360 VIDEO

Jill Boyce

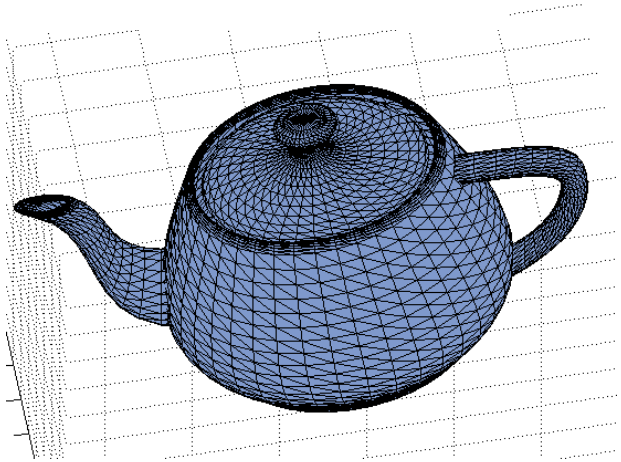
Intel Fellow, VPG Chief Media Architect

Aug 1, 2017

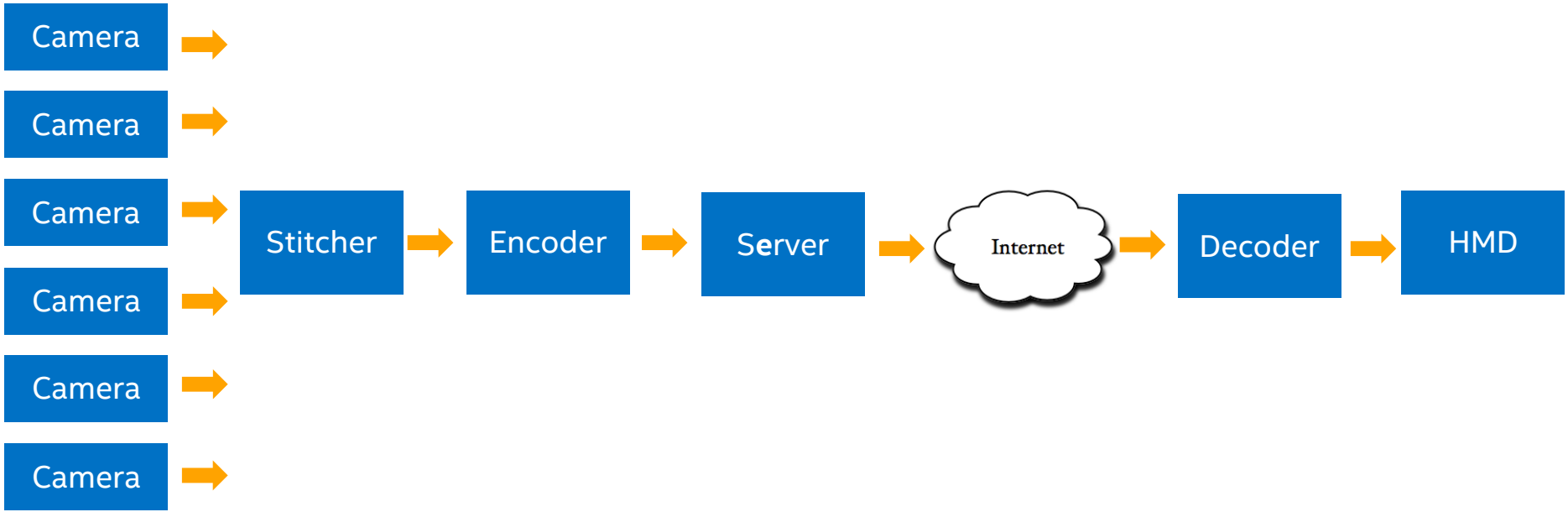
What is 360° Video?

- Also called omnidirectional video, panoramic video, spherical video, VR video
- Not really the same as VR, even though can be viewed using VR Head Mounted Display
  - VR content is rendered by a graphics engine, according to the position/orientation of the HMD
  - 360° video is
    - Captured by cameras, usually with multiple lenses
    - Capture wider field-of-view than is viewed at a particular time
    - Region of the larger field-of-view that is viewed is determined by movement of head mounted display
    - May be stereoscopic or monoscopic

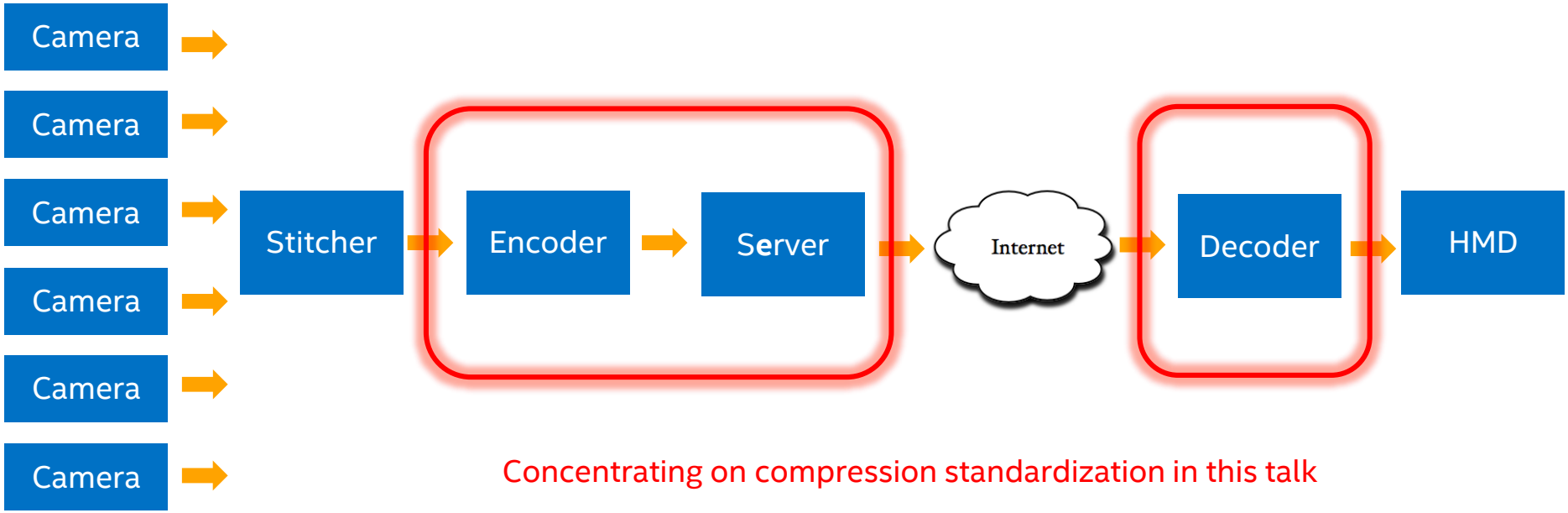
360° Video



360° Video Creation and Distribution



360° Video Creation and Distribution



Concentrating on compression standardization in this talk

Stitching also important for 360 video, but won't be standardized.

360° Video Standards Bodies

- MPEG and ITU-VCEG: Will define codec standards
  - Includes Joint teams: JCT-VC, JVET
- MPEG Liaison organizations: Use MPEG codec standards, will provide requirements, define specific interoperability profiles
  - DVB
  - 3GPP
  - VR Industry Forum
- Others: Khronos, Open XR, etc.

