Status of the Immersive Media Standard:

Related to Test Model for Immersive Video (TMIV) for 3DoF+

Presenter: Jong-Beom Jeong (uof4949@skku.edu)

Jong-Beom Jeong, SoonBin Lee, Il-Woong Ryu, Dongmin Jang, Eun-Seok Ryu

Multimedia Computing Systems Lab. (MCSL)

http://mcsl.skku.edu

Department of Computer Education

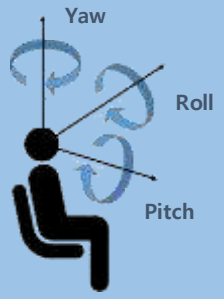
Sungkyunkwan University

Immersive Media Standard

IEEE 3079

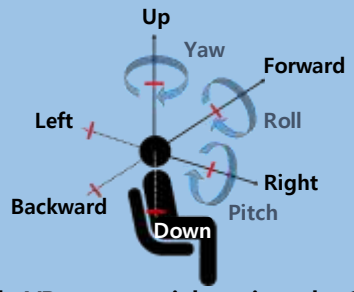
- MPEG-I is responsible for standardizing immersive media in MPEG and specifies the goals of step 3.
- Goal of Revitalizing VR Commercial Service by 2020
- Goal of 6DoF media support by 2022 after completing 3DoF standard by 2017

Step 1



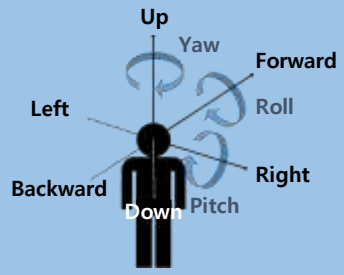
- Complete 3DoF standard by 2017
- Rotate head in a fixed state
- 360 video full streaming by default
- Tiled streaming if possible

Step 2



- Enable VR commercial servicers by 2020
- Allow head rotation and movement within a restricted area
- User-to-user conversations and projection optimization

