Usage of Multiplane Image Information SEI Message for Distribution of Lightweight 3DoF+ Immersive Content with Conventional 2D Decoder

Taoran Lu, Peng Yin, Sejin Oh, Sean McCarthy

DOLBY LABORATORIES INC

MILE HIGH VIDEO 2024

Outline

- Introduction
- Design of the Multiplane Image Information SEI message
- Examples and operation points
- System level considerations

Multiplane Image (MPI) Representation

- A well-established method for novel viewpoint synthesis
- Can handle challenging occlusions and reflections
- Very low computational requirements for 3D scene reconstruction

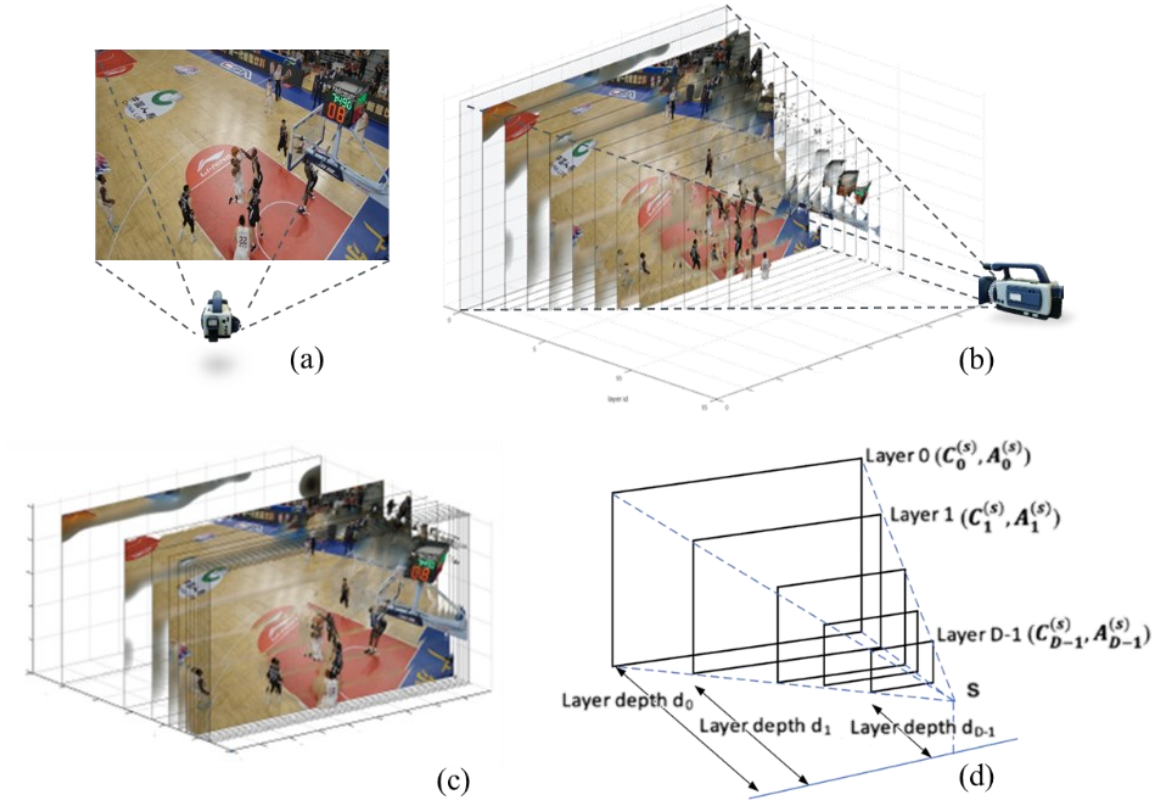


Figure 1: A camera-captured image and its MPI representation.

(a) original camera captured 2D image; (b) MPI representation of the layers in perspective projection with respect to a camera position S; (c) MPI representation illustration with actual layer size and depth information; (d) abstraction of the MPI representation with mathematical notations