360 Video Standardization

Short term: Using HEVC for 360° video

- Targeting October 2017 technical finalization
- New HEVC version with Supplemental Enhancement Information (SEI) messages
- Omnidirectional Media Application Framework (OMAF) systems layer

Long term: New Future Video Coding standard, H.266/MPEG-I video

- Targeting EOY 2020 technical finalization
- Goal of 50% bitrate reduction vs. HEVC
- Includes 360° video, in addition to normal rectangular 2-D video
- Can include new coding algorithms specifically aimed at efficient coding of 360° video

Call for Evidence Future Video Coding standard

MPEG and VCEG issued a Call for Evidence, responses reviewed July 2017

- 2 responses in SDR category
- 2 responses in HDR category
- 3 responses in 360 video category

Significant objective and subjective benefit shown over HEVC in all 3 categories

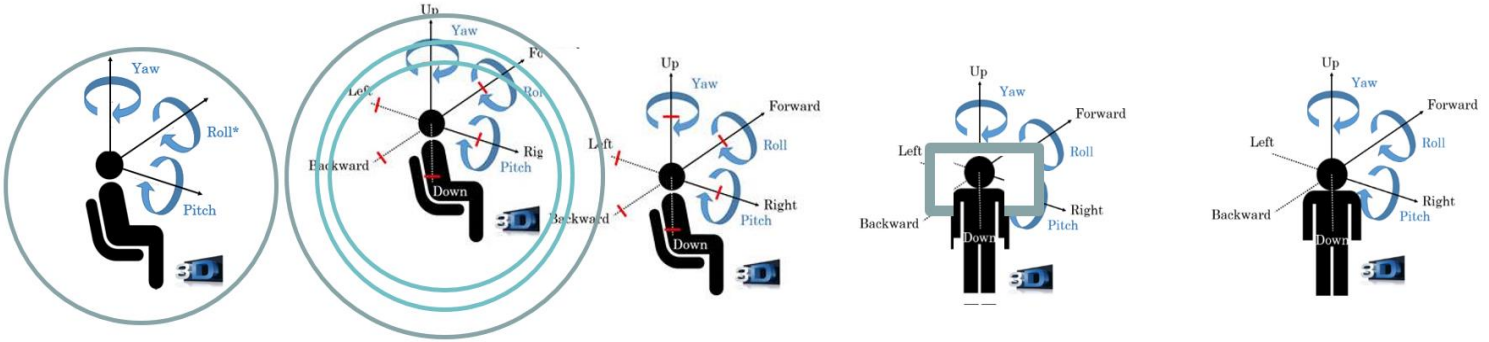
- 30-35% objective gains, higher subjective gains
- Less significant benefit for 360 video than other categories

Call for Proposals to be issued, with responses due Feb 2018

MPEG-I

- A new MPEG-I standard is planned, with multiple parts, with phases to span from normal 2D coding to 360° 3DoF to 6 DoF
  - Part 3: “Immersive Video Coding”, a.k.a. H.266
    - Will include 360 3DoF video
    - May include 3 DoF+, may go into another part
- Free viewpoint TV (FTV) and Light fields requirements harmonized, are multi-view video + depth based
- Point Cloud coding is 3D graphics based

MPEG-I (from April 2017 meeting)



3DoF 3DoF+ windowed 6DoF

omni- 6DoF

directional 6DoF

360 ° Video representation

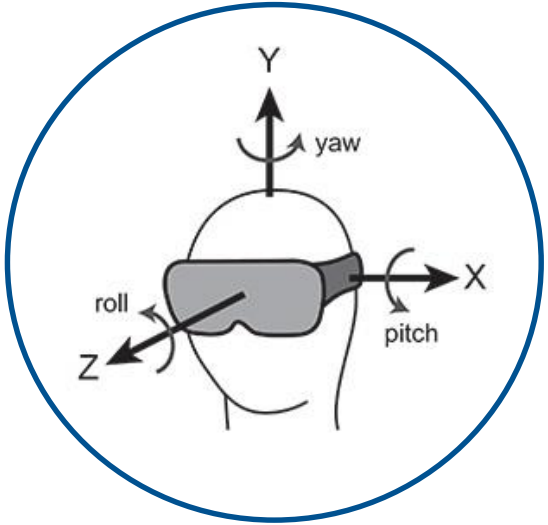
360 video represents a sphere (or sometimes a cylinder)

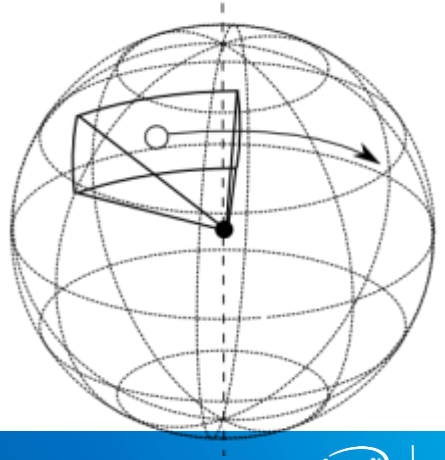
- 360° horizontal x 180° vertical

The viewer is at a fixed position in the center looking out, and selects the viewport with 3 Degrees of Freedom (3 DoF): Yaw, Pitch, Roll

Sphere is mapped to a rectangle for use by video codec

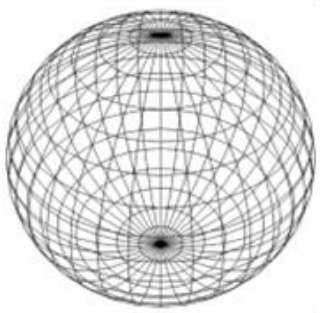
- Equirectangular projection, cube map, etc.

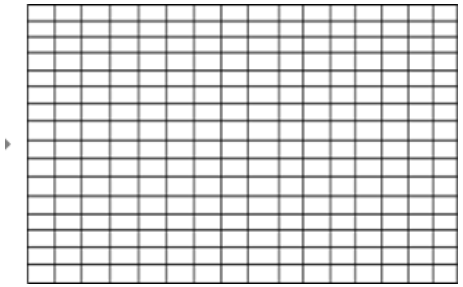


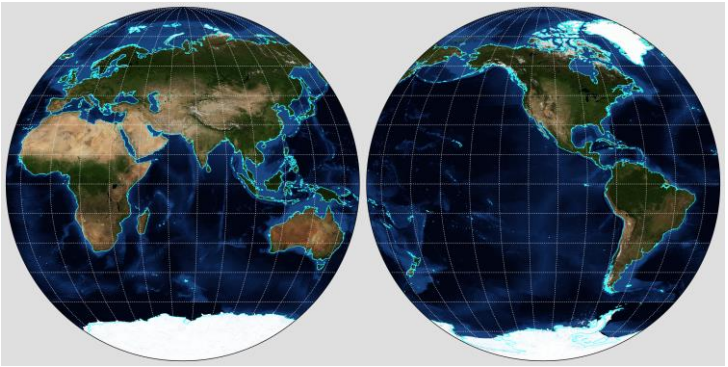


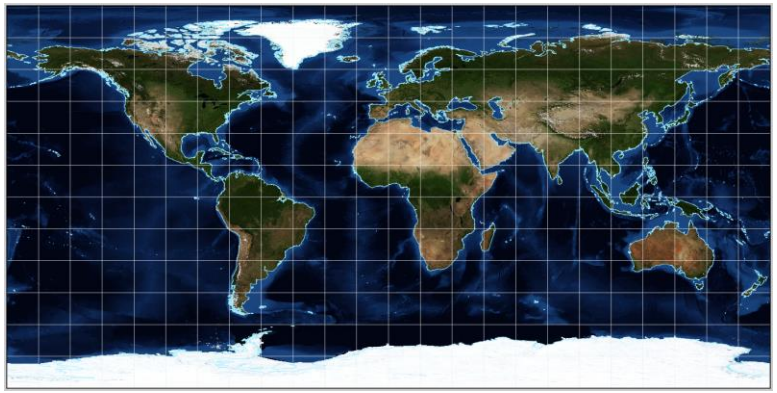
Equirectangular projection Format

Most commonly used 360 video format
Has been used in mapping for centuries
Introduces distortion, especially at poles