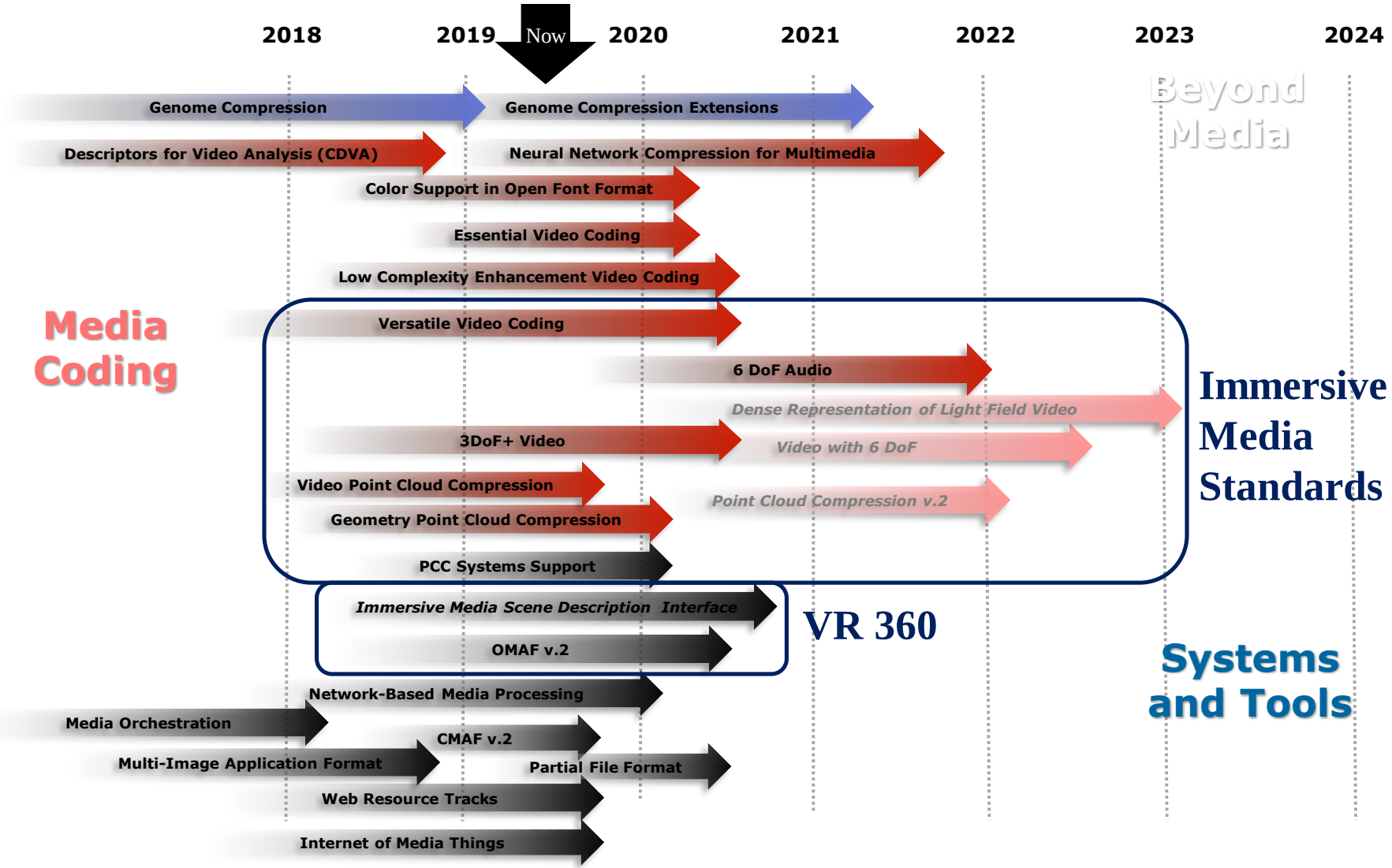
Step 3



- Support 6DoF by 2022
- 6DoF video will reflect user's walking motion
- Support interaction with virtual environments

