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

Immersive Imaging Technologies: from Capture to Display | Tutorial | VCIP 2021, Munich, Germany
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

15:00
Part III: Content Delivery ~20 minutes
Coffee Break

15:30
Continuing Part III: Content Delivery ~20 minutes
Part IV: Rendering and Display Technologies ~20 minutes
5 mins Q&A
Part V: Perception & Quality Evaluation ~40 minutes

17:00
5-10 mins Q&A
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

- The real world
- Sensory experience

- Sensory experience
- Non-sensory experience

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.

Tele-Immersion (Mulligan, 2001)
- “interaction (presence) of people from different places in virtually the same environment.”

Immersion (Takatalo, 2008)
- “concentration to the VE instead of the real world, loss of time”


Imaging Technologies & Immersion

Timeline of imaging and immersion

- Lumière brothers’ train
- Silent films → Audio
- Color films
- High dynamic range
- Immersive imaging technologies
Immersive 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
Immerison & Degrees of freedom (DoF)

- **3DoF**
  - Rotation around 3 axes
  - **No** spectator movement
  - 3 degrees-of-freedom in total

- **3DoF+**
  - Rotation around 3 axes
  - **Limited** 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 handheld display

- **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

Omnidirectional Video

Traditional Display

Light Fields

Augmented Reality

Virtual Reality

Attribution: People illustrations were created by studiogstock - www.freepik.com
Different Imaging Modalities
Theoretical background

The plenoptic function

$P(\theta, \phi, \lambda, t, x, y, z)$

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

Different Imaging Modalities

Light as a field

Light rays parameterized by their intersection with two parallel planes
Different Imaging Modalities

Light fields

Light as a field

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

the camera plane

the sensor / image plane

Light field view $L_{s,t}$
Different Imaging Modalities

Light fields

Light as a field

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

the camera plane

the sensor / image plane
Different Imaging Modalities
Omnidirectional image and video

Omnidirectional (360-degree) image and video

\[ L(\theta, \phi, t) \]
Different Imaging Modalities
Omnidirectional image and video

Omnidirectional (360-degree) image and 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


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)
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

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

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

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

Applications & Creative Experiments

Communication

- Immersive communication
  - “Holoportation”
    - Real-time acquisition and display of volumetric videos
- 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/](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

Next part:

• How to capture immersive images and videos in practice?

Applications for these technologies
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)
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

\[
K = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad P = K \times [R | t] \quad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \times \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}
\]

Based on a single 2D camera

2D imaging: camera lens

\[ \frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} \]

object plane \quad imaging lens \quad image sensor

Attribution: Weston Aenchbacher
Based on a single 2D camera

Light field capture with robotized camera

the camera plane

the sensor / image plane
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

Attribution: Adrian Rosebrock
Based on a single 2D camera

Smartphone capture

- simultaneous localization and mapping (SLAM)

Advanced camera design

Light field capture with a kaleidoscope lens

Advanced camera design

Light field capture with a kaleidoscope lens

Single-Camera Systems

Light field capture with a kaleidoscope lens

Sensor image

Advanced camera design

Light field capture with a kaleidoscope lens

Sensor image

Advanced camera design

Light field capture with lenslet plenoptic camera

Lytro Illum
Plenoptic 1.0

Raytrix
Plenoptic 2.0


Advanced camera design

Lenslet plenoptic camera

object plane \hspace{3cm} \text{imaging lens} \hspace{3cm} \text{image sensor}

Camera

Attribution: Weston Aenchbacher
Advanced camera design

Lenslet plenoptic camera

object plane

imaging lens

lenslet array

image sensor

plenoptic camera

\[ f_{\text{lenslet}} \]

Attribution: Weston Aenchbacher

Immersive Imaging Technologies: from Capture to Display | Tutorial | VCIP 2021, Munich, Germany
Advanced camera design

Lenslet plenoptic camera

![Diagram of lenslet plenoptic camera](image)

The diagram illustrates a lenslet plenoptic camera setup. The camera captures light fields in the form of $L(s, t, u, v)$, where $s$ and $t$ represent the spatial coordinates, and $u$ and $v$ represent the angular coordinates. The camera design is crucial for capturing immersive images.
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 \( L(s, t, u, v) \)

Attribution: Weston Aenchbacher

Advanced camera design

Lenslet plenoptic camera: plenoptic 1.0

Advanced camera design

Lenslet plenoptic camera: plenoptic 1.0

Raw image

Light field views
Advanced camera design

Lenslet plenoptic camera: plenoptic 2.0


Attribution: Weston Aenchbacher
Advanced camera design

Lenslet plenoptic camera: plenoptic 2.0

Raw video

Light field views

Omnidirectional imaging

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


Advanced camera design

Omnidirectional imaging

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

\[
P_s = \frac{P}{\|P\|}
\]

\[
\begin{bmatrix}
  u \\
  v \\
  1
\end{bmatrix} = K \times \left( P_s + \begin{bmatrix}
  0 \\
  0 \\
  \xi
\end{bmatrix} \right)
\]