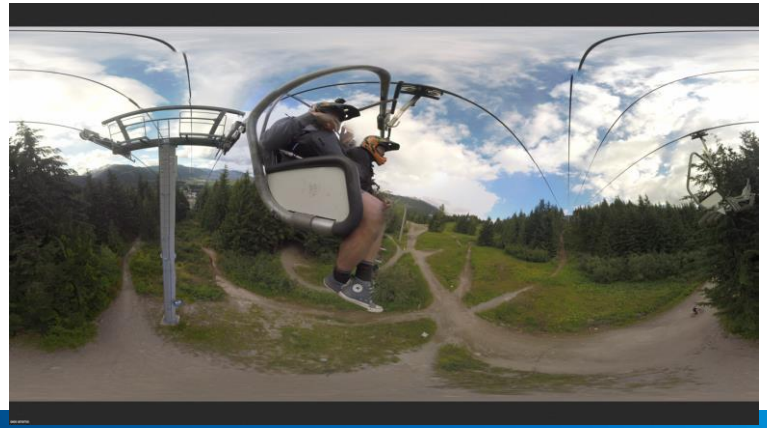








Selected equirectangular content from GoPro 8K



Content from InterDigital

Content from: Nokia/ Tampere University

Selected content from Letin VR 4K

From JVET-D0179

Using HEVC for 360° Video

No change to core codec

- Can be used with any profile, e.g. Main, Main 10

Use Supplemental Enhancement Information (SEI) message to send metadata inband

Define projection format(s)

- Only equirectangular projection and cube map projection formats defined now
- Other projection formats have been proposed

Remaining problem: decode whole video frame but view portion

HEVC SEI Messages

Equirectangular projection SEI

Cube map projection SEI

Omnidirectional viewport

Output documents:

- **JCTVC-AA1005 (decisions of April 2017 meeting)**
- **JCTVC-AB1005 (decisions of July 2017 meeting) – not yet available**

- **All JCT-VC docs publicly available at http://phenix.int-evry.fr/jct/**

HEVC SEI Messages: Equirectagular projection SEI (formerly Omnidirectional projection indication SEI

Provides equations to map equirectangular projection to sphere

Allows indication of angles of coverage of less than a full sphere

- E.g. 180x180, 360x120
- Also allows more than 360° coverage for padding

Stereo support using frame packing

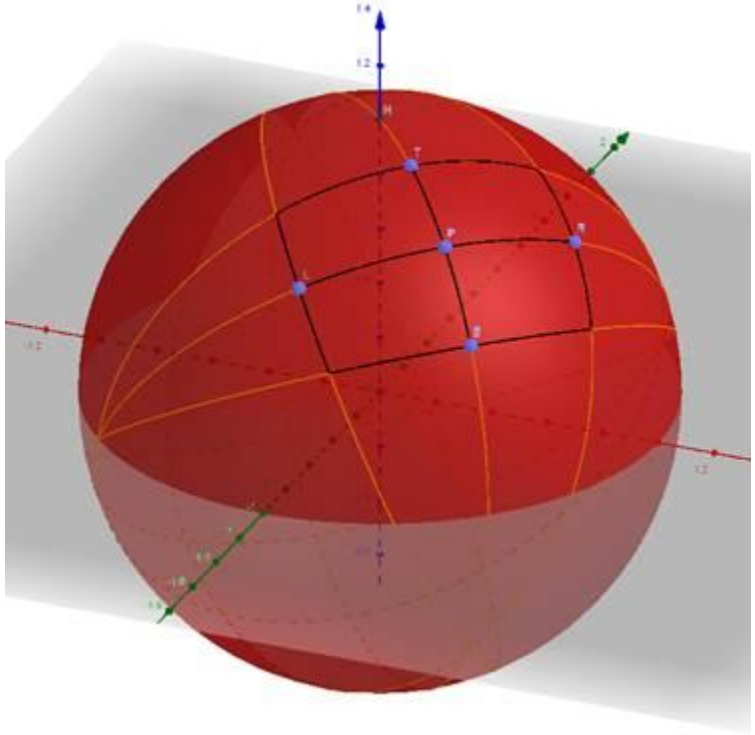
- Three options: left/right, top/bottom, frame sequential

Rotation of region in (yaw, pitch, roll)

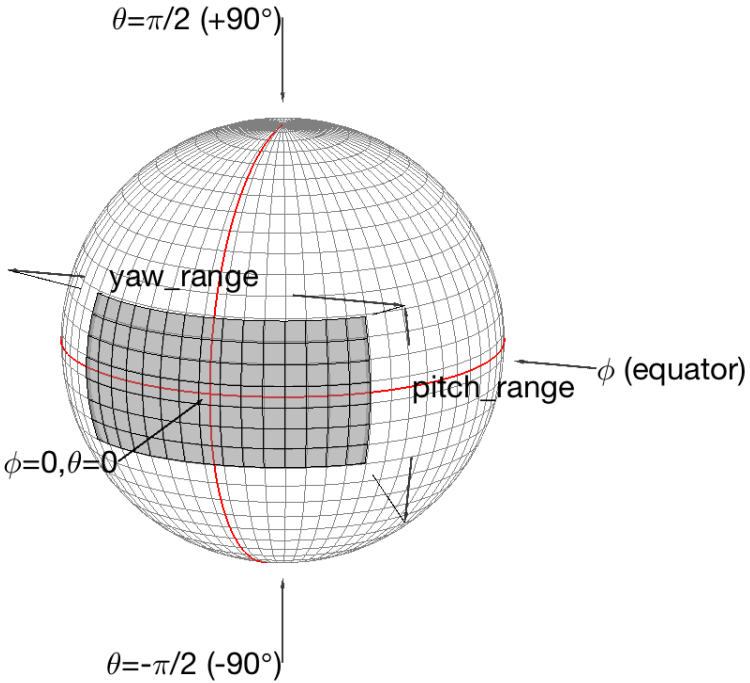
HEVC SEI Messages: Omnidirectional viewport SEI

When viewing omnidirectional video, the viewer typically views only a small portion of the full 360° x 180° video at a given time

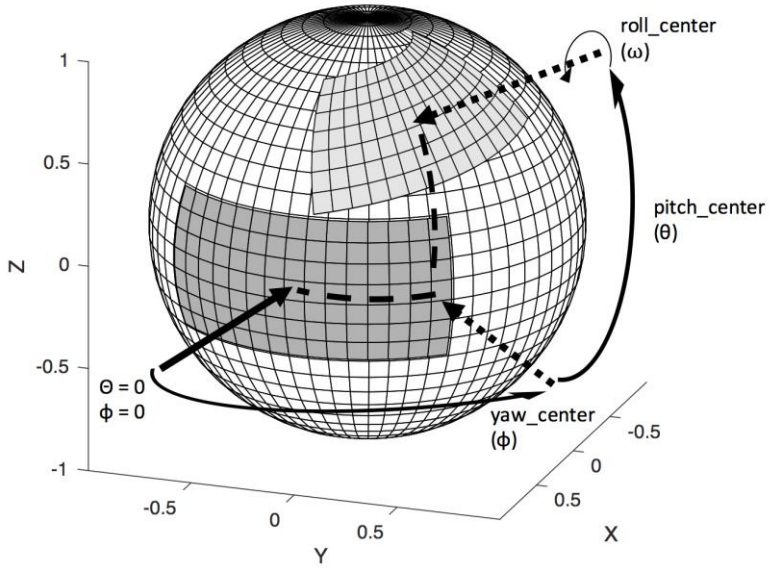
- A viewer may miss the “most interesting” area of the video
- Content creator can recommend one or more regions to view
- Viewer can follow the recommendation or override it
- Can change for every frame