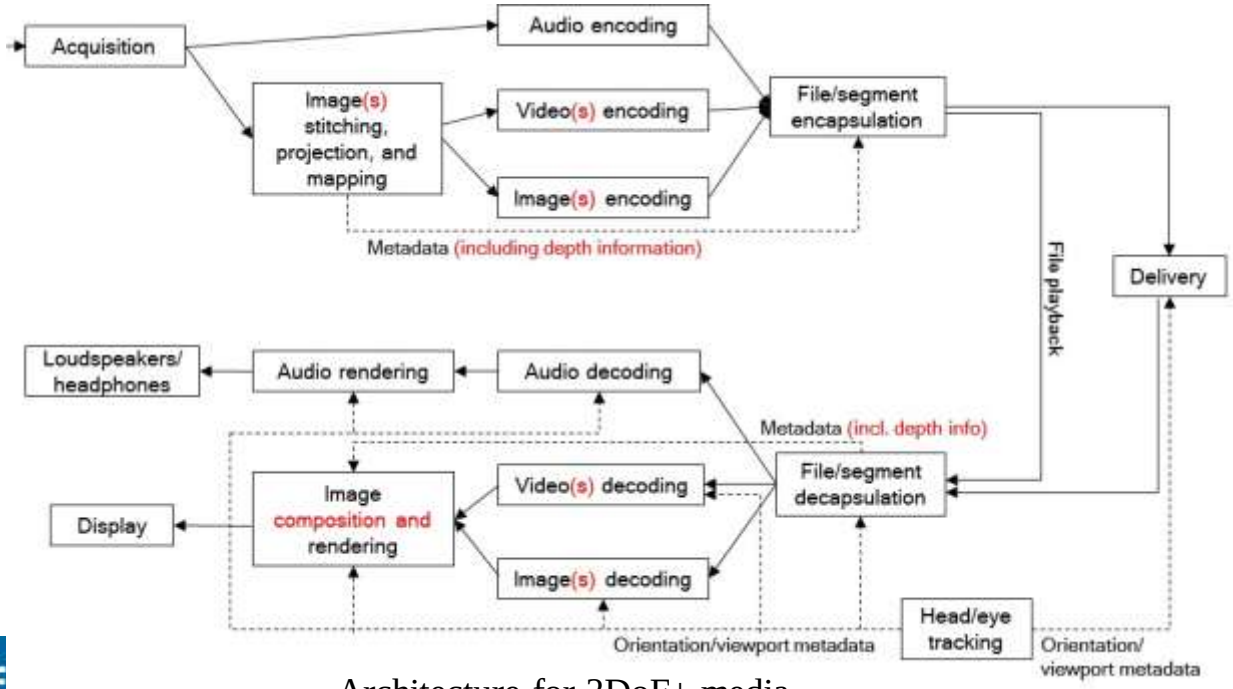
Immersive Media Standard Roadmap

IEEE 3079

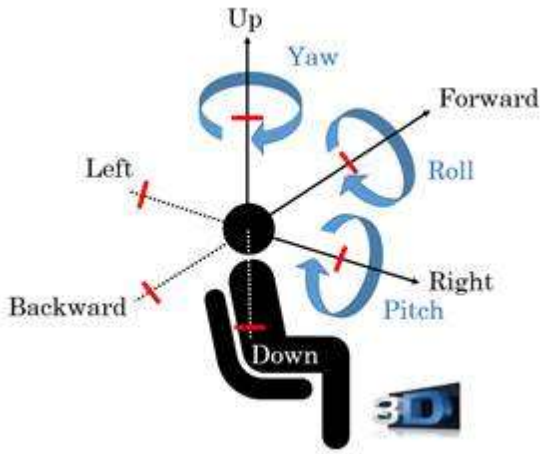


3DoF+ CfP (w18097)

- Background
  - MPEG defined degree of freedom of VR as 3DoF, 3DoF+, and 6DoF
  - Limited movements for user sitting in a chair is available for 3DoF+
- Requirements
  - Solution will be built on HEVC with 3DoF+ metadata(included in MPEG-I part 7)
  - Both objective and subjective quality evaluation will be performed



Architecture for 3DoF+ media



Use case for 3DoF+

3DoF+ CfP (w18097) (-Cont'd)

- Test Sequences

ClassroomVideo

TechnicolorMuseum

TechnicolorHijack

TechnicolorPainter

IntelKermit

Sequence	Class	Resolution	No. of views	Frame count	Frame rate	Source FoV
ClassroomVideo	A	4096x2048	15	120	30	360º x 180º
TechnicolorMuseum	B	2048x2048	24	300	30	180º x 180º
TechnicolorHijack	C	4096x4096	10	300	30	180º x 180º
TechnicolorPainter	D	2048x1088	16	300	30	46º x 25º
IntelKermit	E	1920x1080	13	300	30	77.8º x 77.8º

3DoF+ CfP (w18097) (-Cont'd)

IEEE 3079

- Software

Software name	Location	Tag/branch
RVS	http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/RVS	v3.1
WS-PSNR	http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/WS-PSNR	v2.0
HDRTools	https://gitlab.com/standards/HDRTools	v0.18
360Lib	https://jvet.hhi.fraunhofer.de/svn/svn_360Lib	360Lib-5.1-dev
HM	https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware	HM-16.16

- Target Bitrate

	Target bitrates [Mbit/s]					
Sequence	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5	Rate 6
ClassroomVideo	6.5	10	15	25	40	65
TechnicolorMuseum	10	15	25	40	65	100
TechnicolorHijack	6.5	10	15	25	40	65
TechnicolorPainter	6.5	10	15	25	40	65
IntelKermit	4	6.5	10	15	25	40

