Kjell, et al. "**Qualinet white paper on definitions of quality of experience.**" White Paper, *European Network on Quality of Experience in Multimedia Systems and Services* (2013).

Quality Assessment: General

Quality Assessment

- Necessary to ensure an adequate level of service and user satisfaction
- Definition of "Quality"
  - "set of inherent characteristics, we consider quality in terms of the evaluated excellence or goodness"
- Definition of "Quality of Experience"
  - "the degree of delight or annoyance of the user of an application or service"

Brunnström, Kjell, et al. "**Qualinet white paper on definitions of quality of experience**." White Paper, *European Network on Quality of Experience in Multimedia Systems and Services* (2013).

Subjective by definition

- Subjective quality assessment is the best way for quality assessment
  - Expensive – time, resources
  - Need expertise and instructor guidance

Quality Assessment: General

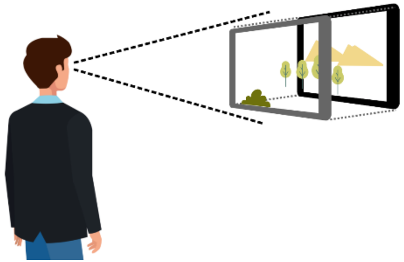
Objective Quality Estimation

- Developed to bypass the need for subjective quality assessment
- Works well if a specific type of distortions are considered (e.g., compression, streaming)

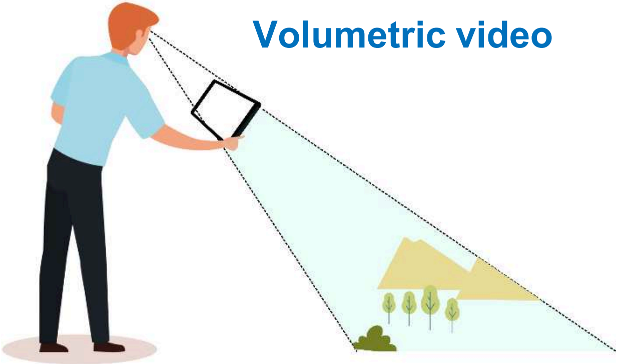
- Classified according to the presence of the availability of the undistorted reference
  - Full-reference
  - Reduced-reference
  - No-reference

Quality Assessment per Modality

Light Fields



360-degree video



Volumetric video

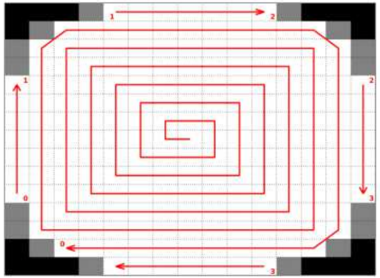
Quality Assessment: Light fields

Subjective quality assessment

- A lot of extra information
- Need new approaches for subjective quality assessment

- Two main approaches
  - Passive approach
    - A set of pre-rendered images is shown to the participants
    - An animation of selected renderings
  - Interactive approach
    - Participants can control the rendering

- Both approaches correlate with one another



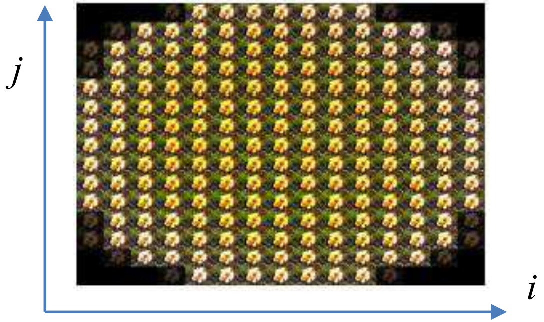
Matysiak, Pierre, et al. "**High quality light field extraction and post-processing for raw plenoptic data**." *IEEE Transactions on Image Processing* 29 (2020): 4188-4203.
Viola, Irene, Martin Řeřábek, and Touradj Ebrahimi. "**A new approach to subjectively assess quality of plenoptic content**." *Applications of Digital Image Processing XXXIX*. International Society for Optics and Photonics, 2016.
Viola, Irene, Martin Řeřábek, and Touradj Ebrahimi. "**Impact of interactivity on the assessment of quality of experience for light field content**." *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.

Quality Assessment: Light fields

Objective quality assessment

- Many sub-aperture images
- Practically goes through the same processing algorithm

- Using traditional image quality estimators and pooling (e.g. averaging) for all sub-aperture images
  - Peak-signal-to-noise-ratio (PSNR)
  - Structural similarity index (SSIM)
  - Visual information fidelity (VIF)



$$PSNR_Y(i,j) = 10\log_{10}\frac{255^2}{MSE_Y(i,j)}$$

$$PSNR_{YUV}(i,j) = \frac{6PSNR_Y + PSNR_U + PSNR_V}{8}$$

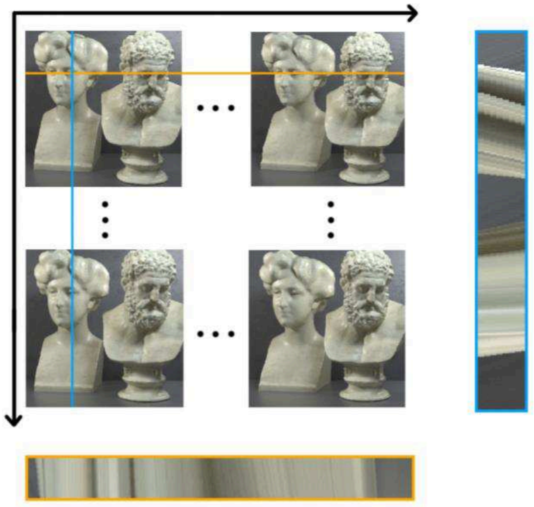
$$PSNR_{YUV_{mean}} = \frac{1}{(I-2)(J-2)}\sum_{i=2}^{I-1}\sum_{j=2}^{J-1} PSNR_{YUV}(i,j)$$

Viola, Irene, and Touradj Ebrahimi. "**Quality assessment of compression solutions for ICIP 2017 Grand Challenge on light field image coding**." *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2018.

Quality Assessment: Light fields

Objective quality assessment

- Epipolar Plane Image (EPI) representations
- Various techniques including finding features from gradients in the EPI representations, including:



- Local Binary Pattern features
- Log-Gabor features
- Convolutional Sparse Coding (CSC) features
- Histogram of Oriented Gradient (HoG) features

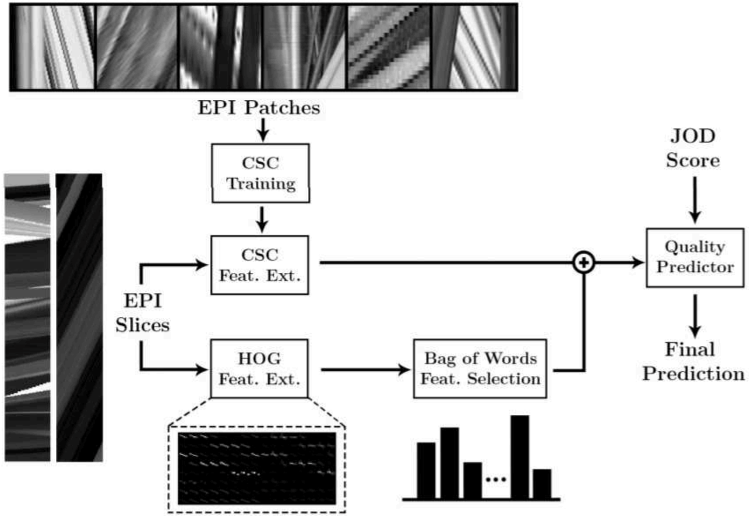


Figure: A sample No-Reference LF quality assessment metric based on CSC and HoG features with Support Vector Regression.

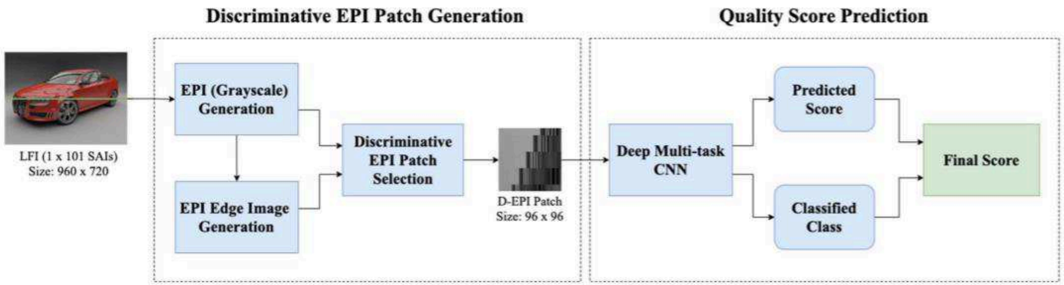
Ak, Ali, Suiyi Ling, and Patrick Le Callet. "**No-reference quality evaluation of light field content based on structural representation of the epipolar plane image**." *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. 2020.

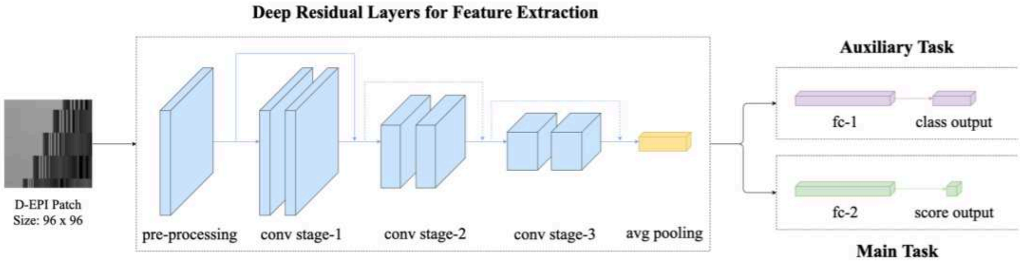
Quality Assessment: Light fields

Learning-based quality assessment

- Zhao et al. (2021) proposes a multi-task approach to no-reference LF quality assessment and EPI patch classification

- First, the Discriminative Epipolar Plane Image (D-EPI) patches are found
- These D-EPI patches are then fed into a deep multi-task CNN, which:
  - Classifies the input D-EPI patches with respect to the distortion types and severity of the distortion
  - Predicts the quality level of the input light field

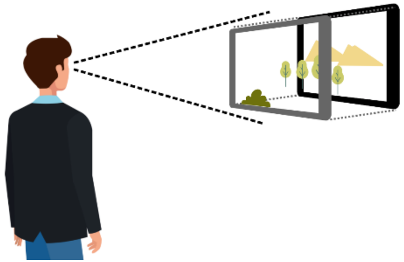




Zhao, Ping, et al. "**DeLFIQE—A Low-Complexity Deep Learning-Based Light Field Image Quality Evaluator**." *IEEE Transactions on Instrumentation and Measurement 70 (2021).*