Indication of region of a sphere



(a) region with yaw and pitch offsets

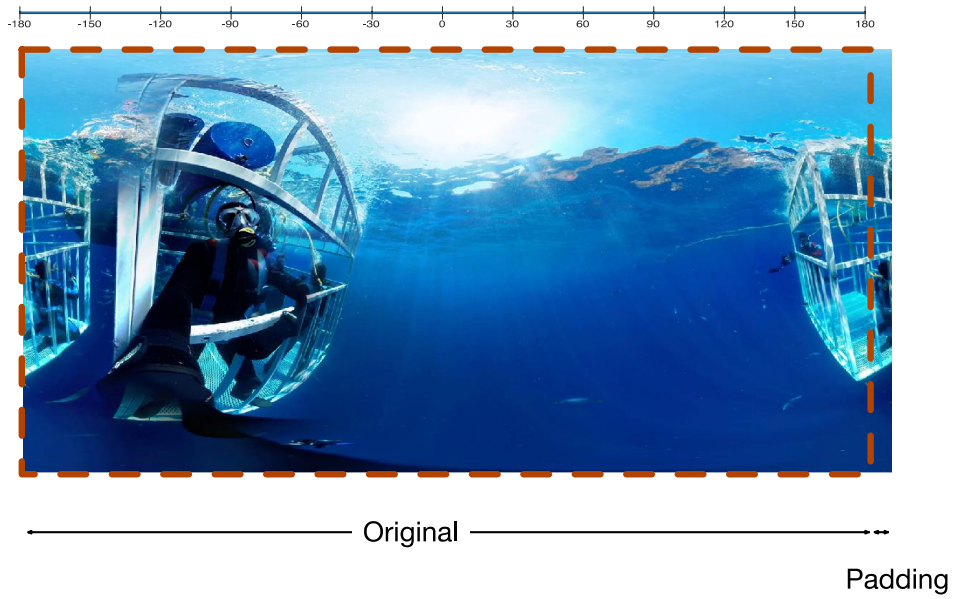


(b) region (lighter shade) with yaw, pitch, and roll offsets

Padding support

Informal subjective tests have shown subjective artifacts along left-right edge border of equirectangular projection format content

- Left and right edges are connected in the spherical representation, but discontinuous in the rectangular ERP format
- Allowing yaw_range to exceed 360 degrees allows padding of content to reduce subjective artifacts along the discontinuous edge

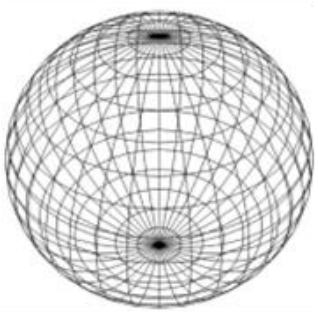


Rotation for improved coding efficiency

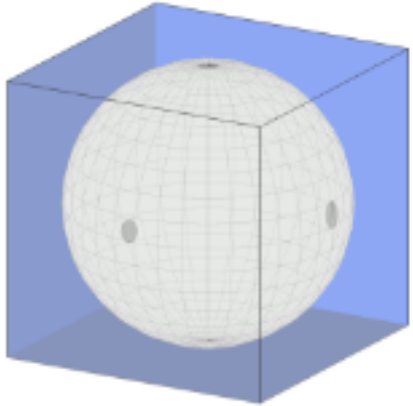
- Coding efficiency can be improved by applying rotation before encoding and after decoding for some equirectangular projection sequences
  - Up to 18% savings
  - Modifies impact of warping
- Used to convey rotation parameters (pitch, yaw, roll)

Projection Formats

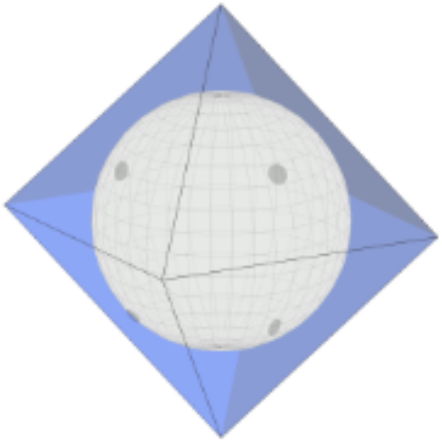
- Warping in equirectangular projection format can hurt video compression efficiency, because motion models in HEVC, AVC assume block translational motion
- Other projection formats have less warping, but introduce more discontinuities, which can also hurt video compression efficiency
- JVET studying coding efficiency of various projection formats
  - Experimental conditions and objective/subjective video quality metrics must be developed



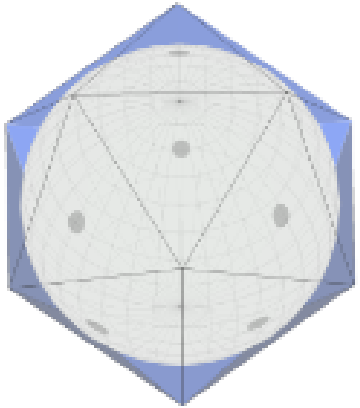
Sphere



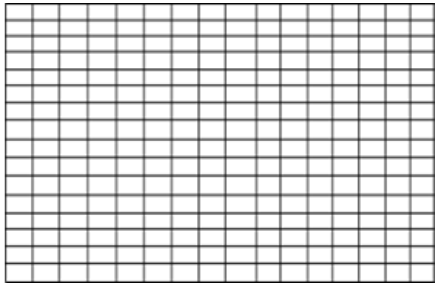
Cube



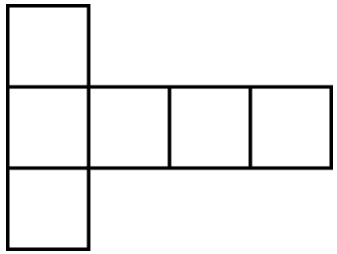
Octahedron



Icosahedron



Equirectangular

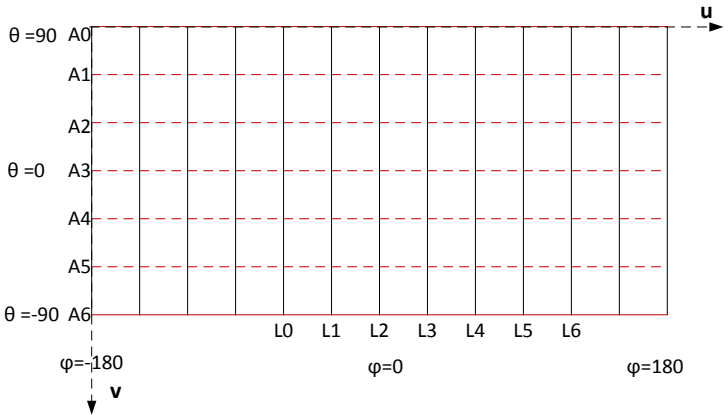


Cubemap

Octahedral

Icosahedral

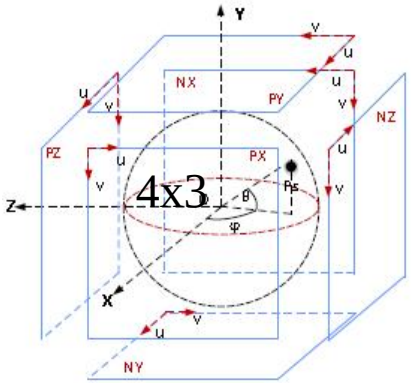
Equirectangular



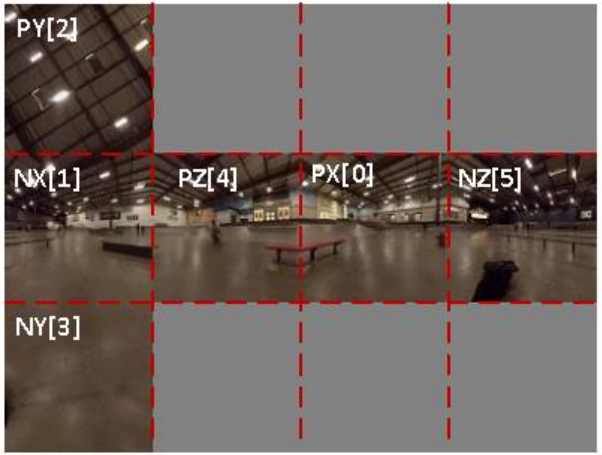
From JVET-D0021

Cube maps with different layouts

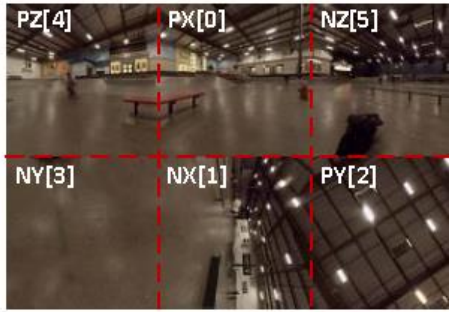
Compact formats can improve coding efficiency and reduce complexity, memory bandwidth