Response on 3DoF+ CfP

IEEE 3079

For response of 3DoF+ CfP, 5 input documents are proposed

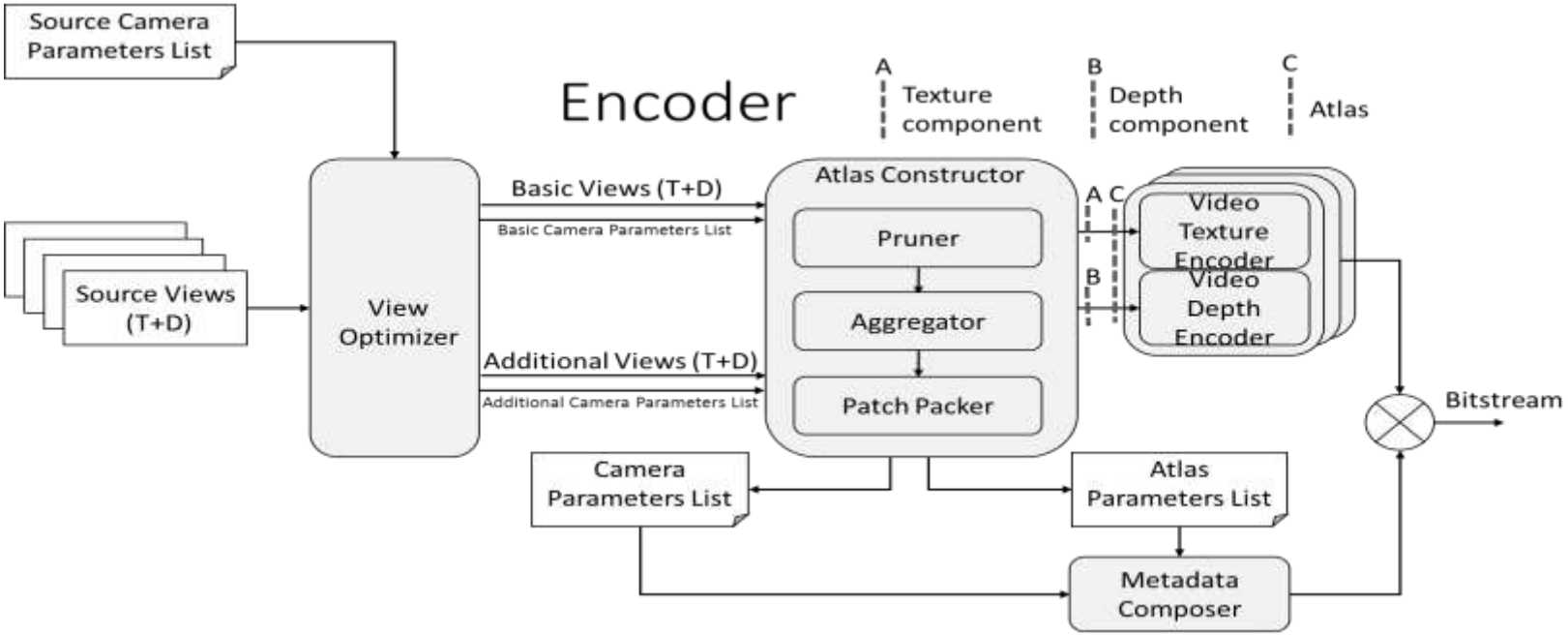
All of proposals introduce pruning and packing architecture

These technologies are included in test model for 3DoF+

Number	Title
m47179	Philips response to 3DoF+ Visual CfP
m42372	Description of Nokia's response to CFP for 3DOF+ visual
m47407	Technical description of proposal for Call for Proposals on 3DoF+ Visual prepared by Poznan University of Technology (PUT) and Electronics and Telecommunications Research Institute (ETRI)
m47445	Technicolor-Intel Response to 3DoF+ CfP
m47684	Description of Zhejiang University's response to 3DoF+ Visual CfP

Test Model for Immersive Video (TMIV)

- Encoder
  - TMIV encoder consists of three main steps:
  - (1) View optimizer
  - (2) Atlas constructor
  - (3) Video encoder & Metadata composer