View Synthesis using MPI

- Given D-layer MPI image {($\boldsymbol{C}_i$, $\boldsymbol{A}_i$) for i=0 … D-1
- **Warping** $T_{\boldsymbol{v}_s,\boldsymbol{v}_t}()$ current view point position ($\boldsymbol{v}_s$) to new view point position ($\boldsymbol{v}_t$)
  - $\begin{bmatrix} u_s \\ v_s \\ 1 \end{bmatrix} = \boldsymbol{K}_s(\boldsymbol{R} - \frac{\boldsymbol{t}\boldsymbol{n}^T}{\boldsymbol{d}})(\boldsymbol{K}_t)^{-1}\begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix}$
  - $\boldsymbol{C}_i^t = T_{\boldsymbol{v}_s,\boldsymbol{v}_t}(d_i, \boldsymbol{C}_i)$
  - $\boldsymbol{A}_i^t = T_{\boldsymbol{v}_s,\boldsymbol{v}_t}(d_i, \boldsymbol{A}_i)$
- **Compositing**
  - $\boldsymbol{W}_i^t = (\boldsymbol{A}_i^t \cdot \prod_{j=i+1}^{D-1}(1 - \boldsymbol{A}_i^t))$
  - Texture in new view $\boldsymbol{C}^t = \sum_{i=0}^{D-1} \boldsymbol{C}_i^t \boldsymbol{W}_i^t$

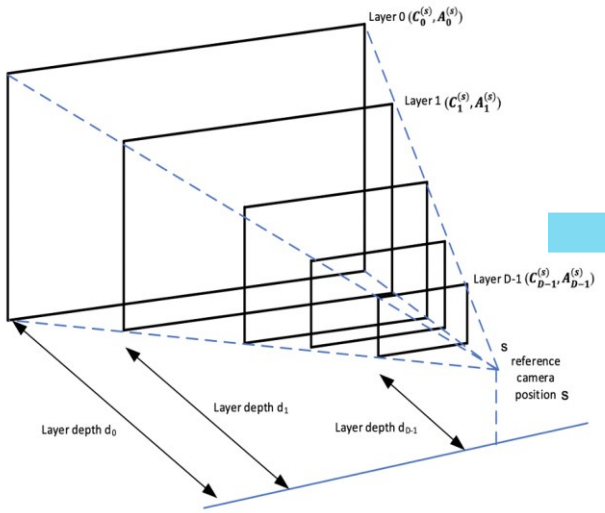
- $\boldsymbol{C}_i$, $\boldsymbol{A}_i$ texture and opacity information
- $\boldsymbol{d}$ layer depth information
- $\boldsymbol{K}_s$,$\boldsymbol{K}_t$ camera intrinsic model
- $\boldsymbol{R}$, $\boldsymbol{t}$ extrinsic camera model for rotation and translation (novel view)

MPI Metadata Carriage in International Standard

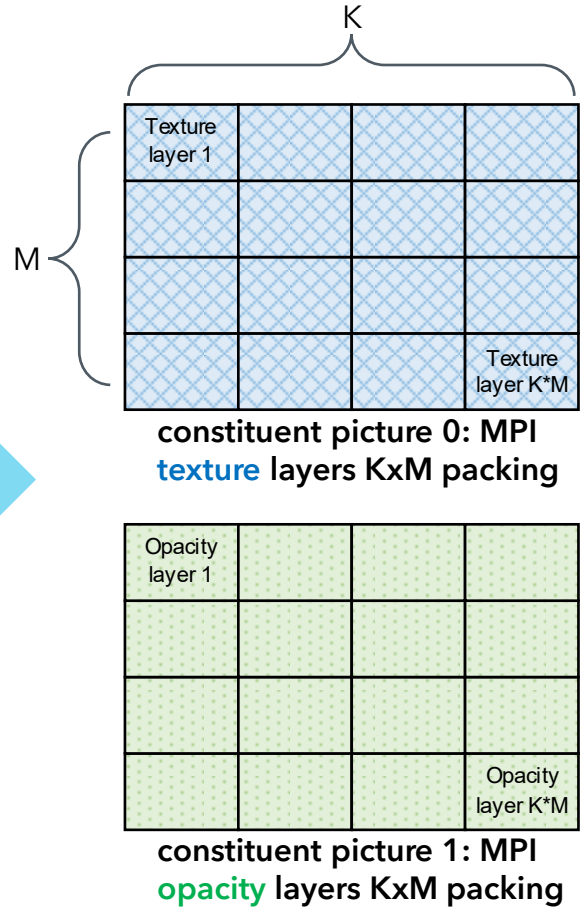
- **Goal**
  - To create a container/carrier for the MPI based volumetric representation
  - To enable distribution of this “container” with existing video distribution infrastructure.
- **Our contribution**
  - Design method to pack high-dimensional volumetric data into “conventional” 2D video
  - Provide the minimal set of metadata for de-packing and enable view synthesis