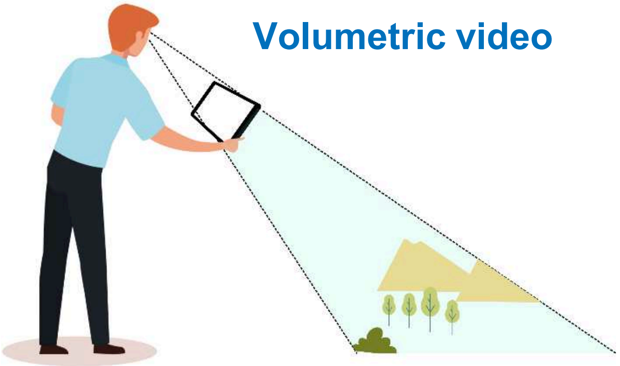
Quality Assessment per Modality

Light Fields



360-degree video

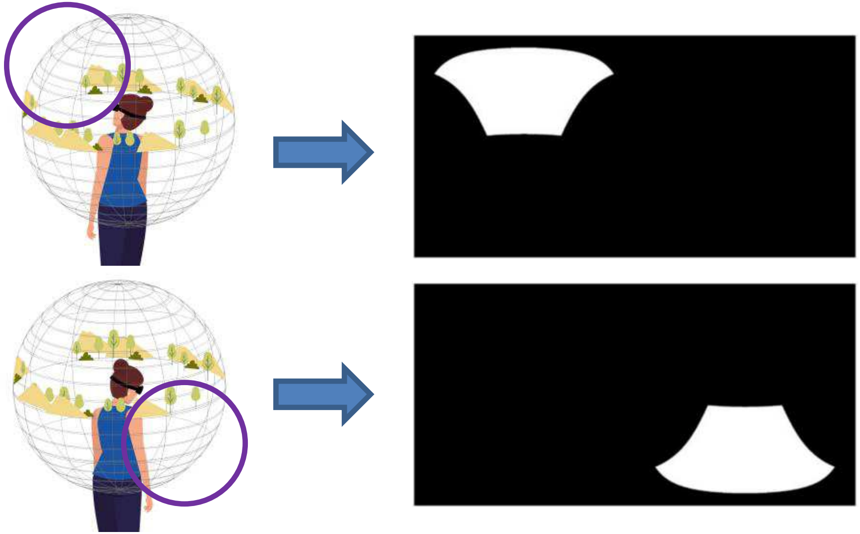
Volumetric video



Quality Assessment: Omnidirectional imaging

Subjective quality assessment

- The medium is spherical, but human field of view (and display devices' FoV) is limited.



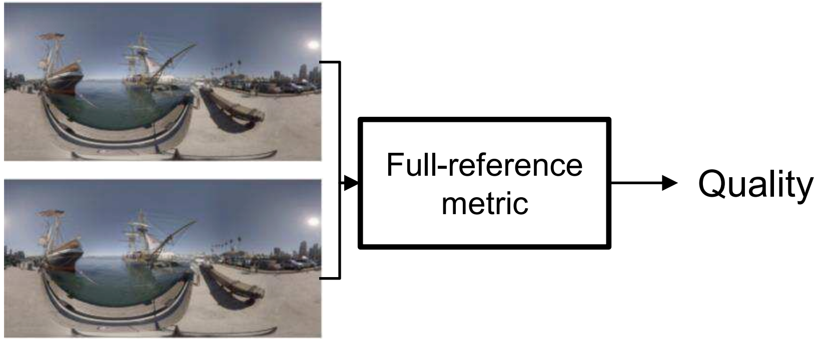
Current approaches

- Modified absolute category rating (M-ACR)

- Double Stimulus Impairment Scale (DSIS)

Singla, Ashutosh, et al. "**Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for UHD and FHD resolution.**" *Proceedings of the on Thematic Workshops of ACM Multimedia* 2017. 2017.
Singla, Ashutosh, Werner Robitza, and Alexander Raake. "**Comparison of Subjective Quality Test Methods for Omnidirectional Video Quality Evaluation.**" *IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2019.

Quality Assessment: Omnidirectional imaging



Objective quality assessment

- There are two approaches
  - Using traditional metrics on the projected image/video
  - Projecting the image/video to the sphere and computing on sphere

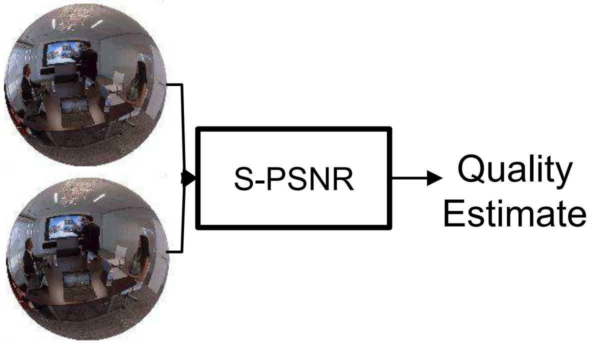
- Traditional metrics on equirectangular projected (ERP) or cubemap (CMP) data
  - PSNR
  - SSIM / MS-SSIM
  - VIF
  - VMAF

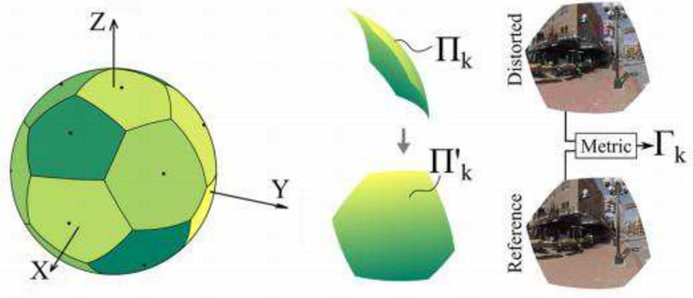
Quality Assessment: Omnidirectional imaging

Objective quality assessment

- There are two approaches
  - Using traditional metrics on the projected image/video
  - Projecting the image/video to the sphere and computing on sphere

- Sphere-based approaches
  - Spherical PSNR (S-PSNR)
  - Weighted Spherical PSNR (WS-PSNR)
  - Craster parabolic projection PSNR (CPP-PSNR)
  - Voronoi-cell-based metrics





Sun, Yule, Ang Lu, and Lu Yu. "**Weighted-to-spherically-uniform quality evaluation for omnidirectional video**." *IEEE Signal Processing Letters* 24.9 (2017): 1408-1412.

Zakharchenko, Vladyslav, Kwang Pyo Choi, and Jeong Hoon Park. "**Quality metric for spherical panoramic video**." *Optics and Photonics for Information Processing X*. International Society for Optics and Photonics, 2016.

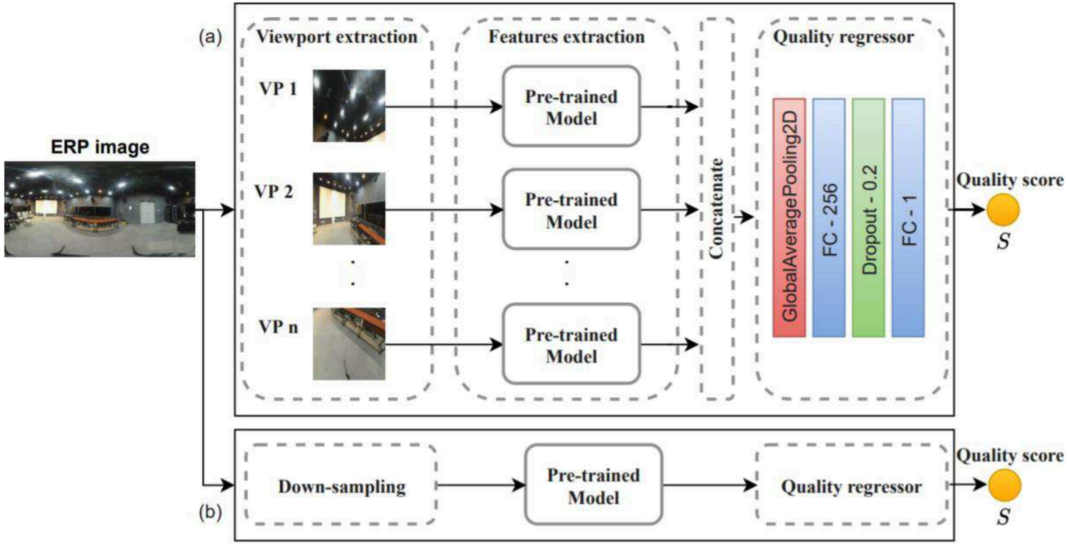
Yu, Matt, Haricharan Lakshman, and Bernd Girod. "**A framework to evaluate omnidirectional video coding schemes**." *IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2015.

Croci, Simone, et al. "**Voronoi-based objective quality metrics for omnidirectional video**." *Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019.

Quality Assessment: Omnidirectional imaging

Learning-based methods

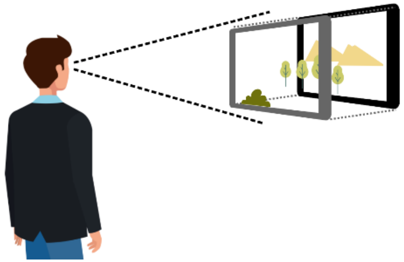
- Application of CNN quality estimators on various scenarios:
  - Patch-based
    - Patches are either selected on a grid from the given projection (e.g., ERP) or selected via an algorithm
  - Viewport-based
    - Viewports are selected using a prediction algorithm
  - Projection-based



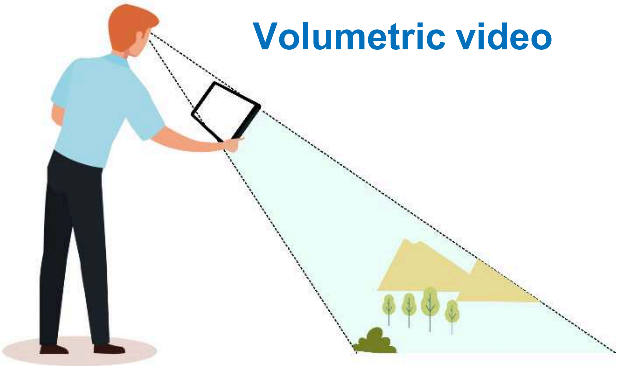
Sendjasni, Abderrezzaq, Mohamed-Chaker Larabi, and Faouzi Alaya Cheikh. "**Convolutional Neural Networks for Omnidirectional Image Quality Assessment: Pre-Trained or Re-Trained?**." *IEEE International Conference on Image Processing (ICIP)*. 2021.
Kim, Hak Gu, Heoun-Taek Lim, and Yong Man Ro. "**Deep virtual reality image quality assessment with human perception guider for omnidirectional image**." *IEEE Transactions on Circuits and Systems for Video Technology* 30.4 (2019): 917-928.