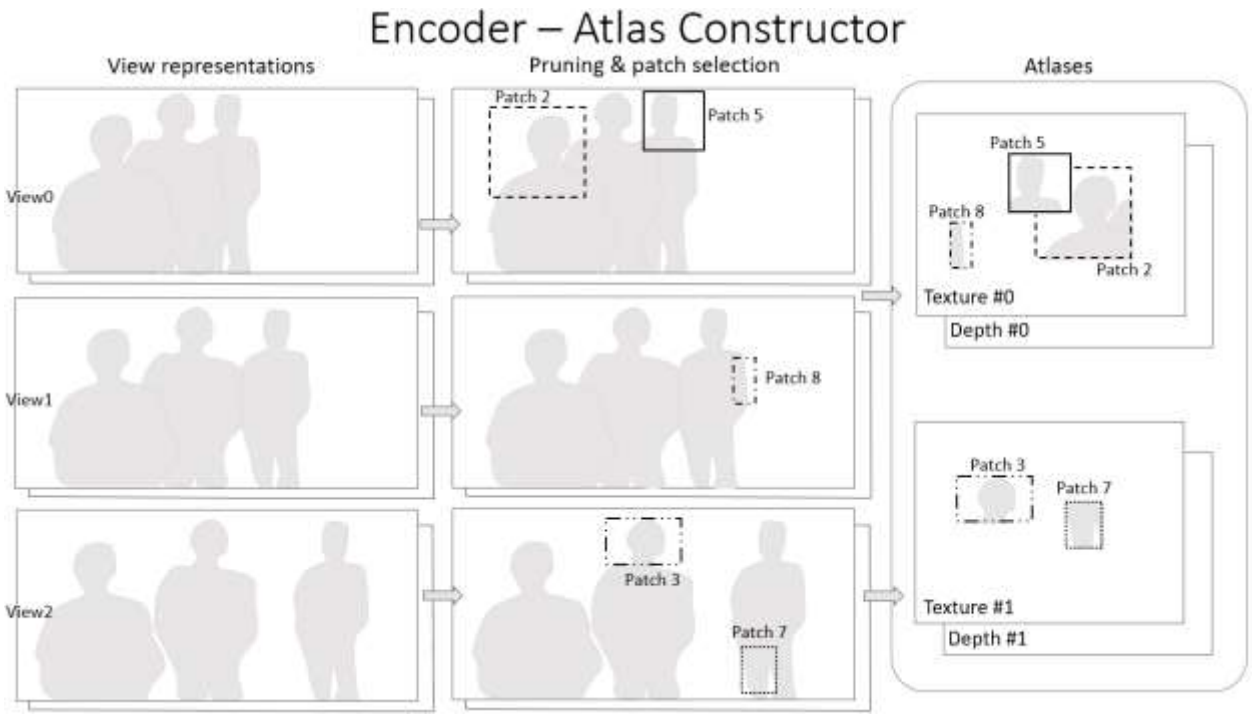


Architecture for TMIV encoder[1]

[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Test Model for Immersive Video (TMIV) (-Cont'd)

- Encoder
  - Atlas constructor references the occupancy map of each views
  - Minimizes the atlas size by overlapping the patches

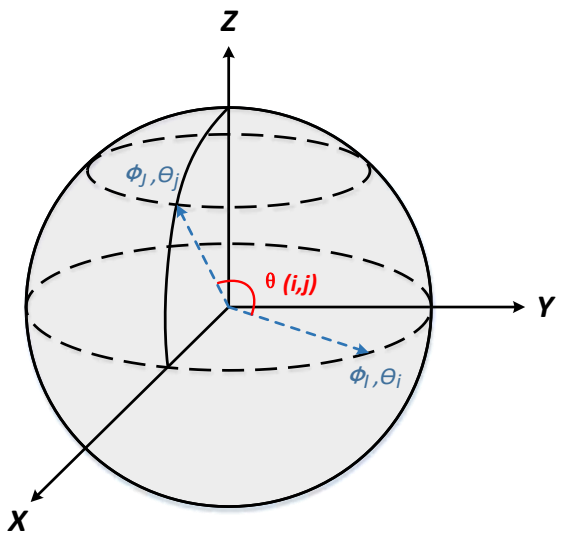


Example of TMIV encoder process[1]

[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Test Model for Immersive Video (TMIV) (-Cont'd)

- View Optimizer
  - Determines the basic views and the additional views
  - Select two views (=one pair) which have the largest direction deviation
  - They have the maximum value of θ(i, j) where i, j are the indices the source views
  - If overlap $\geq 0.5 * \min(FOV_i, FOV_j)$, only one basic view is selected
  - Else, multiple basic views including view m and view n are selected



Example of the directions deviations[1]