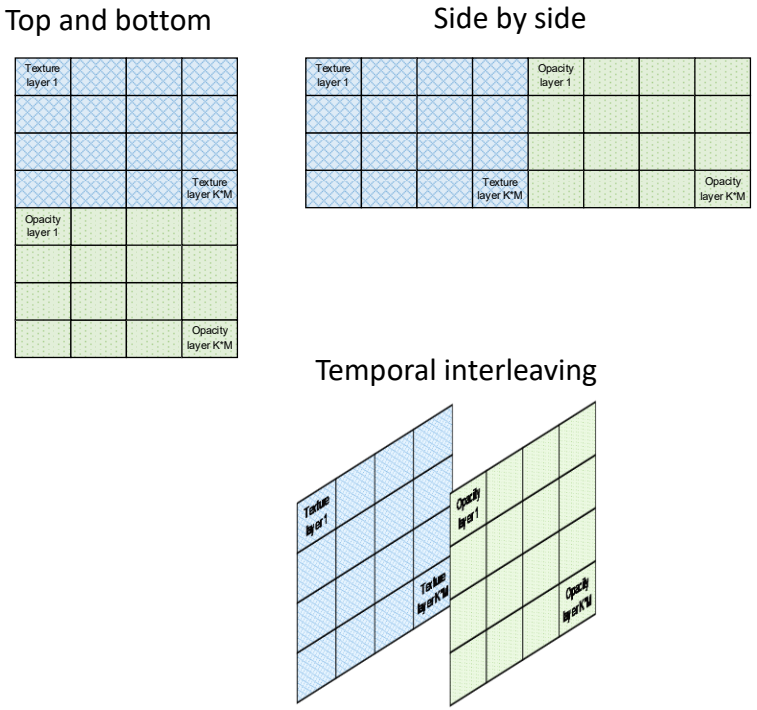
MPI layer packing and arrangement

- MPI texture and opacity layers are first spatially packed by D=KxM arrangement to form *constituent pictures*
- two constituent pictures can be packed by three different options



D layers (D = K x M)





MPI Metadata: What are essential metadata for single view use case?

- **Basic MPI info:**
  - Number of MPI layers
  - Depth of MPI layers
- **Packing/arrangement:**
  - Texture/alpha arrangement: texture/alpha temporal interleaving or spatially packed, side-by-side or top-and-bottom
  - Layer arrangement: KxM on a 2D grid

multiplane_image_information(payloadSize) {	Descriptor
mpii_num_layers_minus1	ue(v)
mpii_layer_depth_equal_distance_flag	u(1)
if(mpii_layer_depth_equal_distance_flag) {	
depth_rep_info_element(ZNearSign, ZNearExp, ZNearMantissa, ZNearManLen)	
depth_rep_info_element(ZFarSign, ZFarExp, ZFarMantissa, ZFarManLen)	
} else	
for(i = 0; i <= mpii_num_layer_minus1; i++)	
depth_rep_info_element(ZSign[i], ZExp[i], ZMantissa[i], ZManLen[i])	
mpii_texture_opacity_interleave_flag	u(1)
if(mpii_texture_opacity_interleave flag == 0)	
mpii_texture_opacity_arrangement_flag /* 0: TaB, 1:SbS */	u(1)
mpii_frame_num_layers_in_height_minus1	ue(v)
}	

Layer depth related: depth_rep_info_element() has exact same definition as in DRI SEI (ITU-T Rec. H.274 clause 8.22.1.1)

Layer packing & arrangement related

Temporal interleaved packing

- Temporal interleaved packing substantially outperforms the spatial packing for the capability to support the decoding of same MPI with lower decoder level requirements.