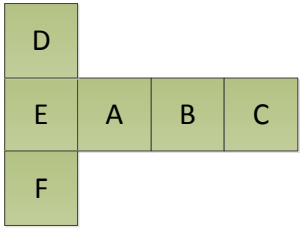


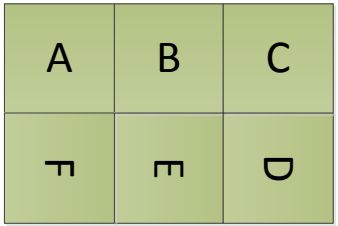
Native 4x3



Compact 3x2

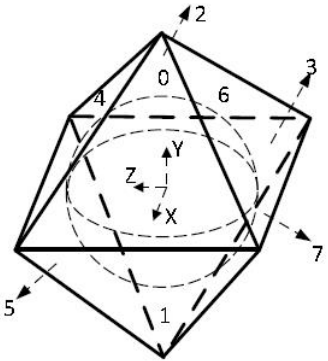


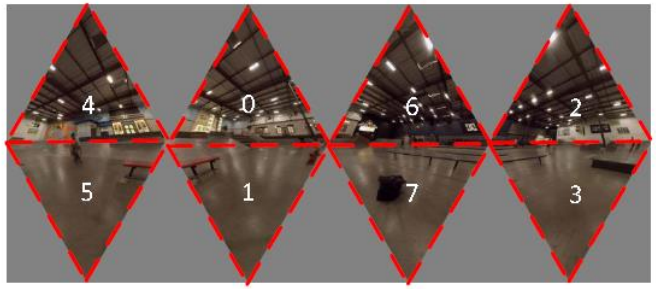




From JVET-D0021

Octahedron projection (8 sided solid)



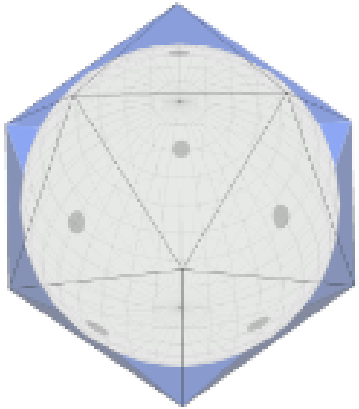


Non-compact frame packing

Compact frame packing

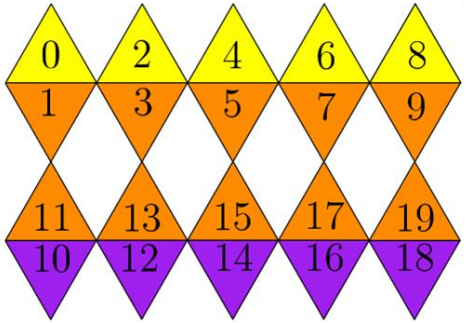
From JVET-D0021

Proposed projection and packing formats

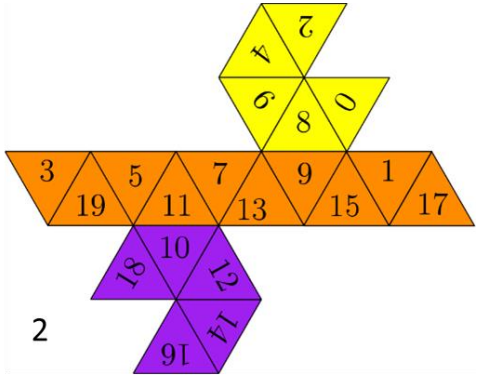


Icosahedron

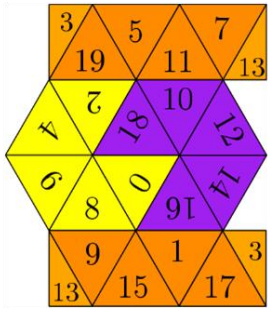
1



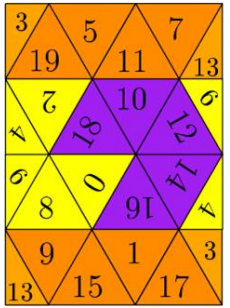
2



3

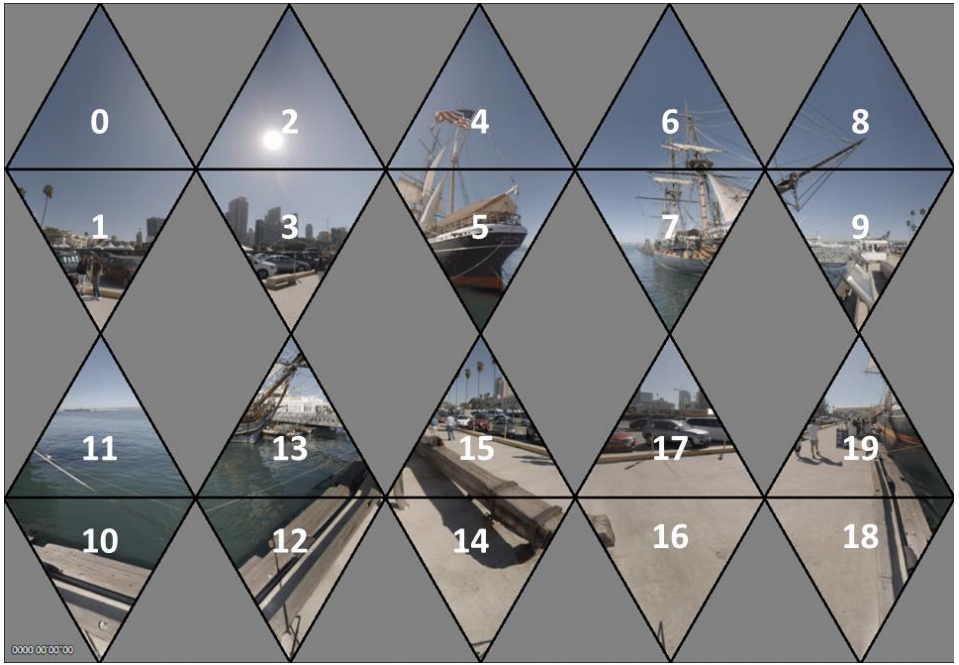


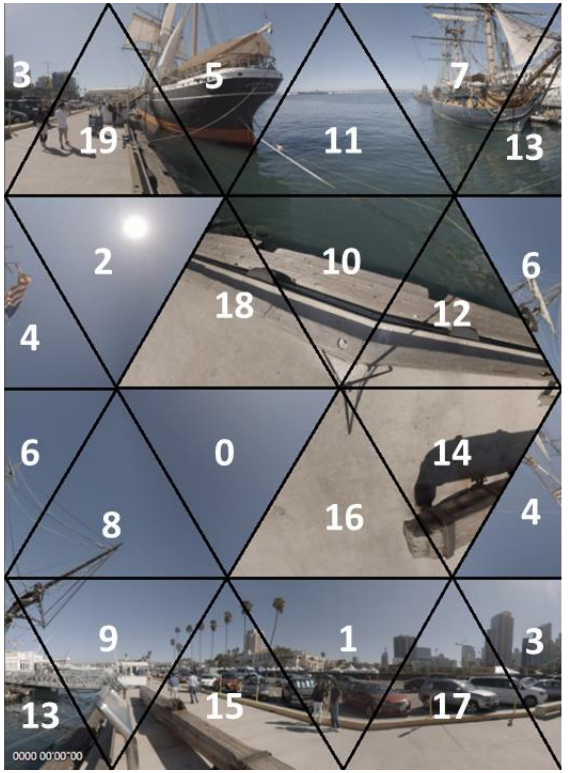
4



From JVET-E0029

Compact Icosahedron





From JVET-E0029

Comparison of projection formats vs Equirectangular

Projection	WS_PSNR End-to-End		
	Y	U	V
CMP	-1.7%	-1.5%	-1.6%
EAP	8.4%	-2.9%	-3.8%
OHP1	4.5%	10.4%	10.5%
ISP1	-1.3%	1.8%	1.6%
OHP2	4.7%	14.0%	13.8%
SSP Hor	-7.7%	-3.6%	-4.3%