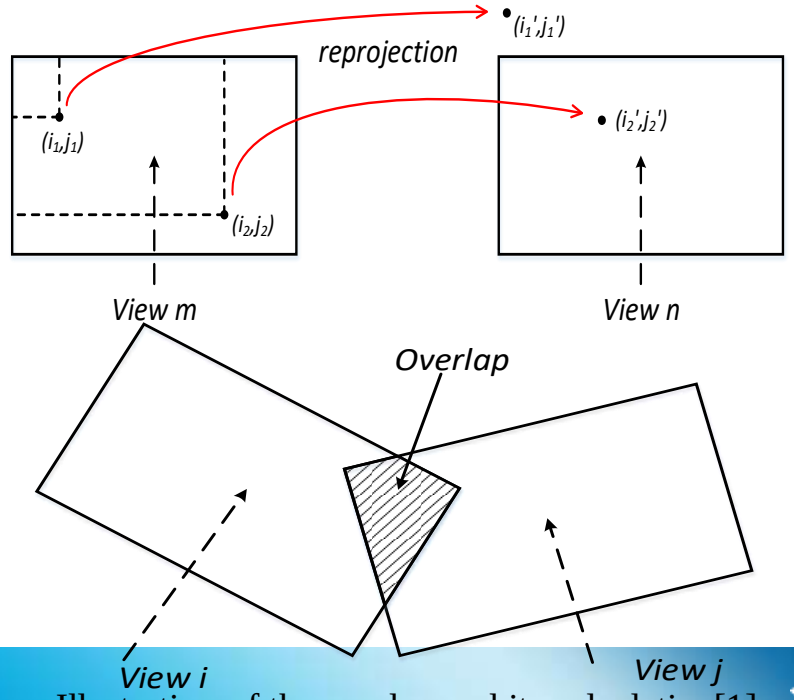
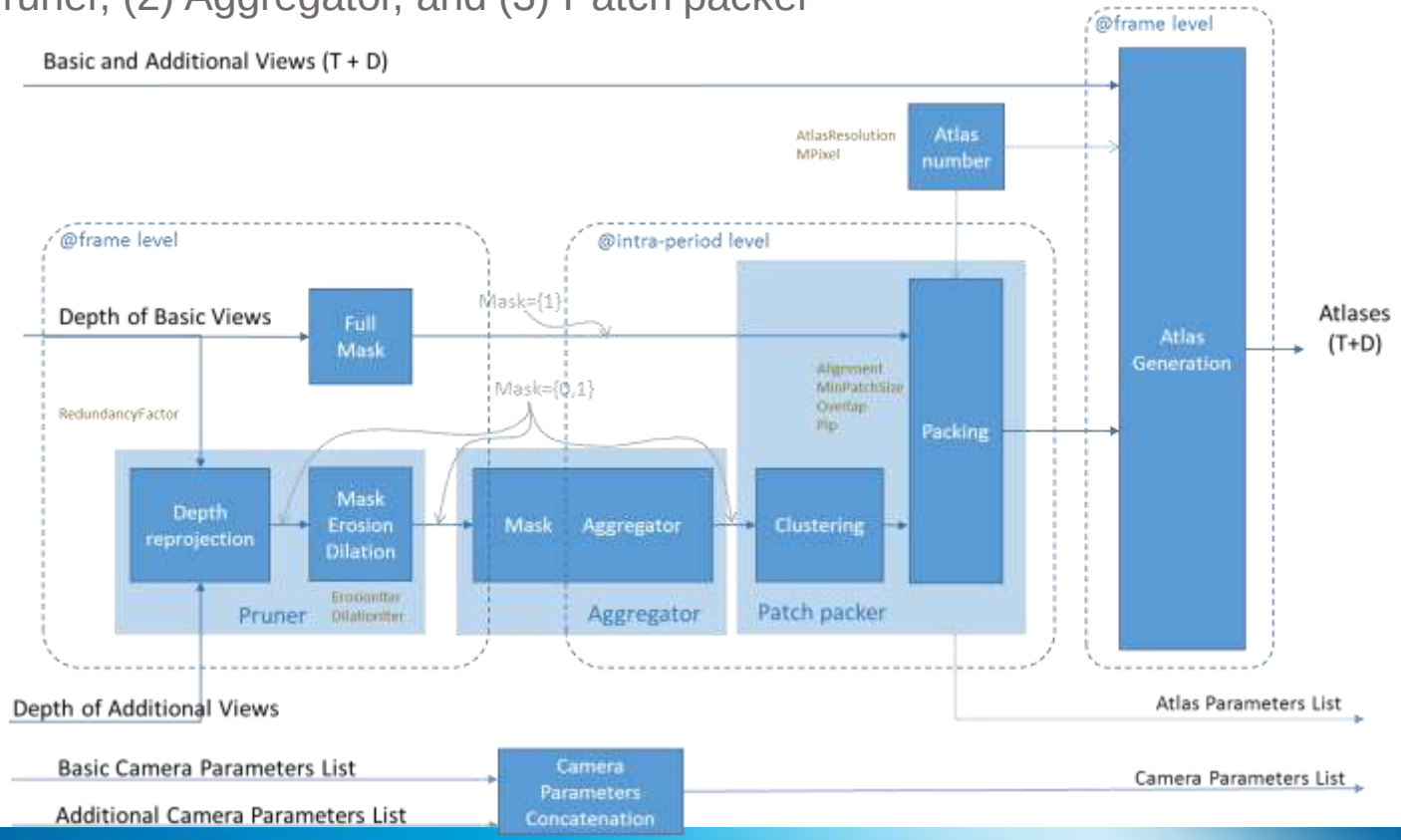