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

Multi-Camera Systems

Towards spherical light field capture

- Cameras positioned on a portion of a sphere


Planar light field capture

- Camera arrays

Stanford, 2004


Technicolor, 2017

Saarland university, 2018
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


Multi-Camera Systems

3D models & Volumetric video

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

Multi-Camera Systems

Combining advanced cameras

- Plenoptic multiview
- Plenoptic panorama


Multi-Camera Systems

Combining advanced cameras

- Grid of omnidirectional cameras
- Allows 6DoF


Attribution: project.inria.fr/ftv360/
Data Formats and Representations

Light fields
- Collection of images

Omnidirectional imaging
- Projections from sphere to plane

3D models & Volumetric video
- Textured meshes
- Point clouds
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)

Data Formats and Representations

Light fields

- Multi-Plane Images (MPIs)

Data Formats and Representations

Light fields

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

Data Formats and Representations

Omnidirectional imaging

- Projections from sphere to plane
- Different projection methods
  - Equirectangular projection
  - Cubemap projection
  - Octahedron projection
  - Icosahedron 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

Learning-based methods

Neural volumes

Learning-based methods

Pixel-Aligned Implicit Function (PiFU)

Learning-based methods

Monocular Real-Time Volumetric Capture

- PiFU-based system
- Texture atlas

Learning-based methods

Neural Radiance Fields (NeRF)

Learning-based methods

Neural Radiance Fields (NeRF)

Learning-based methods

Neural Radiance Fields (NeRF)

\[(x, y, z, \theta, \phi) \rightarrow F_\theta \rightarrow (RGB\sigma)\]

(a) Coordinate-based MLP

Summary: Acquisition & Data Format

Key concepts of immersive imaging

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

Next part:
- How to encode and transmit all this data?

Data Formats

- Light field images
- Omnidirectional images
- Textured meshes and point clouds
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 ➔ Encoder ➔ Multi-rate encoding
Part III: Content Delivery

Video streaming

Captured video ➔ Encoder ➔ Packager ➔ Origin Server

Multi-rate encoding
Part III: Content Delivery

Video streaming

1. Captured video
2. Multi-rate encoding
3. Encoder
4. Packager
5. Origin Server
6. Network
7. HTTP Request
8. HTTP Get
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

Acquisition → Representation → Encoding → Decoding → Rendering → Display

Storage/Transmission
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


**Light field coding**

**Based on a sparse subset**

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

---


Light field coding

Based on a sparse subset

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

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

Light field coding

Based on standard video coding

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


Light field coding
Based on standard video coding


Light field coding
Based on standard video coding

Raster
Spiral

Serpentine
Zig-zag

Perpendicular
Hilbert

U-shape
Lozenge


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

Light field coding
Based on multi-view video coding

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

- Warping and Sparse Prediction (WaSP)

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


Light field coding

Learning-based coding

- Utilizing learning based multiplane images (MPI)

Light field streaming

• Only a few views are transmitted for free viewpoint rendering

• Only subset of SAIs are transmitted
Light field streaming

Cloud – Light field storage

Light field streaming

User side – Light field visualisation

User side – Light field visualisation

Cloud – Light field storage

Light field streaming

Scenario A: Render on server side and transmit rendered image

User side – Light field visualisation

Cloud – Light field storage

Light field streaming

Scenario B: Transmit full light field and render on user side

User side – Light field visualisation

Cloud – Light field storage

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

Light field streaming

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

Part III: Content Delivery
Delivery of 360-degree video

Acquisition → Tiling → Encoding → Storage/Transmission

Display ← Rendering ← Decoding
360-degree video coding

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

360-degree video streaming

Source video → Tiling → Encoding and DASHify → HTTP server

Internet → Selection

Figure: Viewport-aware adaptive streaming using tiles method

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

Resolution 1

Resolution 2

Resolution 3

(tile, resolution, rate) combination
360-degree video streaming

Server-side optimization

Spherical  Planar Projection  Tiling  Encoding Ladder

Figure: Overview of the different formats and representations

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

360-degree video streaming

Server-side optimization

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

360-degree video streaming

Server-side optimization

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

360-degree video streaming
Server-side optimization

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

360-degree video streaming

Dynamic tiling

![Image of 360-degree video streaming and dynamic tiling]
360-degree video streaming

Dynamic tiling
360-degree video streaming

Dynamic tiling

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

360-degree video streaming

Visual attention

One user

Several users
360-degree video streaming

Bitrate-allocation

Target bitrate of DASH representation \( \rightarrow R \) Bits/s

\[
R_1 = \varphi_1 R \\
R_2 = \varphi_2 R \\
R_3 = \varphi_3 R \\
R_4 = \varphi_4 R
\]