Quality Assessment per Modality

Light Fields



360-degree video

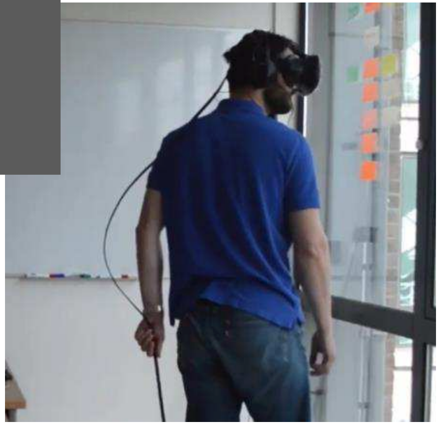


Volumetric video

Quality Assessment: Volumetric video

Subjective quality assessment

- There are infinitely many viewpoints to look from
- Similar to the light fields, there are two main approaches
  - Passive approach
    - Show a pre-rendered video with a selected trajectory
  - Interactive approach
    - Let the viewer select the viewpoint either using mouse or HMD

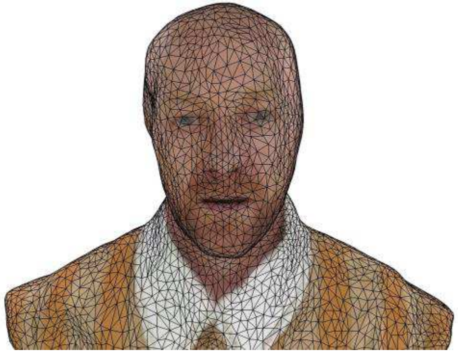


Zerman, Emin, et al. "**Textured Mesh vs Coloured Point Cloud: A Subjective Study for Volumetric Video Compression**.“ *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020.
Subramanyam, Shishir, et al. "**Comparing the Quality of Highly Realistic Digital Humans in 3DoF and 6DoF: A Volumetric Video Case Study**." *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2020.

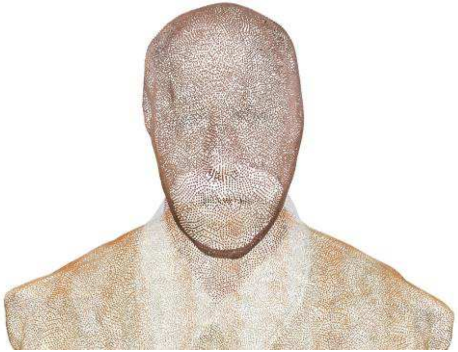
Quality Assessment: Volumetric video

Objective quality assessment

- Volumetric videos can be represented with two different representations:
  - Textured polygonal meshes
  - Coloured point clouds



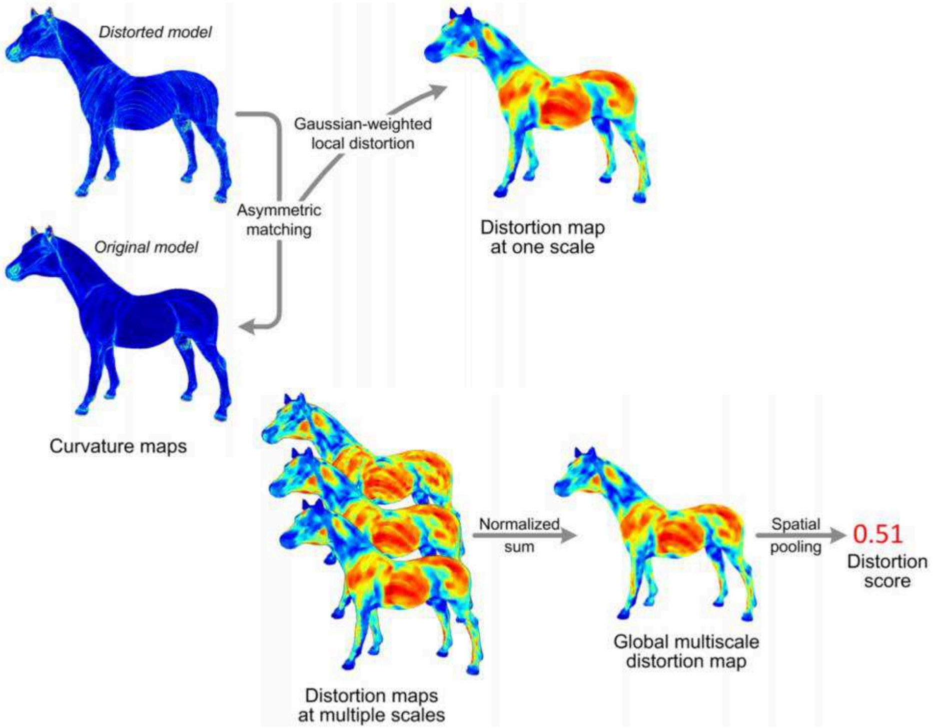
- Thus, objective quality assessment for volumetric videos depends on the data representation
  - Mesh metrics
  - Point cloud metrics
  - Representation-agnostic approach



Quality Assessment: Volumetric video

Mesh metrics

- A lot of works in the computer graphics field in the last two decades
- Mainly focuses on geometry errors
- They generally look at:
  - Positional errors – RMS distance
  - Curvature – MSDM2
  - Roughness – $3DWPM_1, 3DWPM_2$
  - Angles, etc.
- Joint models also exist

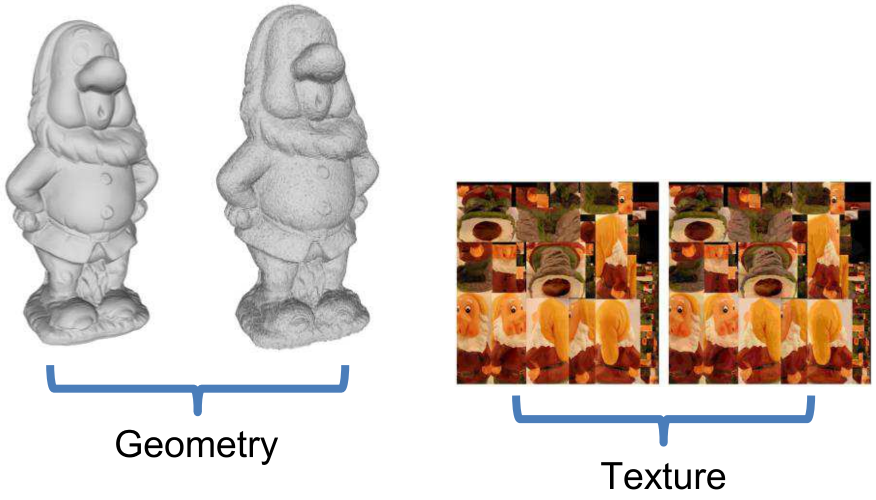


Lavoué, Guillaume. "**A multiscale metric for 3D mesh visual quality assessment**." *Computer Graphics Forum*. Vol. 30. No. 5. Oxford, UK: Blackwell Publishing Ltd, 2011.
Corsini, Massimiliano, et al. "**Perceptual metrics for static and dynamic triangle meshes**." *Computer Graphics Forum*. Vol. 32. No. 1. Oxford, UK: Blackwell Publishing Ltd, 2013.

Quality Assessment: Volumetric video

Mesh metrics

- A lot of works in the computer graphics field in the last two decades
- Mainly focuses on geometry errors
- They generally look at:
  - Positional errors – RMS distance
  - Curvature – MSDM2
  - Roughness – $3DWPM_1, 3DWPM_2$
  - Angles, etc.
- **Joint models also exist**



$$CM = \alpha Q_G + (1-\alpha) Q_T ,$$

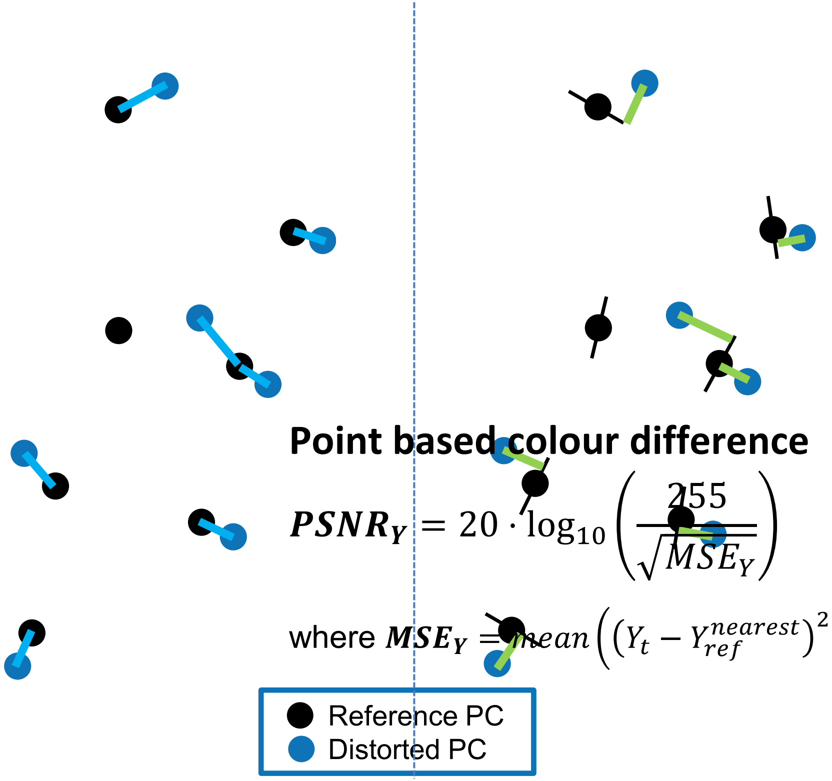
$$CM_1 = \alpha_1 MSDM2 + (1-\alpha_1) MS-SSIM$$
$$CM_2 = \alpha_2 SDCD + (1-\alpha_2) MS-SSIM$$

Guo, Jinjiang, et al. "**Subjective and objective visual quality assessment of textured 3D meshes**." *ACM Transactions on Applied Perception (TAP)* 14.2 (2016): 1-20.

Quality Assessment: Volumetric video

Point cloud metrics

- Mostly point-based computations
- Some of the state-of-the-art metrics
  - **Point-to-point geometry difference** –
  - Color difference (Point-to-point)
  - **Point-to-plane geometry difference** –
  - Plane-to-plane geometry difference
  - PC-MSDM
- Different pooling methods
  - Root mean square (RMS) distance
  - Mean squared error (MSE)
  - Hausdorff distance



Point based colour difference

$$\boldsymbol{PSNR_Y} = 20 \cdot \log_{10}\left(\frac{255}{\sqrt{MSE_Y}}\right)$$

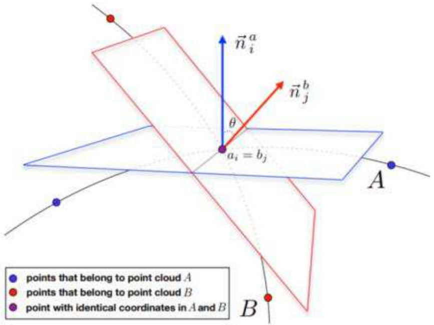
where $\boldsymbol{MSE_Y} = mean\left(\left(Y_t - Y_{ref}^{nearest}\right)^2\right)$