Illustration of the overlap and its calculation[1]

[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Test Model for Immersive Video (TMIV) (-Cont'd)

- Atlas Constructor
  - Atlas : aggregation of patches after packing process
  - The atlas constructor is composed of three parts:
  - (1) Pruner, (2) Aggregator, and (3) Patch packer



Block diagram of the Atlas Constructor[1]

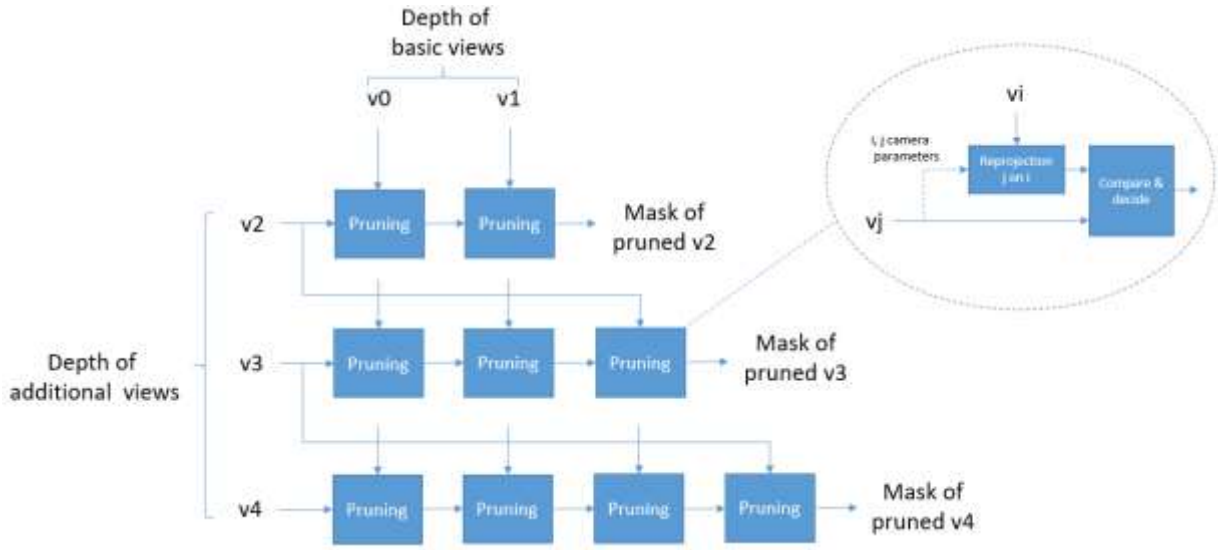
[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Test Model for Immersive Video (TMIV) (-Cont'd)

- Pruner
  - De-project depth from view j to 3D space and project them to the reference view I
  - A sample is pruned if it is already covered by the previous view, determined by :