Computation of \( \varphi_i = \frac{\sum_{i=1}^{M_t} \sum_{j=1}^{N_t} v_{d(i,j)}^k}{M_tN_t} \)

\[
\sum_i R_i^j \leq R
\]
360-degree video streaming

Visual-attention based metric

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

\[
\text{MSE} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} e(i,j)^2}{M \times N}
\]

\[
\text{VA-MSE} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} e(i,j)^2 \cdot v_a(i,j)}{s_{vp}}
\]

Visual attention map
360-degree video streaming

Spherical weighted metric

Accounts for the geometrical distortion of the ERP

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

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

360-degree video streaming

Time-series model with Passive Aggressive Regression

360-degree video streaming

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

Part III: Content Delivery

Light Fields

360-degree video

Volumetric video
Delivery of volumetric video

Acquisition → Reconstruction → Encoding → Storage/Transmission → Display → Rendering → Decoding

Point Cloud/Mesh
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

Volumetric video coding

Video-based PCC

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

Video-based PCC

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
Volumetric video coding
Based on deep learning

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

VoxelDNN network architecture

Volumetric video streaming

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

Volumetric video streaming

Bitrate-allocation

1) Distance to user

2) Size of the object

3) Visual area within the viewport

Volumetric video streaming
Based on 3D tiles

Volumetric video streaming

Volumetric video streaming

Video conferencing scenario

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

Volumetric video streaming

Figure: Architecture of the developed video conferencing system

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.

Summary

Brief overview of content delivery systems for traditional digital imaging

Content delivery mechanisms, i.e.:
- Compression
- Streaming

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

Various immersive imaging modalities:
- Light fields
- Omnidirectional imaging
- Volumetric videos
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


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

Viewport rendering from light field using MPIs/MSIs

Rendering algorithms

Deep 3D Capture

Rendering algorithms

PiFU

Rendering algorithms

NeRF

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

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

- Tablets
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

- 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 ($\geq 90$Hz)
• Eye-tracking
  • Foveated rendering
• Lenses map the up-close display to a wide field of view
Near-eye light field displays

- Similar to integral imaging

Head-Mounted Displays for Virtual Reality

Near-eye light field displays

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

Head-Mounted Displays for Augmented Reality

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

Attribution: Youtube: Microsoft HoloLens: Partner Spotlight with Case Western Reserve University
Head-Mounted Displays for Augmented Reality

More powerful
Heavier and bulkier

Increased FoV, eye tracking
Expensive

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

Easy to wear.
Low spatial resolution and FoV.
Head-Mounted Displays Comparison

Human Stereoscopic Vision

Human Peripheral Vision

Oculus Rift (~125°)

Playstation VR (~100°)

HoloLens 1 (~34°)

Magic Leap 1 (~54°)

HoloLens 2 (~54°)
Head-Mounted Displays Comparison

Playstation VR (~100°)
Head-Mounted Displays for VR/AR

Pass-through AR
“passes” outside view through

VR Mode
Does NOT pass the outside view through

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

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
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

Vertical FoV = ~125°

Horizontal line

Resting sight line

Eye rotation limits

Vertical
Visual Perception

Visual angle

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

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”

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”

**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

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)

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

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

\[
PSNR_{YUVmean} = \frac{1}{(I-2)(J-2)} \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} PSNR_{YUV}(i,j)
\]

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.

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

Quality Assessment per Modality

- **Light Fields**
- **360-degree video**
- **Volumetric video**
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)


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
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
Qualitv 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

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


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

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 – $3DWP_{M_1}, 3DWP_{M_2}$
  - Angles, etc.

- Joint models also exist

\[ CM = \alpha Q_G + (1 - \alpha) Q_T, \]
\[ CM_1 = \alpha_1 MSDM2 + (1 - \alpha_1) MS - SSIM \]
\[ CM_2 = \alpha_2 SDCD + (1 - \alpha_2) MS - SSIM \]

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

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

where

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

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


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 assess 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

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


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

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

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


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

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 pursuit)

Visual Attention: Overview

Participant 1

Eye gaze trajectories

Individual visual attention and fixation maps

Participant N

Visual attention map
Visual Attention: Light fields

Visual Attention: Light fields

Focal sweep

All-in-focus

Refocus
Visual Attention: Light fields

Focus Guided Saliency Estimation Pipeline

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

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

Visual Attention: Light fields

Saliency estimation for light fields

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


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


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

---


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)


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

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

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

Quality


Visual Attention: Omnidirectional imaging

Saliency estimation for omnidirectional imaging

- Dividing into patches


Visual Attention: Omnidirectional imaging

Saliency estimation for omnidirectional imaging

- Panoramic convolutions

Recent deep-learning advances


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.

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

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.

Visual Attention: Volumetric video

Visual attention for polygon meshes

- Center bias

Summary: Perception & Quality

Basic principles of visual perception

Quality assessment for
- Light fields
- Omnidirectional imaging
- Volumetric videos

Many Thanks!

Visual attention for
- Light fields
- Omnidirectional imaging
- Volumetric videos

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