Level	Max luma sample rate MaxLumaSr (samples/sec)	Max luma picture size MaxLumaPs (samples)
5.1	534773760	8912896
5.2	1069547520	8912896
6	1069547520	35651584
6.1	2139095040	35651584
6.2	4278190080	35651584

(a) HEVC/VVC level constraints

SbS/TaB packing			
640x480	960x540	1280x720	1920x1080
Y	Y	Y	
Y	Y	Y	
Y	Y	Y	

(b) capability to support with SbS/TaB packing

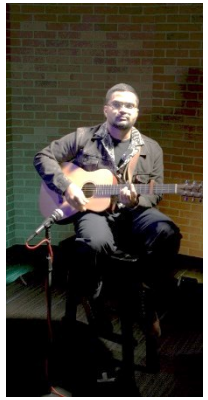
temporal interleaved packing			
640x480	960x540	1280x720	1920x1080
Y	Y		
Y	Y		
Y	Y	Y	
Y	Y	Y	Y
Y	Y	Y	Y

(c) capability to support with temporal interleaved packing

Content and compression settings for MPI coding experiment

Breakfast ©InterDigital

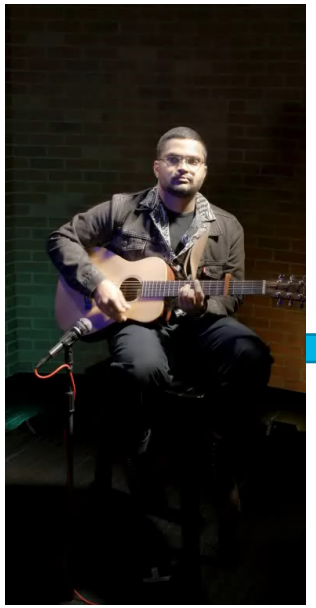
CBAbasketball ©Alibaba



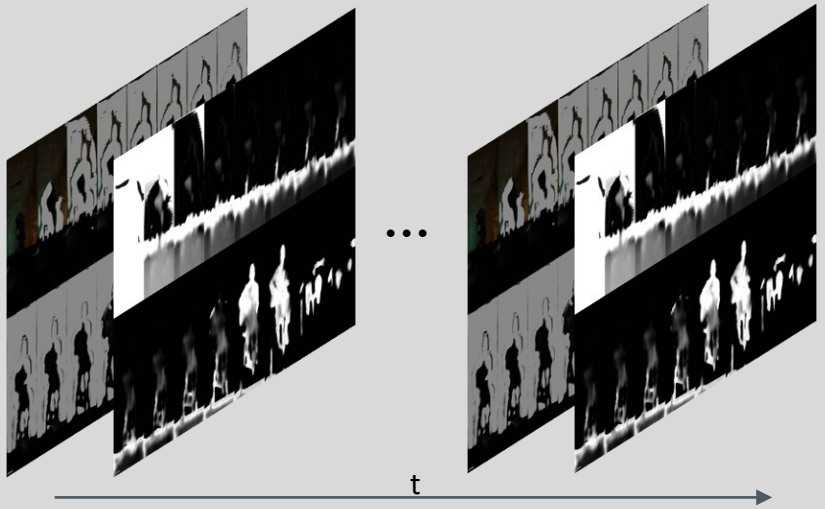
GuitarMan ©Dolby

	source image size	MPI layer arrangement	packed picture: width x height @ fps Interleaved packing
Breakfast	896x512	4x4	3584x2048@60p
CBAbasketball	896x512	4x4	3584x2048@60p
GuitarMan	384x768	2x8	3072x1536@60p

Example: GuitarMan



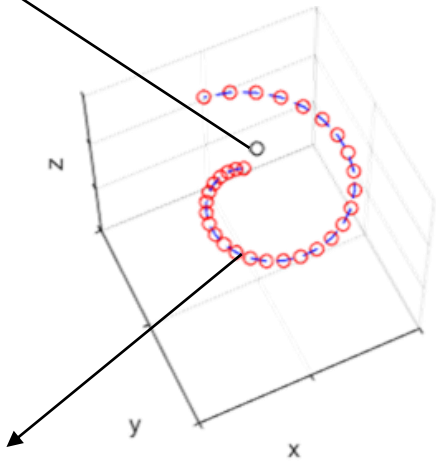
source video to MPI representation and SEI



multiplane_image_information(payloadSize) {	Descriptor
mpii_view_id	ue(v)
mpii_num_layers_minus1	ue(v)
mpii_layer_depth_equal_distance_flag	u(1)
if(mpii_layer_depth_equal_distance_flag) {	
depth_rep_info_element(ZNearSign, ZNearExp, ZNearMantissa, ZNearManLen)	
depth_rep_info_element(ZFarSign, ZFarExp, ZFarMantissa, ZFarManLen)	
} else	
for(i = 0; i <= mpii_num_layer_minus1; i++)	
depth_rep_info_element(ZSign[i], ZExp[i], ZMantissa[i], ZManLen[i])	
mpii_texture_opacity_interleave flag	u(1)
if(mpii_texture_opacity_interleave flag == 0)	
mpii_texture_opacity_arrangement_flag /* 0: TaB, 1:SbS */	u(1)
mpii_picture_num_layers_in_height_minus1	ue(v)
}	