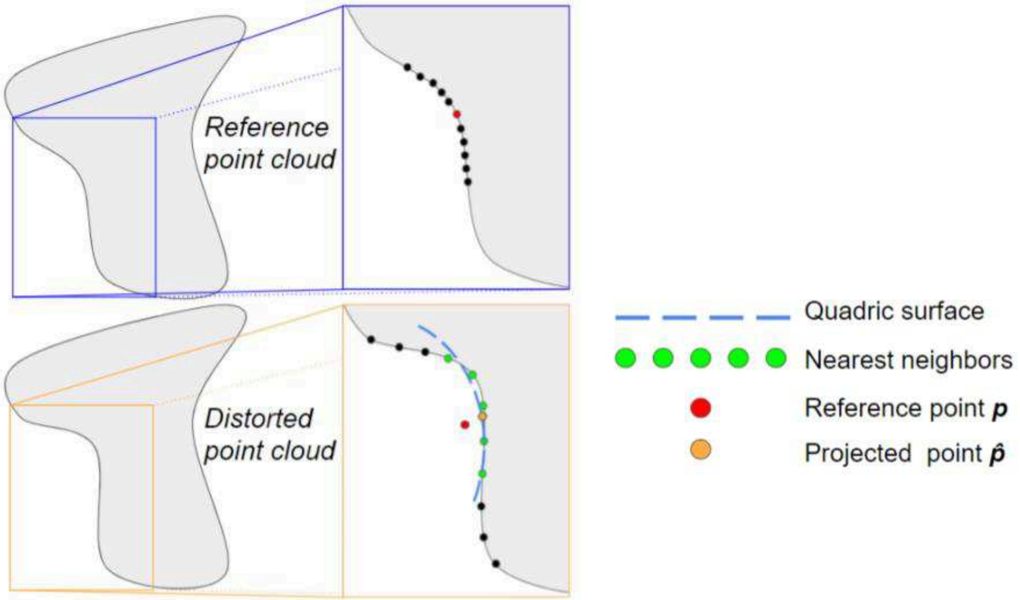
Mekuria, R. N., et al. "**Evaluation criteria for PCC (point cloud compression)**." *ISO/IEC JTC 1/SC29/WG11 Doc. N16332*, (2016).
Tian, Dong, et al. "**Geometric distortion metrics for point cloud compression**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.

Quality Assessment: Volumetric video

Point cloud metrics

- Mostly point-based computations
- Some of the state-of-the-art metrics
  - Point-to-point geometry difference
  - Color difference (Point-to-point)
  - Point-to-plane geometry difference
  - **Plane-to-plane difference**
  - **PC-MSDM**





Alexiou, Evangelos, and Touradj Ebrahimi. "**Point cloud quality assessment metric based on angular similarity**." *IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018.
Meynet, Gabriel, Julie Digne, and Guillaume Lavoué. "**PC-MSDM: A quality metric for 3D point clouds**." *Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019.

Quality Assessment: Volumetric video

Point cloud metrics

- Metrics for joint assessment of geometry and colour
- These studies are built on only one quality database and need to be studied further for generality.

- Recent studies which jointly assesses the quality
  - PCQM (extended from PC-MSDM)
    - Based on curvature
    - Including colour and lightness measures
    - Joint (geometry+colour) quality is found by weighted sum of the features
  - Viola et al. (2020)
    - Point cloud colours are taken into account by taking the histogram of the luminance channel
    - Joint quality is found by weighted sum of geometry and colour distortion values

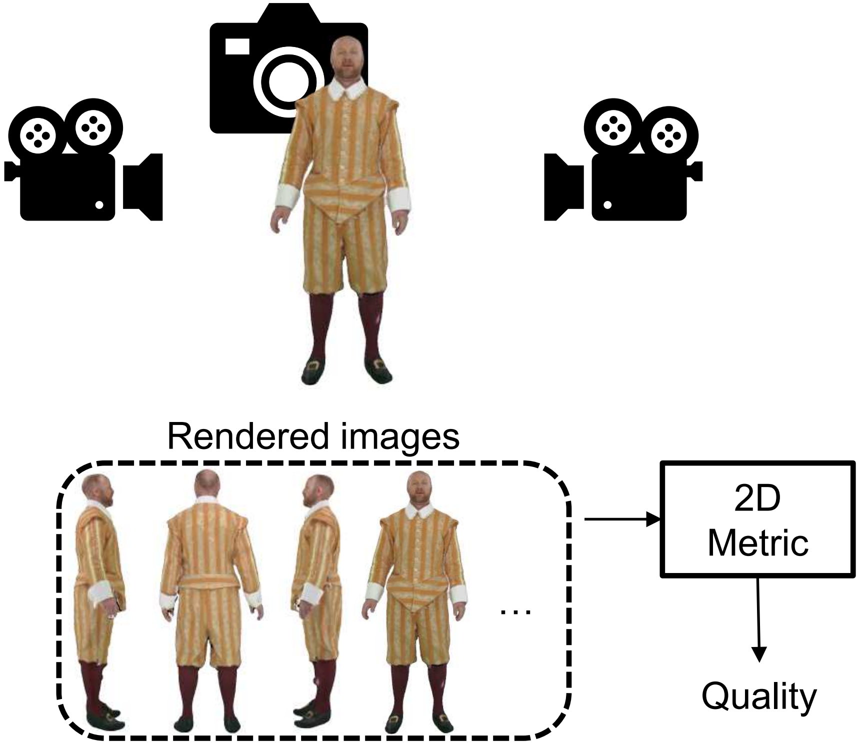
Meynet, Gabriel, et al. "**PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds**." *12th International Conference on Quality of Multimedia Experience (QoMEX 2020)*. 2020.
Viola, Irene, Shishir Subramanyam, and Pablo Cesar. "**A color-based objective quality metric for point cloud contents**." *12th International Conference on Quality of Multimedia Experience (QoMEX 2020)*. 2020.

Quality Assessment: Volumetric video

Representation-agnostic approach

- Advantages:
  - The underlying data representation is not important
  - Similar to human visual perception
  - Years of scientific research on 2D image QA

- Disadvantages:
  - Relies on rendering
  - Different rendering parameters might affect the performance



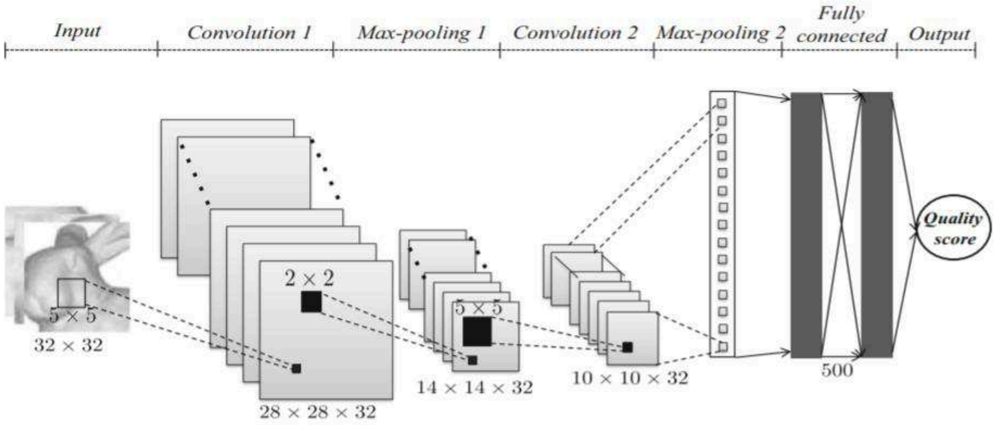
Lavoué, Guillaume, Mohamed Chaker Larabi, and Libor Váša. "**On the efficiency of image metrics for evaluating the visual quality of 3D models**." *IEEE Transactions on Visualization and Computer Graphics* 22.8 (2015): 1987-1999.

Torlig, Eric M., et al. "**A novel methodology for quality assessment of voxelized point clouds**." *Applications of Digital Image Processing XLI*. International Society for Optics and Photonics, 2018.

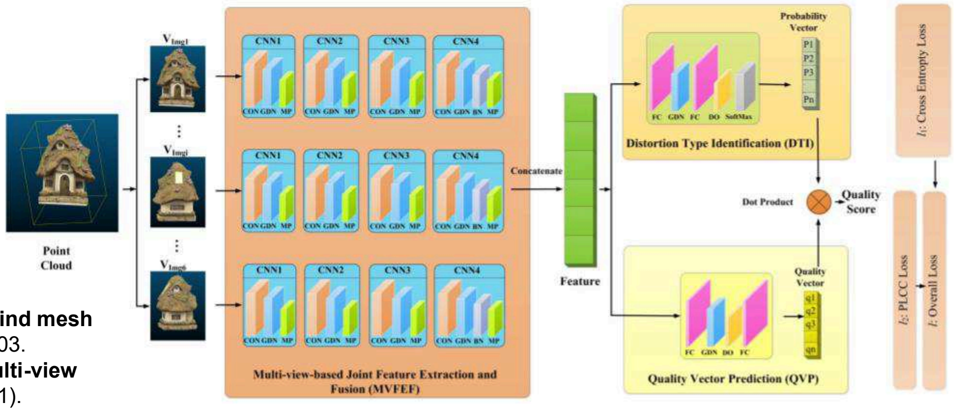
Quality Assessment: Volumetric video

Learning-based methods

- Some methods utilize the existing 2D neural network structure and estimate the point cloud or mesh quality with 2D projections



+ Lower entry barrier
+ Larger knowledge base
- Needs rendering
- Is not fully based on the data



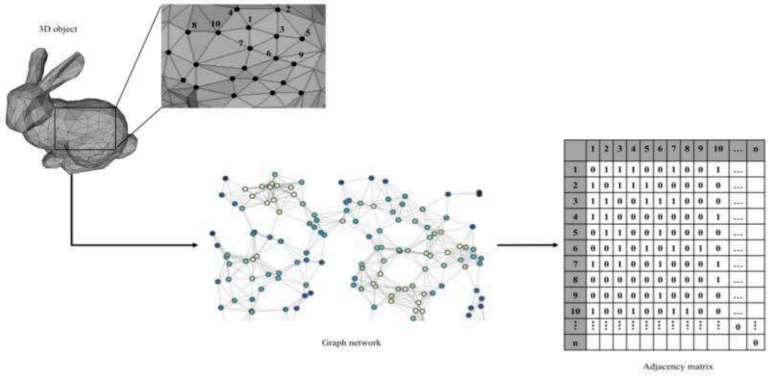
Abouelaziz, Ilyass, et al. "**3D visual saliency and convolutional neural network for blind mesh quality assessment**." *Neural Computing and Applications* 32.21 (2020): 16589-16603.
Liu, Qi, et al. "**PQA-Net: Deep No Reference Point Cloud Quality Assessment via Multi-view Projection**." *IEEE Transactions on Circuits and Systems for Video Technology* (2021).