CMP vs. ERP	WS-PSNR (End to End)		
	Y	U	V
Train	11.3%	12.6%	10.9%
Skateboarding_trick	9.7%	9.3%	9.3%
Skateboarding_in_lot	-8.5%	-9.6%	-10.8%
Chairlift	-18.2%	-13.3%	-14.2%
KiteFlite	-4.9%	-1.5%	-3.8%
Harbor	1.1%	-3.0%	-2.1%
PoleVault	-2.4%	-5.8%	-4.9%
AerialCity	4.2%	3.7%	3.6%
DrivingInCity	4.8%	7.2%	9.2%
DrivingInCountry	-14.6%	-14.3%	-13.6%
Overall	-1.7%	-1.5%	-1.6%

Negative numbers show bitrate reduction

Best format very content dependent

Averages mask per sequence differences

From JVET-E0008

From JVET-D0040

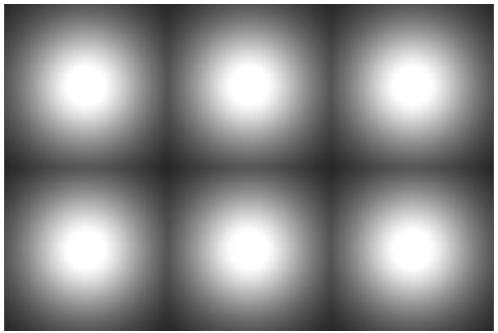
Objective Test Metrics for 360 Video

All methods are full reference: compare output video sequence to input

- PSNR (Peak Signal-to-Noise Ratio)
  - Typical for “normal” video
  - All pixels given equal weight
- New metrics (including WS-PSNR) consider spherical representation, warping of rectangular formats, to apply uneven weighting to the pixels in rectangular format to better represent equal weighting in a sphere

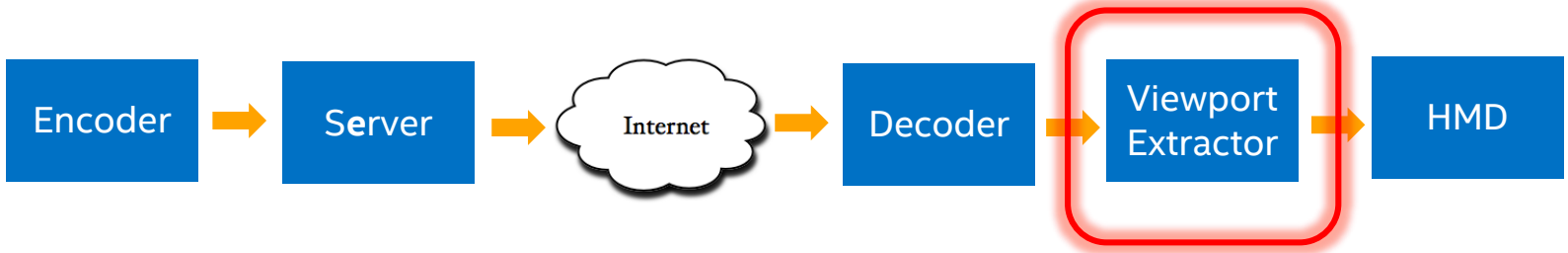
WS-PSNR weights

for equirectangular



For cube map

Viewports for 360 video

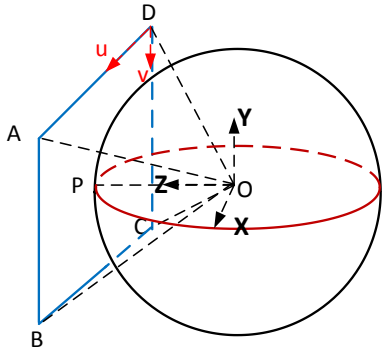


- Only a small viewport from full omnidirectional video is actually viewed at a particular time
- Places large burden on network bandwidth, decoder complexity to code and transmit high resolution video, but view low resolution video
  - 360° x 180 ° video coded
  - Viewport of ~110 ° x 90° or 90° x 90° viewed
- HEVC standard normally requires that entire coded picture be decoded
  - Many proposed approaches to reduce network and decoder burden

Viewport generation

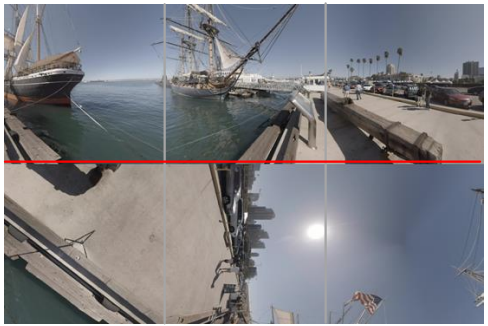
View only a portion of the full 360 video

Equirectangular projection example



Subjective Testing of 360 video

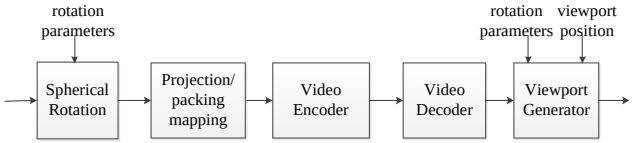
- Static viewports
- Dynamic viewports
- Problem: Mapping spherical video to rectangle leads to discontinuous edges, which cause compression artifacts
  - Different projection formats have discontinuous edges at different locations
- “Evil viewports” force discontinuous edges to occur at same location



Dynamic viewport Example

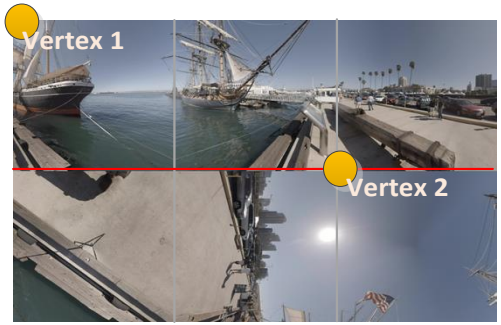
<Video>

Creation of “Evil Viewport”