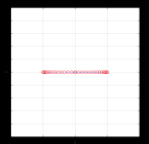
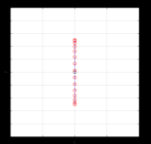
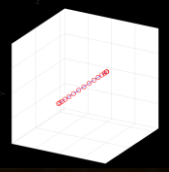
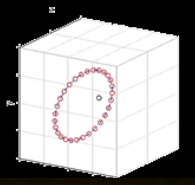
MPI compression & transmission with SEI

MPI rendering (at ref camera position)



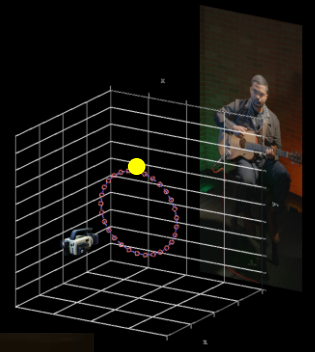
MPI rendering (along a virtual camera position path)

Novel view rendering examples: virtual path can be set freely at device-side




Novel view rendering at different compression quality (VVC main 10)



uncompressed

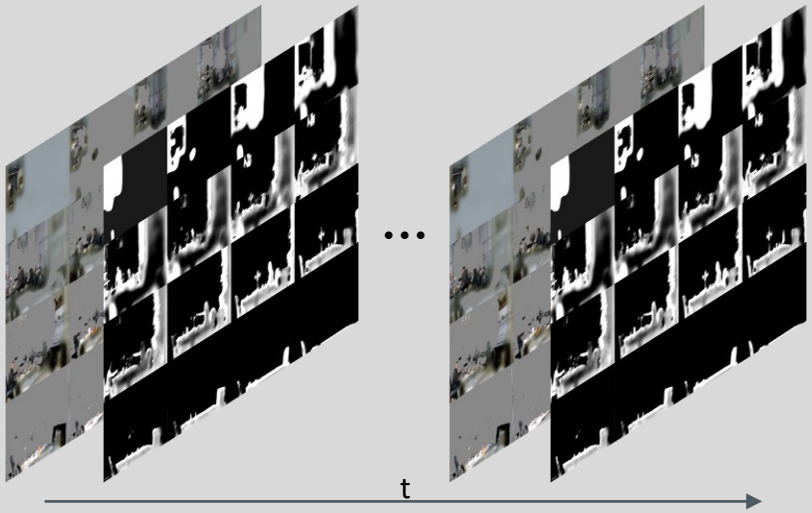
High BR
10.2 Mbps

Mid BR
5.8 Mbps

Low BR
3.2 Mbps

Example: Breakfast

source video to MPI representation and SEI



multiplane_image_information(payloadSize) {	Descriptor
mpii_view_id	ue(v)
mpii_num_layers_minus1	ue(v)
mpii_layer_depth_equal_distance_flag	u(1)
if(mpii_layer_depth_equal_distance_flag) {	
depth_rep_info_element(ZNearSign, ZNearExp, ZNearMantissa, ZNearManLen)	
depth_rep_info_element(ZFarSign, ZFarExp, ZFarMantissa, ZFarManLen)	
} else	
for(i = 0; i <= mpii_num_layer_minus1; i++)	
depth_rep_info_element(ZSign[i], ZExp[i], ZMantissa[i], ZManLen[i])	
mpii_texture_opacity_interleave flag	u(1)
if(mpii_texture_opacity_interleave flag == 0)	
mpii_texture_opacity_arrangement_flag /* 0: TaB, 1:SbS */	u(1)
mpii_picture_num_layers_in_height_minus1	ue(v)
}	

MPI compression & transmission with SEI

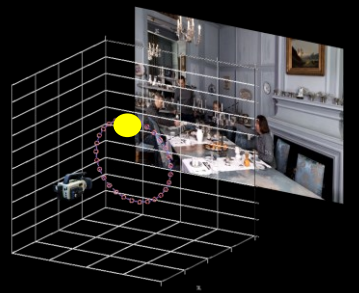
MPI rendering (at ref camera position)

MPI rendering (along a virtual camera position path)