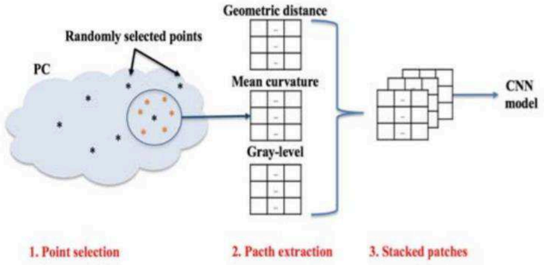
Quality Assessment: Volumetric video

Learning-based methods

- More recent methods use graph concepts and ideas from graph signal processing or graph convolutional network and other ways to order the unordered 3D representation (compared to the ordered image representations – e.g., pixels)

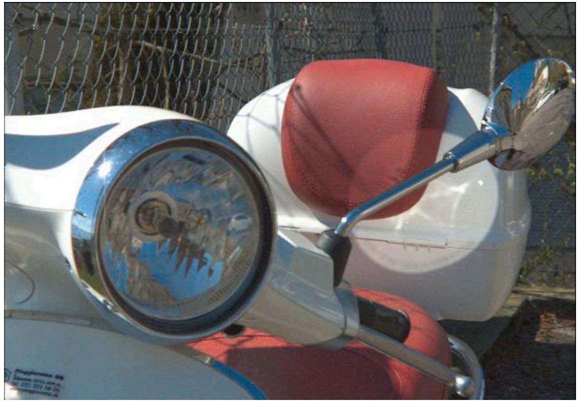
+ Using the original data representation
+ QA might be faster as there is no need to visualize the 3D models
- Some features may need to be selected by hand
- Generally, they focus only on geometry

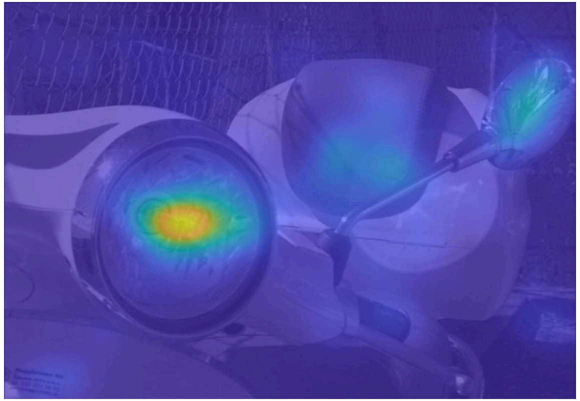




Abouelaziz, Ilyass, et al. "**Learning graph convolutional network for blind mesh visual quality assessment**." *IEEE Access* 9 (2021): 108200-108211.
Chetouani, Aladine, et al. "**Deep Learning-Based Quality Assessment Of 3d Point Clouds Without Reference**." *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2021.

Visual Attention: Overview





Visual attention

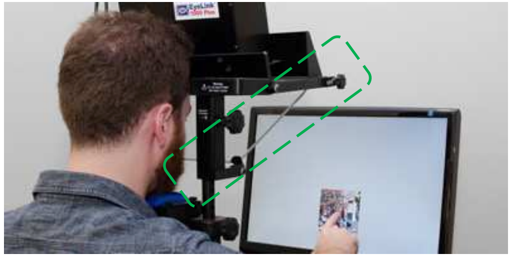
- Analysis of regions of interest within a visual stimulus, which would gather the viewer's attention
- Visual saliency
  - Detection/Estimation of important regions in an image or video
- Applications:
  - Segmentation
  - Cropping/Re-targeting
  - Compression
  - Visual quality assessment
  - Post-production

Visual Attention: Overview

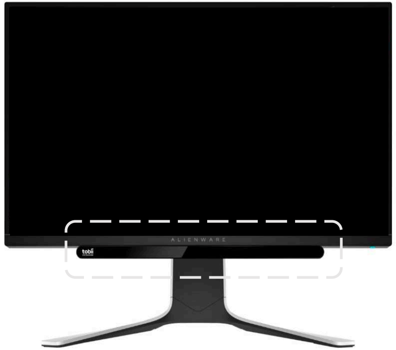
Visual attention

- How are the visual attention data collected?

- Eye tracking data is captured while showing the participants a stimulus
- There are various options for eye tracking
  - Webcam
  - Mirrored systems
  - IR cameras and sensor
    - Eyelink
    - Tobii
    - Pupil Labs
    - Others

SR Research Eyelink, https://www.sr-research.com/eyelink-1000-plus/



Tobii Eye Tracker, https://gaming.tobii.com/tobii-eye-tracker-4c/

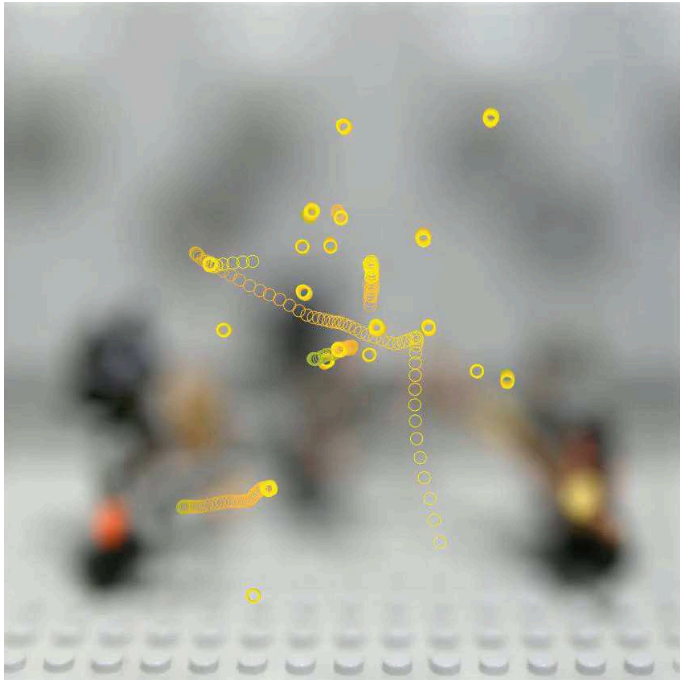
Pupil Labs, https://pupil-labs.com/products/vr-ar/

Visual Attention: Overview

Visual attention

- How are the visual attention data collected?

- Eye tracking data is captured while showing the participants a stimulus
- There are various options for eye tracking
  - Webcam
  - Mirrored systems
  - IR cameras and sensor
    - Eyelink
    - Tobii
    - Pupil Labs
    - Others

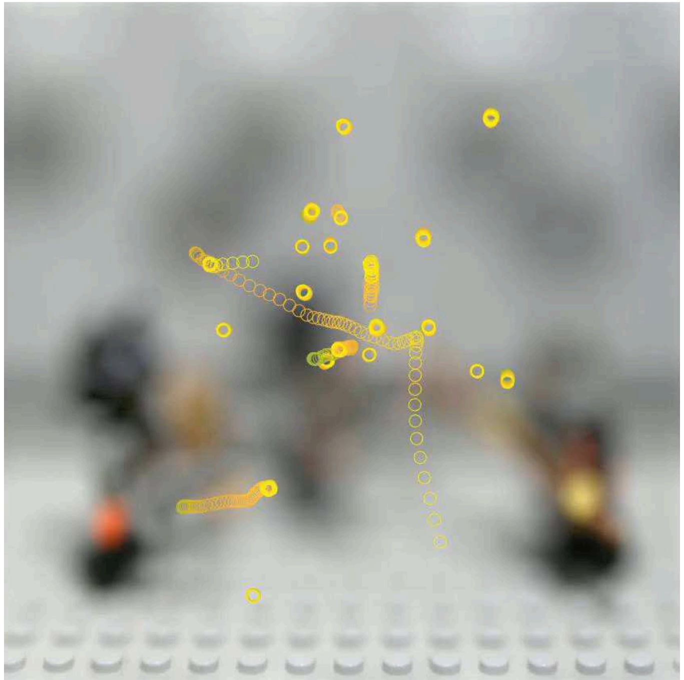


Visual Attention: Overview

Types of eye movement

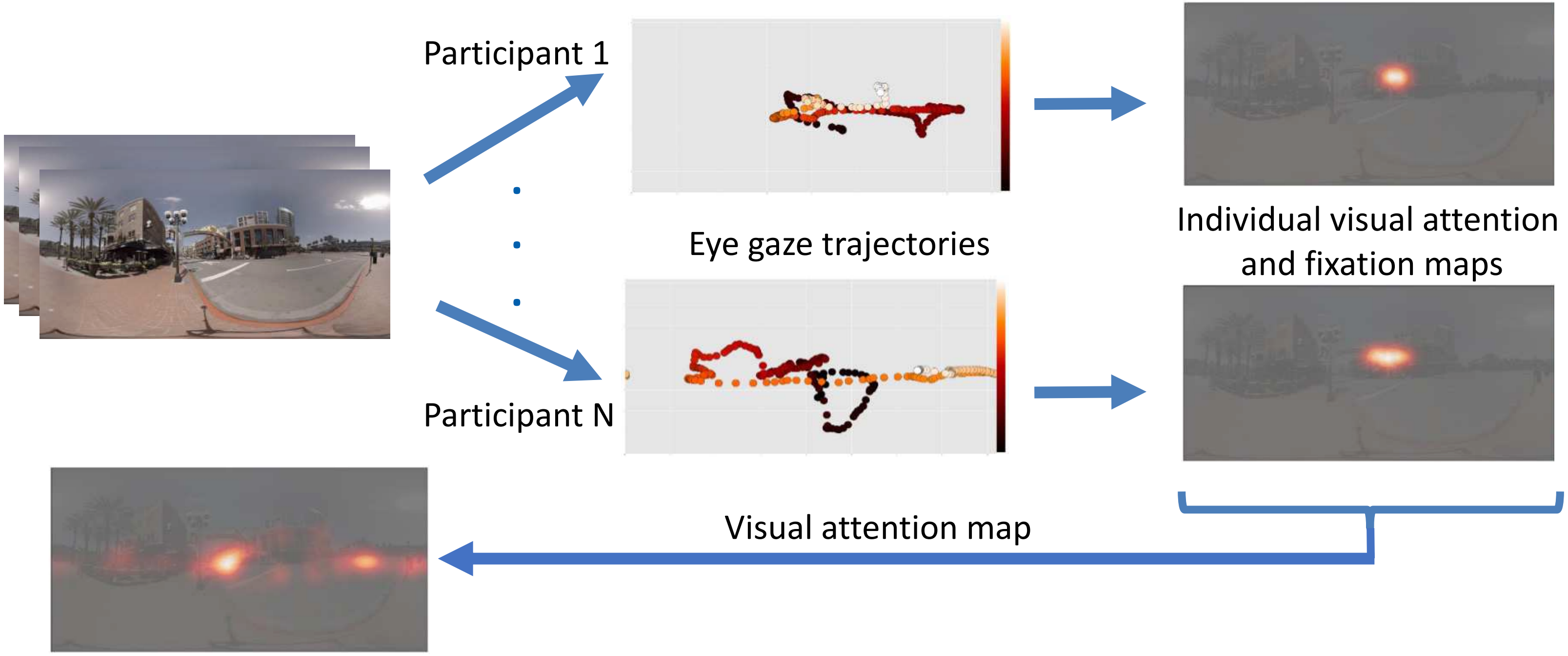
- Fixations
- Saccades

- Events (periods, blinks, smooth pursit)