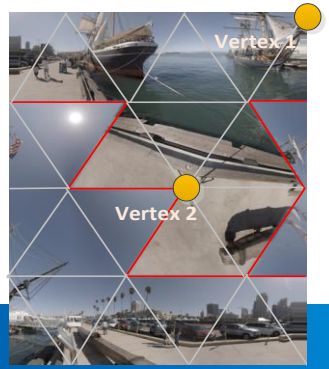
ERP

Rotated ERP for evil viewport

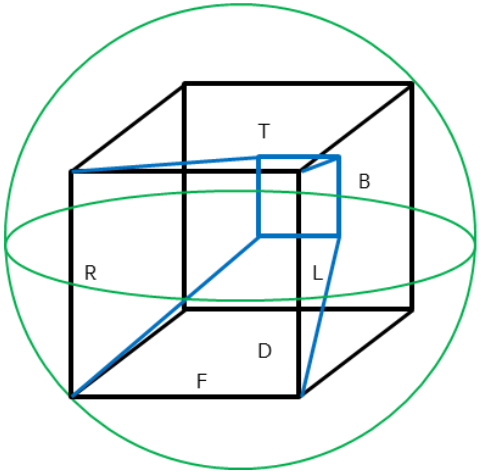


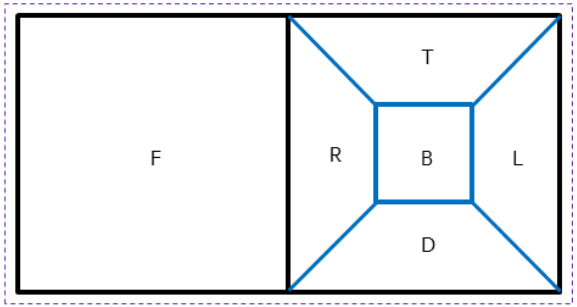
CMP

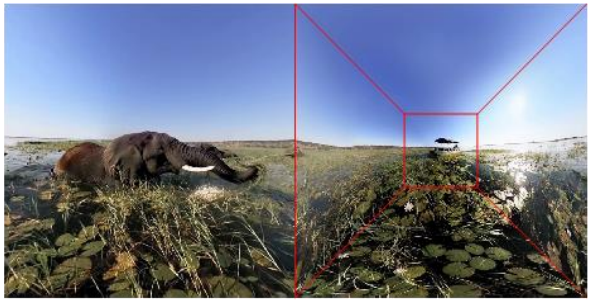
Rotated CMP for evil viewport



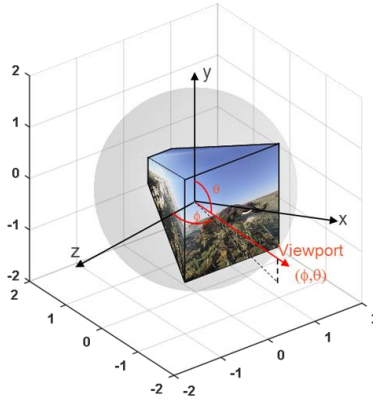
Viewport based projections: Truncated square pyramid

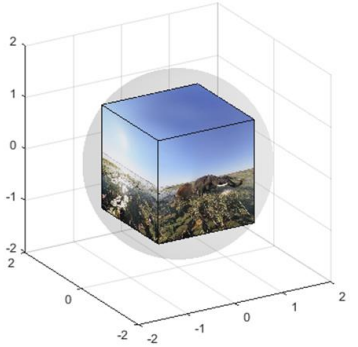






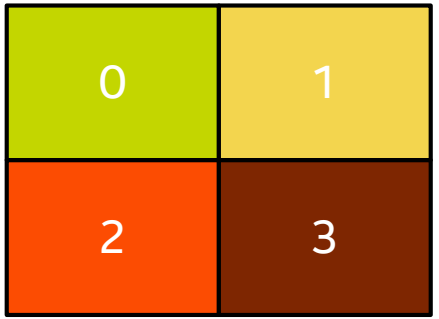
360tools



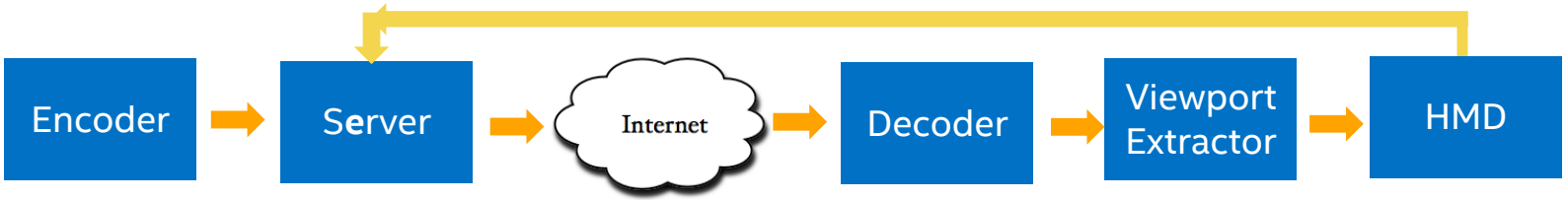


From JVET-D0071

Tile-based decoding approaches
Defined in informative annex in OMAF



- HEVC tiles normally can't be individually decoded for a sequence
  - For P and B frames, motion vectors can point outside of current tile
    - Predictions allowed from other tiles in previously coded frame
- HEVC already has "Temporal Motion Constrained Tile Sets" SEI message
  - Optional for an encoder to use
  - Neighboring tiles are grouped into tile sets, which persist for entire coded video sequence (Each CVS typically begins with an I frame)
  - Enables decoding of individual tile sets for the sequence by restricting motion between tile sets



MPEG-I

Part 1: Architecture for Immersive Media (Jan 2018)

Part 2: Application Framework for Omnidirectional Media (Oct 2017)

Part 3: Immersive Video (Oct 2020)

Part 4: Immersive Audio (?)

Part 5: Point Cloud compression (Apr 2019?)

Part 6: 3DoF+, 6 DoF, Light fields (2020/2021?)

OMAF: Omnidirectional Media Application Framework

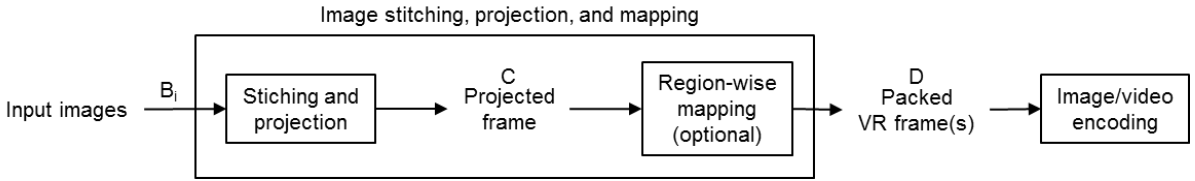
Targeting October 2017 finalization, EOY 2017 standardization

Standardized method for signaling metadata for distribution of 360 video using HEVC

- Support of projections
- Signaling of ROI
- Signaling of packaging and region based mapping