Novel view rendering at different compression quality

uncompressed

High BR: 8.6 Mbps

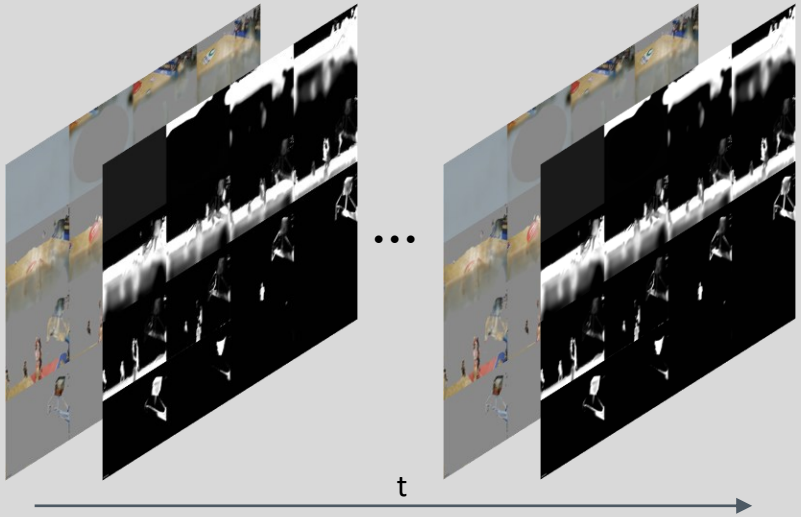


Mid BR: 3.1 Mbps

Low BR: 1.6 Mbps

Example: CBAbasketball

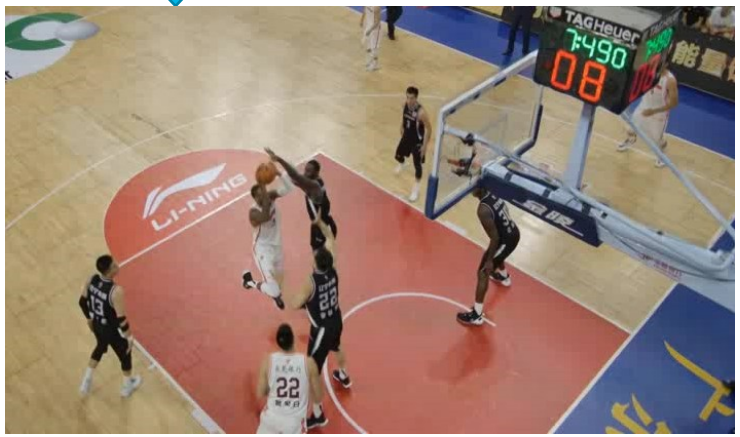
source video to MPI representation and SEI



multiplane_image_information(payloadSize) {	Descriptor
mpii_view_id	ue(v)
mpii_num_layers_minus1	ue(v)
mpii_layer_depth_equal_distance_flag	u(1)
if(mpii_layer_depth_equal_distance_flag) {	
depth_rep_info_element(ZNearSign, ZNearExp, ZNearMantissa, ZNearManLen)	
depth_rep_info_element(ZFarSign, ZFarExp, ZFarMantissa, ZFarManLen)	
} else	
for(i = 0; i <= mpii_num_layer_minus1; i++)	
depth_rep_info_element(ZSign[i], ZExp[i], ZMantissa[i], ZManLen[i])	
mpii_texture_opacity_interleave flag	u(1)
if(mpii_texture_opacity_interleave flag == 0)	
mpii_texture_opacity_arrangement_flag /* 0: TaB, 1:SbS */	u(1)
mpii_picture_num_layers_in_height_minus1	ue(v)
}	

MPI compression & transmission with SEI

MPI rendering (at ref camera position)

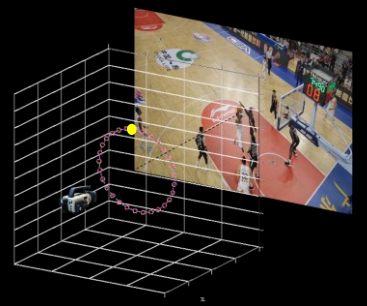


MPI rendering (along a virtual camera position path)

Novel view rendering at different compression quality

uncompressed

High BR: 8.4 Mbps



Mid BR: 2.5 Mbps

Low BR: 1.2 Mbps

System level considerations

Q & A

Dolby