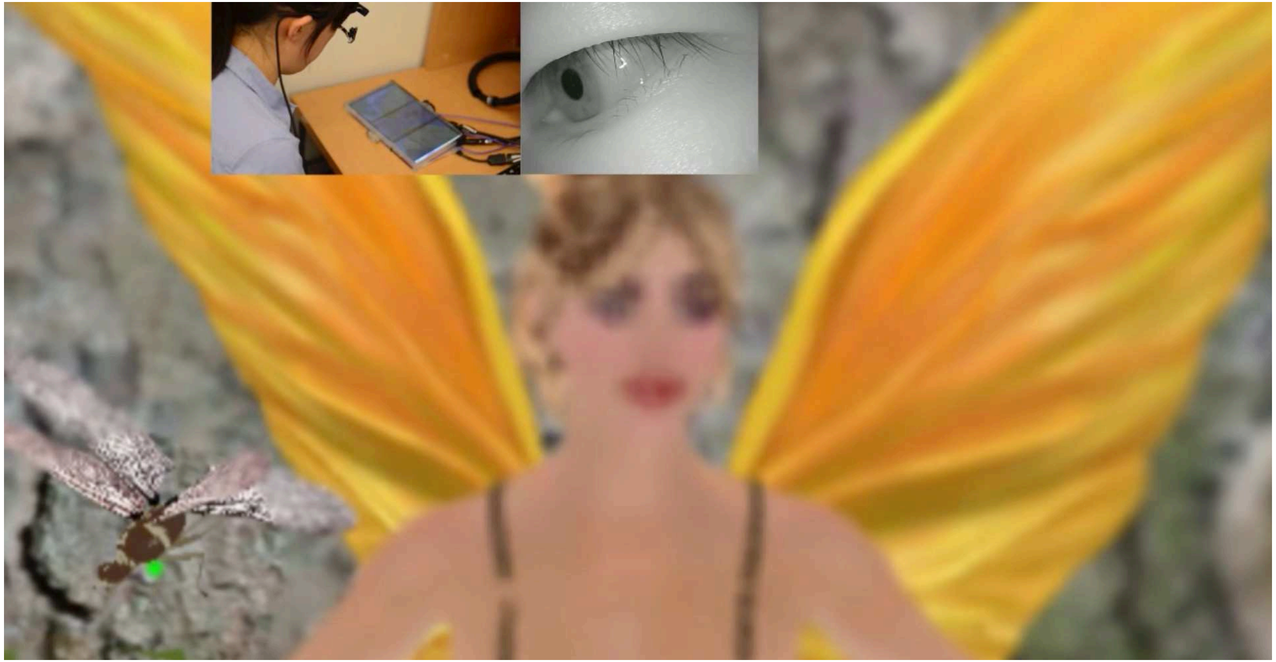


Rayner, Keith. "**The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search.**" Quarterly journal of experimental psychology 62.8 (2009): 1457-1506.

Visual Attention: Overview

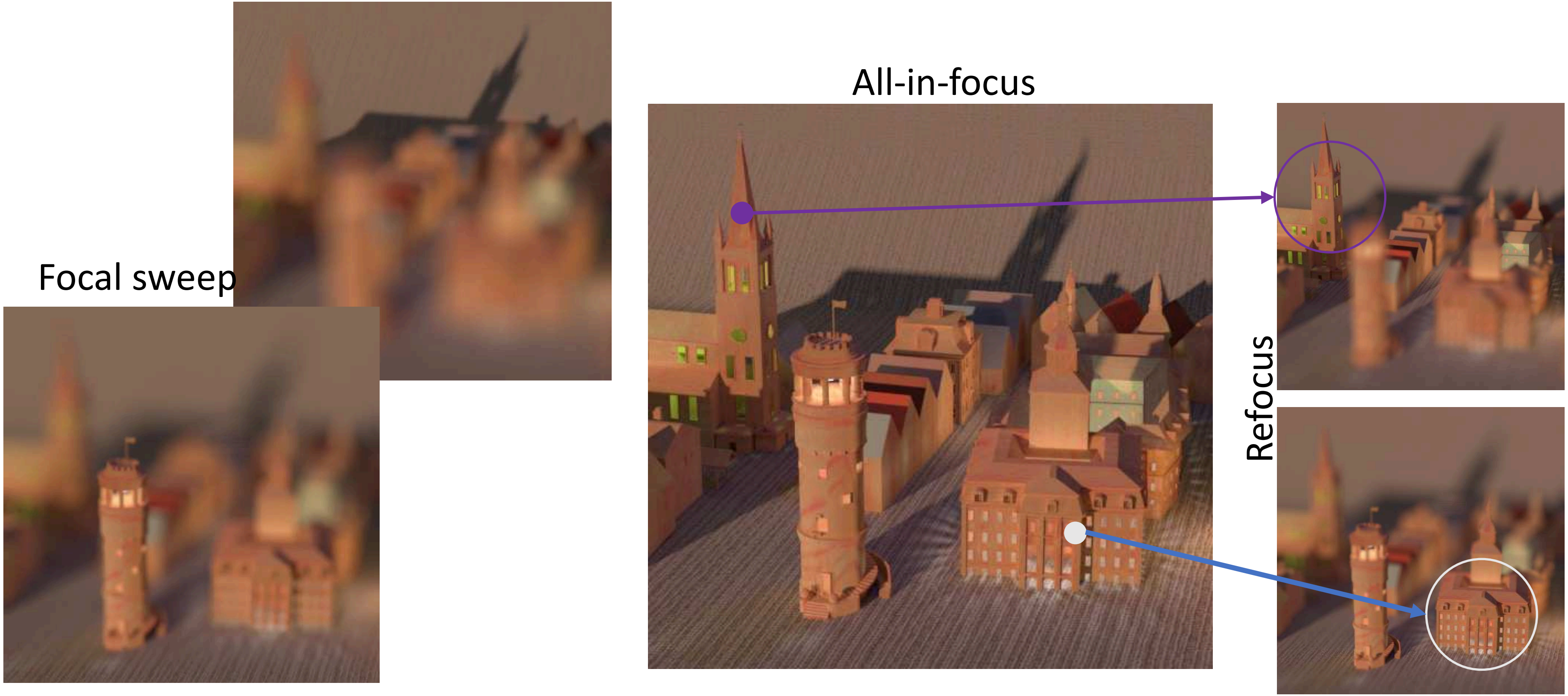


Visual Attention: Light fields



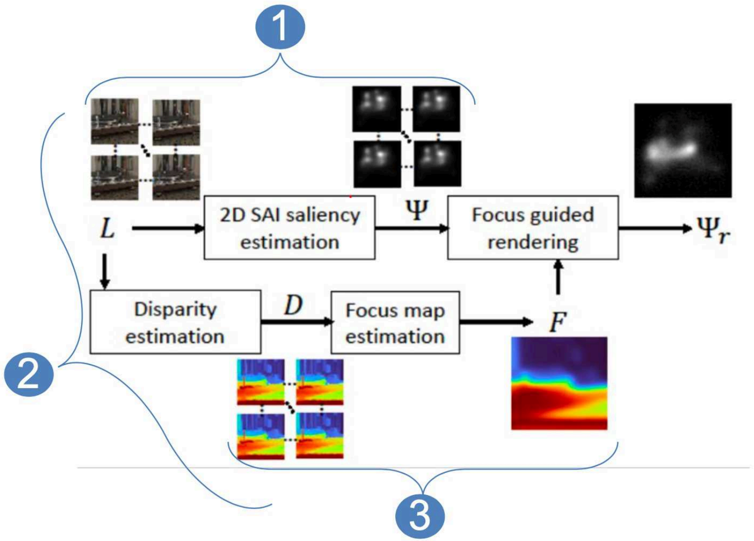
Sun, Qi, et al. "**Perceptually-guided foveation for light field displays.**" ACM Transactions on Graphics (TOG) 36.6 (2017): 1-13.

Visual Attention: Light fields



Visual Attention: Light fields

Focus Guided Saliency Estimation Pipeline



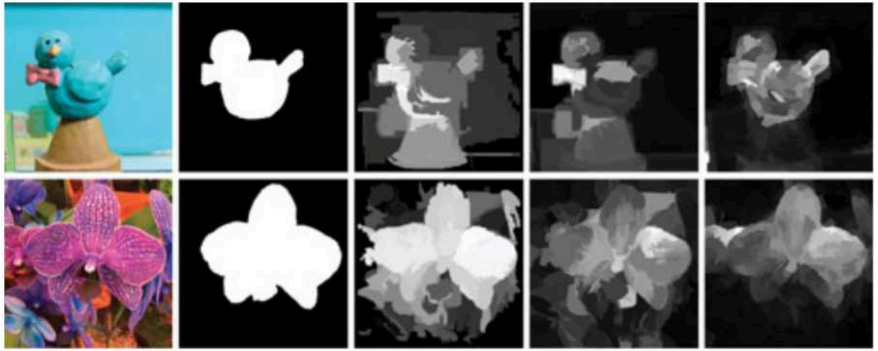
- 2D saliency estimator DeepGaze II
- Classical refocus rendering algorithm

Gill, Ailbhe, Emin Zerman, Martin Alain, Mikael Le Pendu, and Aljosa Smolic. "**Focus Guided Light Field Saliency Estimation.**" In 2021 13th International Conference on Quality of Multimedia Experience (QoMEX), pp. 213-218. IEEE, 2021.

Visual Attention: Light fields

Saliency estimation for light fields

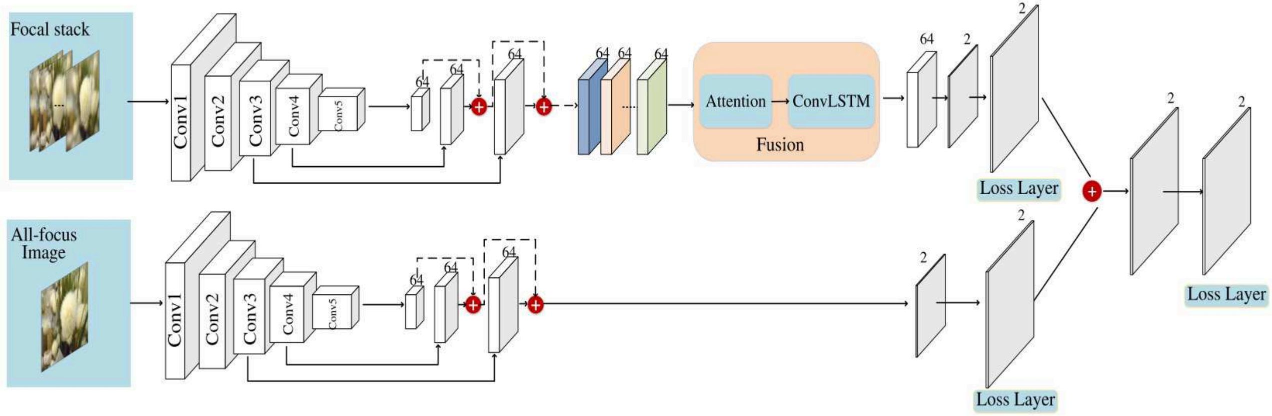
- Saliency estimation for 2D images has been studied by many.
  - Top-down – training based
  - Bottom-up – low-level features
  - Centre prior
  - Sharpness