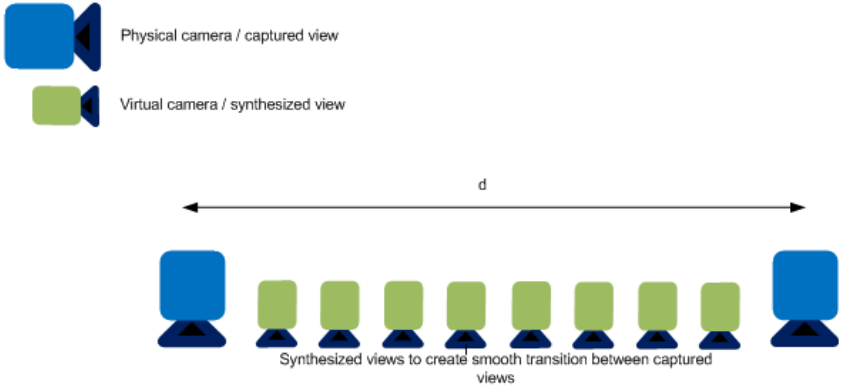
Aligned with JCT-VC SEI message syntax

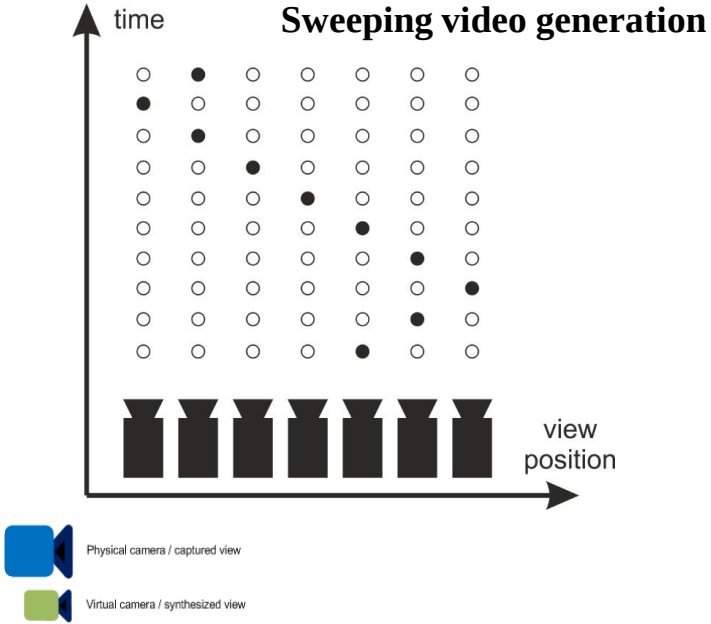
- MANY different proposed projection formats (>15)
  - Only equirectangular, cube map, and fisheye defined
- Informative section describing viewport dependent approaches
  - Motion constrained tile sets
  - Simulcast and scalable HEVC (SHVC)



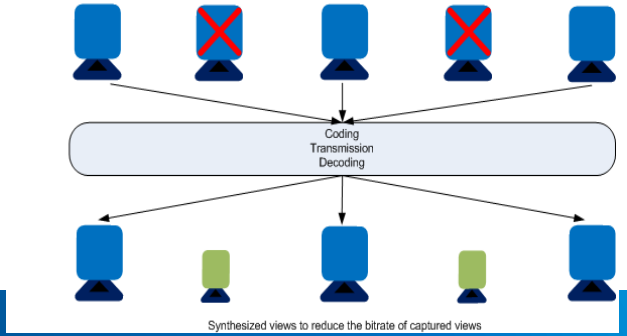
Lightfield Camera

Free view point TV





- Multi-camera capture of views and depth
- Similar to MV-HEVC and 3D-HEVC
- View synthesis for virtual camera views



Point Cloud Coding

- More recent attention within the standards group
- Call for Proposals issued, with responses due Oct 2017

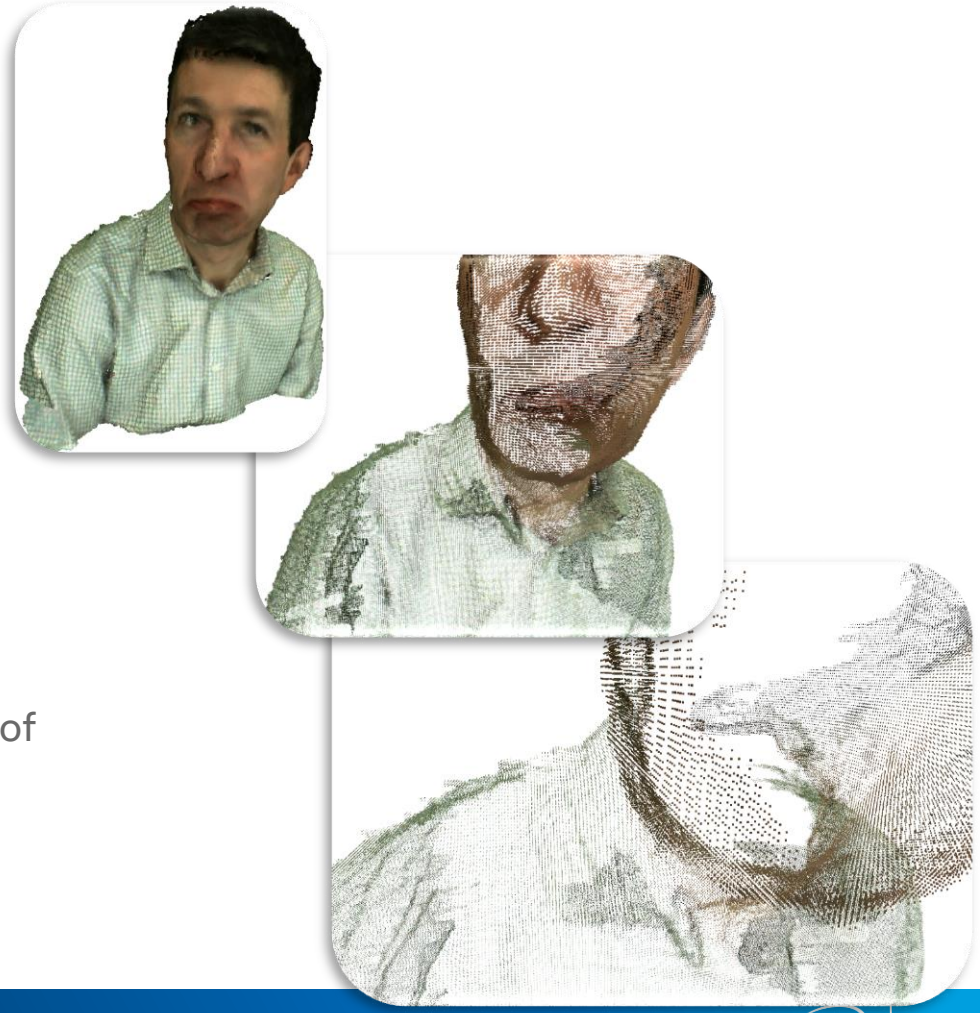



Video sequences from 8i

Point Cloud Video Example

The point cloud format

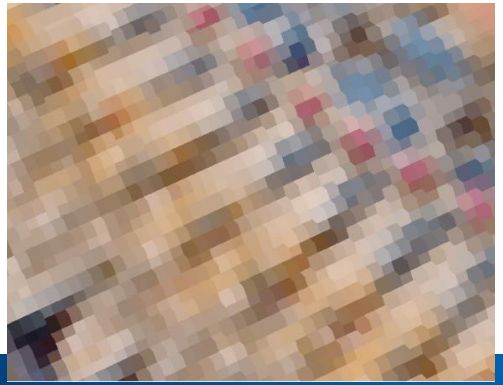
- **A collection of points**
- **Not related to each other**
  - no order
  - no local topology (no mesh!)
- **Each is point is the given of**
  - a position (X,Y,Z)
  - a color (R,G,B) or (Y,U,V)
  - possibly other things like transparency, time of acquisition, etc.

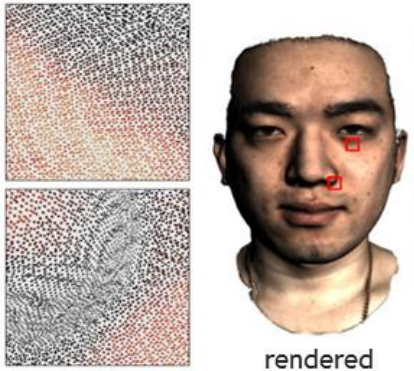


From m40715

How to render point clouds?

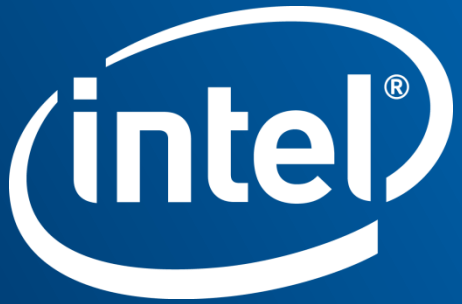
- **Giving size to points**
  - Splats, rectangles, cubes (=3D pixels)
  - Trade-off size vs. texture high frequency
- **A demo using PCC contents (and renderer)**
  - Mitsubishi content
  - 8i content





Summary

- 360° Video can be compressed & distributed using HEVC codec, with extensions to provide metadata describing pre- and post-processing
- Future efforts aim to provide more immersive experiences with 6 DoF video


intel®