Wang, Tiantian, et al. "**Deep learning for light field saliency detection.**" Proceedings of the IEEE International Conference on Computer Vision. 2019.
Zhang, Jun, et al. "**Light Field Saliency Detection With Deep Convolutional Networks.**" IEEE Transactions on Image Processing 29 (2020): 4421-4434.
Zhang, Miao, et al. "**LFNet: Light Field Fusion Network for Salient Object Detection.**" IEEE Transactions on Image Processing 29 (2020): 6276-6287.
Piao, Yongri, et al. "**Exploit and Replace: An Asymmetrical Two-Stream Architecture for Versatile Light Field Saliency Detection.**" AAAI. 2020.
Wang, Wenguan, et al. "**Salient object detection in the deep learning era: An in-depth survey**." *arXiv preprint* arXiv:1904.09146 (2019).
Li, Nianyi, et al. "**Saliency detection on light field**." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

Visual Attention: Light fields

Saliency estimation for light fields

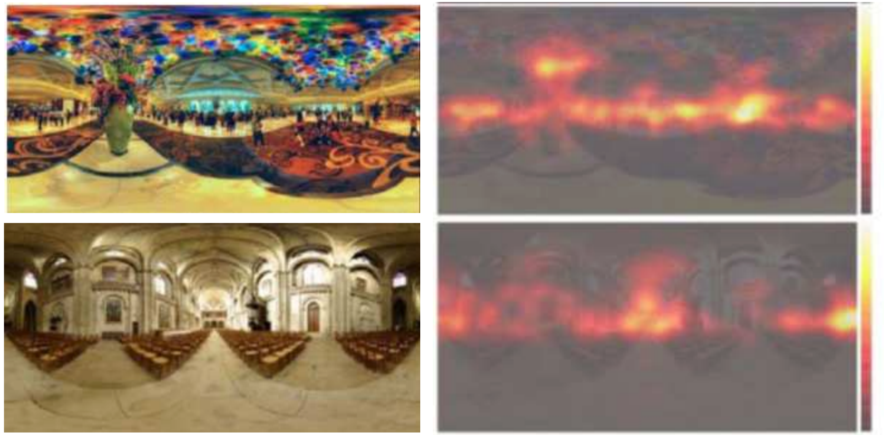


Wang, Tiantian, et al. "**Deep learning for light field saliency detection.**" Proceedings of the IEEE International Conference on Computer Vision. 2019.

Visual Attention: Omnidirectional imaging

Visual attention for omnidirectional imaging

- A number of user studies has been done
  - HMDs: HTC Vive, Oculus Rift, etc.
  - Head and/or eye trajectories are considered
- User exploration studies show that
  - Viewers generally spend their time in visually comfortable regions
    - Equatorial region

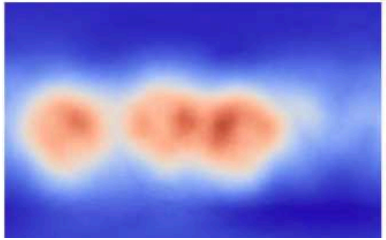


Singla, Ashutosh, et al. "**AhG8: Measurement of User Exploration Behavior for Omnidirectional (360°) Videos with a Head Mounted Display**." *Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 118th Meeting*: Macao, CN. 2017.
Rai, Yashas, Jesús Gutiérrez, and Patrick Le Callet. "**A dataset of head and eye movements for 360 degree images.**" Proceedings of the 8th ACM on Multimedia Systems Conference. 2017.
Xu, Mai, et al. "**State-of-the-art in 360 video/image processing: Perception, assessment and compression**." IEEE Journal of Selected Topics in Signal Processing 14.1 (2020): 5-26.
Ozcinar, Cagri, and Aljosa Smolic. "**Visual attention in omnidirectional video for virtual reality applications**." 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2018.
De Abreu, Ana, Cagri Ozcinar, and Aljosa Smolic. "**Look around you: Saliency maps for omnidirectional images in VR applications**." *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.
Corbillon, Xavier, Francesca De Simone, and Gwendal Simon. "**360-degree video head movement dataset**." *Proceedings of the 8th ACM on Multimedia Systems Conference*. 2017.

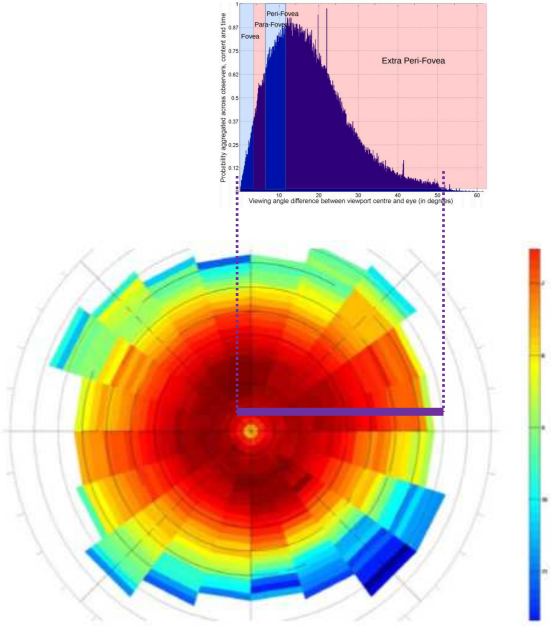
Visual Attention: Omnidirectional imaging

Visual attention for omnidirectional imaging

- Another study compared different weighting methods
  - Oculus Rift DK2 + eye tracking camera
  - Head + eye tracking



Centre of Viewport w/ Gaussian filter



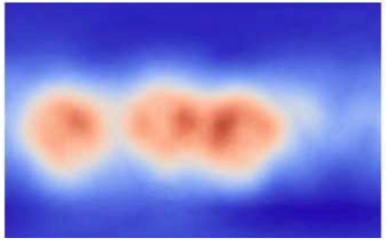
Rai, Yashas, Patrick Le Callet, and Philippe Guillotel. "**Which saliency weighting for omni directional image quality assessment?**." *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.

John, Brendan, et al. "**A Benchmark of Four Methods for Generating 360∘ Saliency Maps from Eye Tracking Data.**" International Journal of Semantic Computing 13.03 (2019): 329-341.

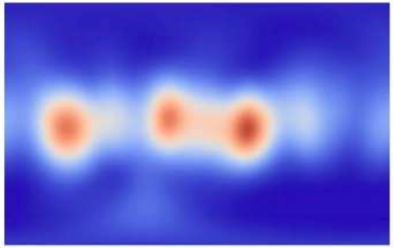
Visual Attention: Omnidirectional imaging

Visual attention for omnidirectional imaging

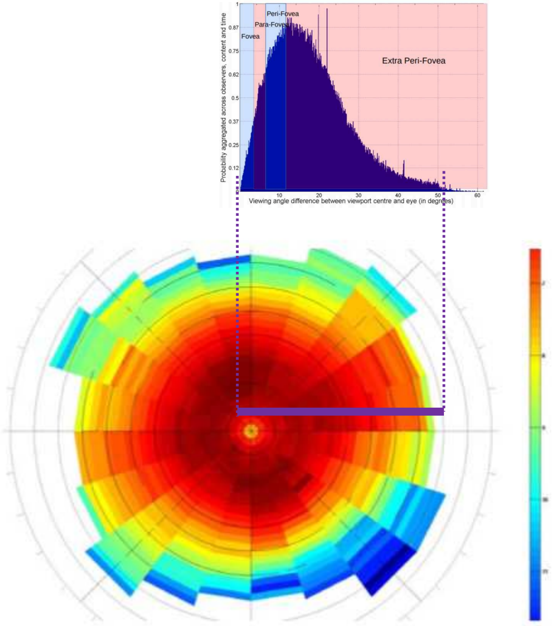
- Another study compared different weighting methods
  - Oculus Rift DK2 + eye tracking camera
  - Head + eye tracking



Centre of Viewport w/ Gaussian filter



Centre of Viewport w/ the new filter



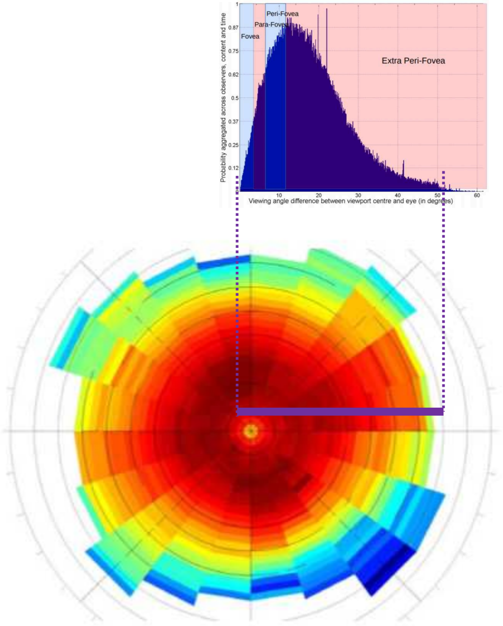
Rai, Yashas, Patrick Le Callet, and Philippe Guillotel. "**Which saliency weighting for omni directional image quality assessment?**." *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.

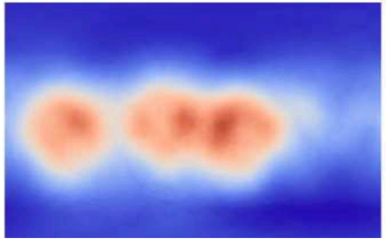
John, Brendan, et al. "**A Benchmark of Four Methods for Generating 360∘ Saliency Maps from Eye Tracking Data.**" International Journal of Semantic Computing 13.03 (2019): 329-341.

Visual Attention: Omnidirectional imaging

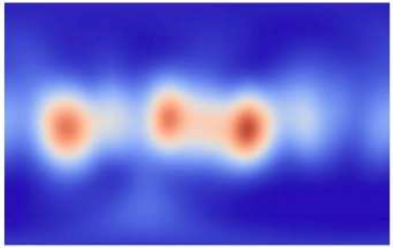
Visual attention for omnidirectional imaging

- Another study compared different weighting methods
  - Oculus Rift DK2 + eye tracking camera
  - Head + eye tracking

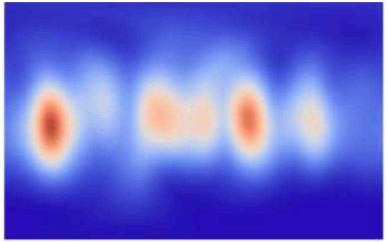




Centre of Viewport w/ Gaussian filter



Centre of Viewport w/ the new filter



Real Visual Attention (head + eye tracking)

Rai, Yashas, Patrick Le Callet, and Philippe Guillotel. "**Which saliency weighting for omni directional image quality assessment?**." *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017.

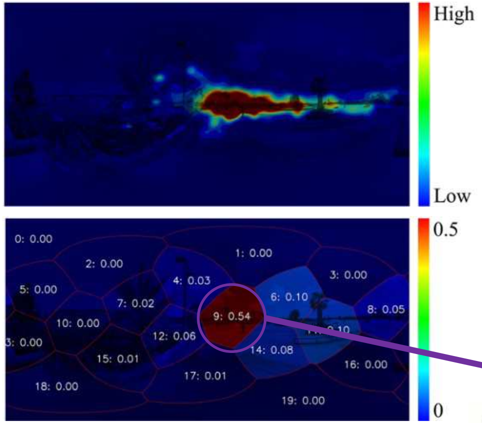
John, Brendan, et al. "**A Benchmark of Four Methods for Generating 360∘ Saliency Maps from Eye Tracking Data.**" International Journal of Semantic Computing 13.03 (2019): 329-341.

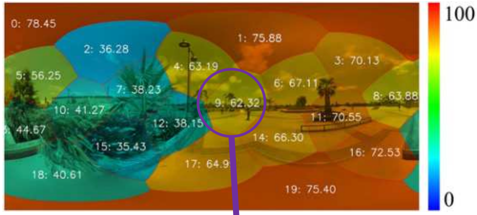
Visual Attention: Omnidirectional imaging

Visual attention for omnidirectional imaging

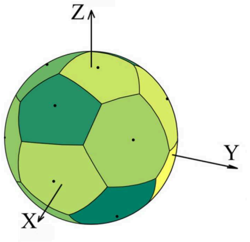
- Use in quality assessment
- Visual attention can be used to
  - Compute weighted sum of distortion/ quality measure (e.g., MSE, VMAF, etc.)
  - Find the Voronoi cells with higher visual attention to get a weighted sum

$$MSE_{VA} = \frac{\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}\left(I(i,j)-\hat{I}(i,j)\right)^2 h_{i,j}}{\sum_{i=0}^{H-1}\sum_{j=0}^{W-1} h_{i,j}}$$





$$T_i = \frac{\sum_{k=0}^{M-1} v_{i,k}\Gamma_{i,k}}{\sum_{k=0}^{M-1} v_{i,k}}$$



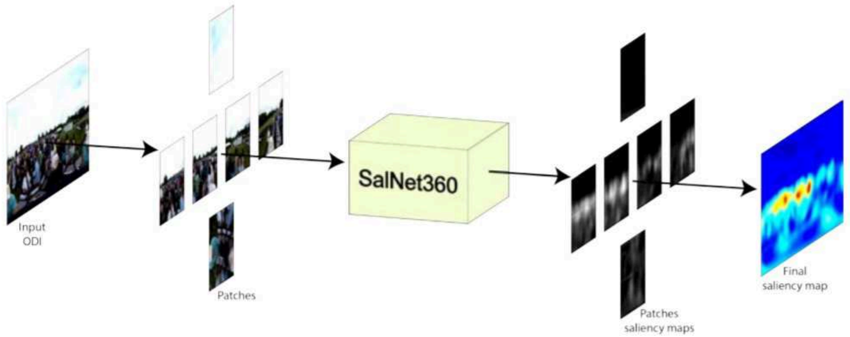
Quality

Upenik, Evgeniy, and Touradj Ebrahimi. "**Saliency Driven Perceptual Quality Metric for Omnidirectional Visual Content**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.
Croci, Simone, et al. "**Visual attention-aware quality estimation framework for omnidirectional video using spherical Voronoi diagram**." Quality and User Experience. 5.4 (2020).
Ozcinar, Cagri, Julián Cabrera, and Aljosa Smolic. "**Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality.**" IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9.1 (2019): 217-230.

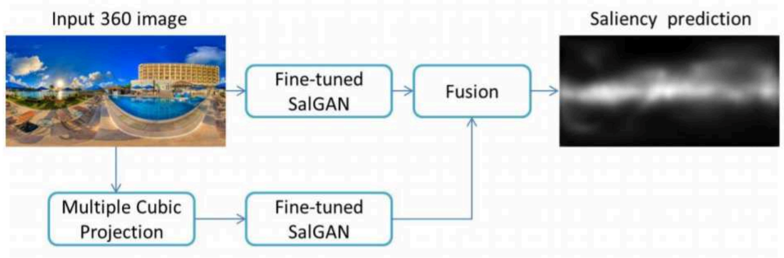
Visual Attention: Omnidirectional imaging

Saliency estimation for omnidirectional imaging

- Dividing into patches



Monroy, Rafael, et al. "**Salnet360: Saliency maps for omni-directional images with CNN.**" *Signal Processing: Image Communication* 69 (2018): 26-34.



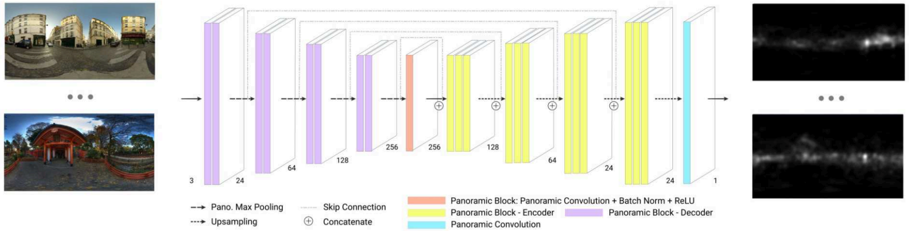
Chao, Fang-Yi, et al. "**SALGAN360: Visual saliency prediction on 360 degree images with generative adversarial networks.**" *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2018.

Salient360!—Visual Attention Modeling for 360 Content. Retrieved from https://salient360.ls2n.fr/grand-challenges/.

Visual Attention: Omnidirectional imaging

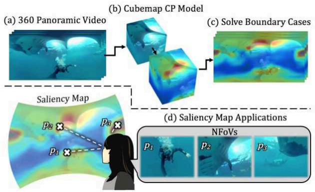
Saliency estimation for omnidirectional imaging

- Panoramic convolutions

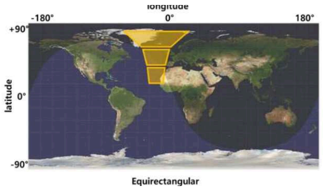


Martin, et al. "**Panoramic convolutions for 360 single-image saliency prediction**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop. 2020.

Recent deep-learning advances



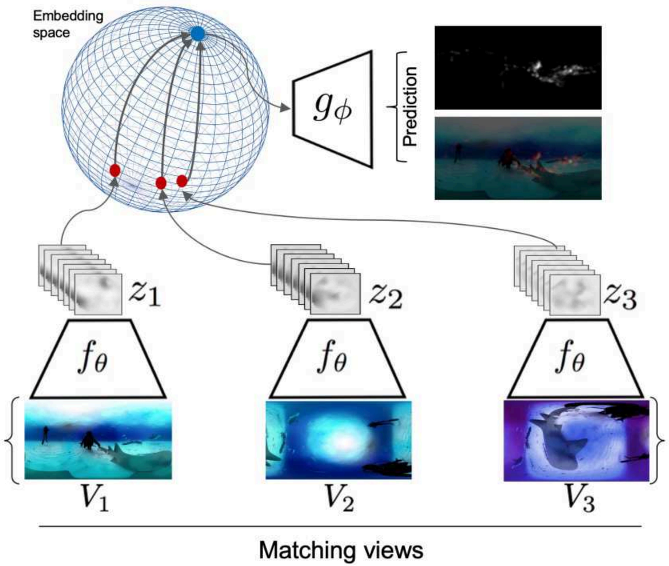
Cheng, Hsien-Tzu, et al. "**Cube padding for weakly-supervised saliency prediction in 360 videos.**" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.



Lee, Yeonkun, et al. "**SpherePHD: Applying CNNs on a spherical polyhedron representation of 360deg images.**" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

Eder, Marc, et al. "**Tangent Images for Mitigating Spherical Distortion.**" Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Visual Attention: Omnidirectional imaging



A deep representation is learnt by maximizing the mutual information between views of the same scene in the embedding space, while discarding views of different scenes.

Djilali, Yasser Abdelaziz Dahou, Tarun Krishna, Kevin McGuinness, and Noel E. O'Connor. "**Rethinking 360 Image Visual Attention Modelling With Unsupervised Learning.**" In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 15414-15424. 2021.

Visual Attention: Volumetric video

Visual attention for point clouds

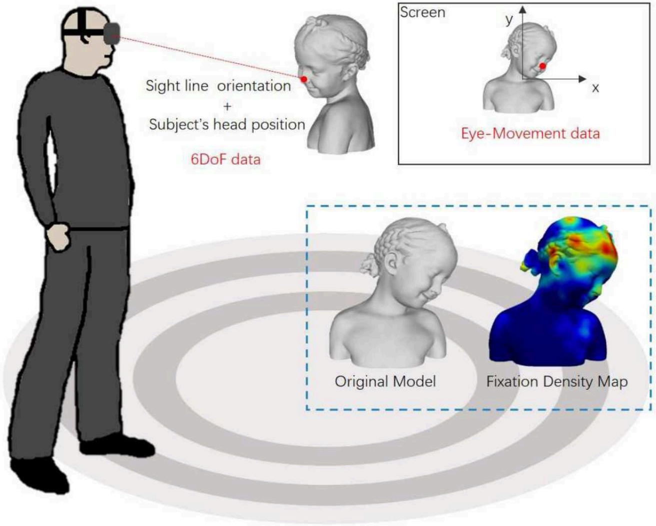
- A custom environment in Unity 3D
- HTC Vive Pro headset + Pupil Labs hardware for eye tracking
- Static 3D models

Alexiou, Evangelos, Peisen Xu, and Touradj Ebrahimi. "**Towards Modelling of Visual Saliency in Point Clouds for Immersive Applications**." *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

Visual Attention: Volumetric video

Visual attention for polygon meshes

- HTC Vive headset + aGlass
- 3D meshes consist of four different types: humans, animals, familiar objects, and mechanical parts.

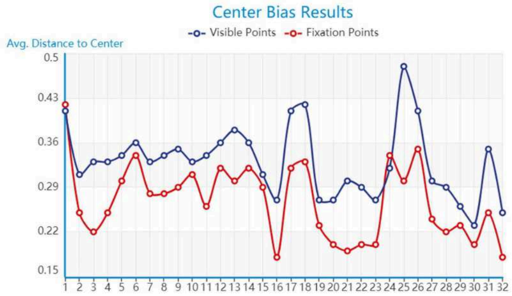


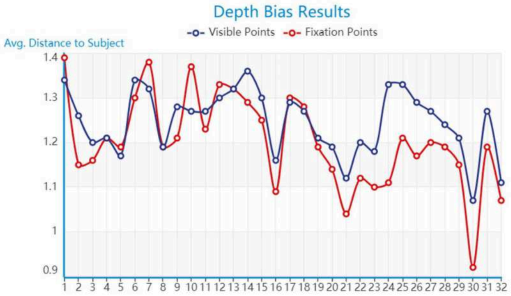
Ding, Xiaoying, and Zhenzhong Chen. "**Towards Mesh Saliency Detection in 6 Degrees of Freedom.**" arXiv preprint arXiv:2005.13127 (2020).

Visual Attention: Volumetric video

Visual attention for polygon meshes

- Center bias





Ding, Xiaoying, and Zhenzhong Chen. "**Towards Mesh Saliency Detection in 6 Degrees of Freedom.**" arXiv preprint arXiv:2005.13127 (2020).

Summary: Perception & Quality

Basic principles of visual perception

Quality assessment for

- Light fields
- Omnidirectional imaging
- Volumetric videos

Visual attention for

- Light fields
- Omnidirectional imaging
- Volumetric videos

Many Thanks!

References:

- Check out the tutorial website:
- https://v-sense.scss.tcd.ie/lectures/tutorial-on-immersive-imaging-technologies/