    $|z - z_p| < RedundancyFactor * \min(z, z_p)$

  - Default value of RedundancyFactor is 0.02

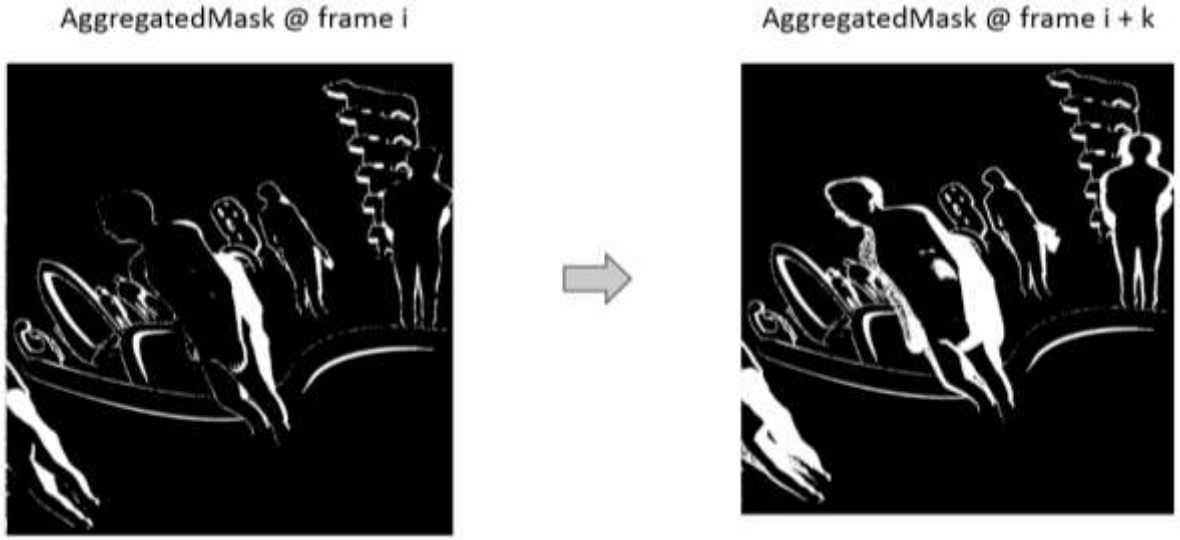


Example of pruning with 2 basic views and 3 additional views[1]

[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Test Model for Immersive Video (TMIV) (-Cont'd)

- Aggregator
  - Accumulation for each intra period
  - Implemented by logical operation OR



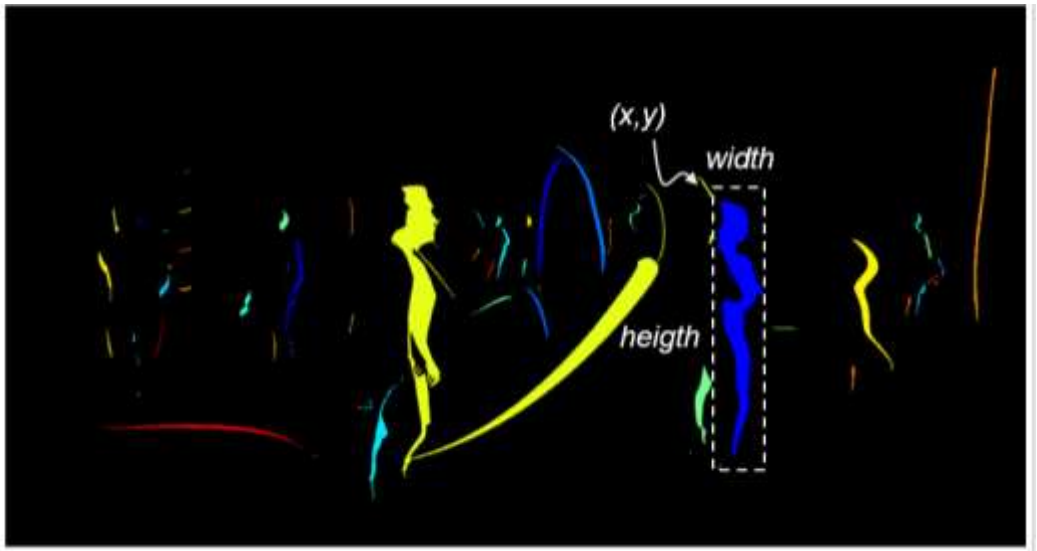
[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Test Model for Immersive Video (TMIV) (-Cont'd)

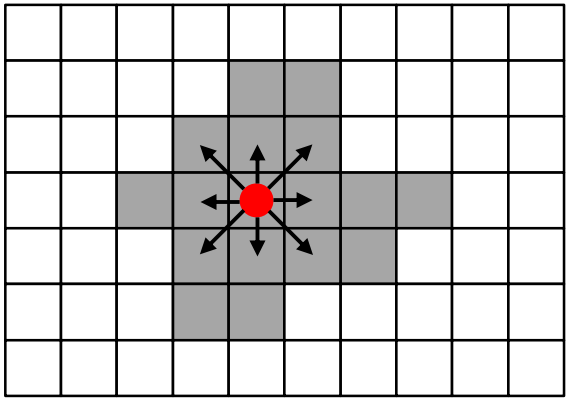
- Patch Packer
  - 8-pixel neighborhood region growing is conducted to generate a cluster
  - Then, the clusters are sorted by a decreasing size order
  - The number of atlases is determined by :

$$N_{Atlas} = ceil(\frac{MPixel * 1024^2}{AtlasWidth * AtlasHeight})$$

where maximum size of all atlases is expressed in Mpixels (default value : 20)



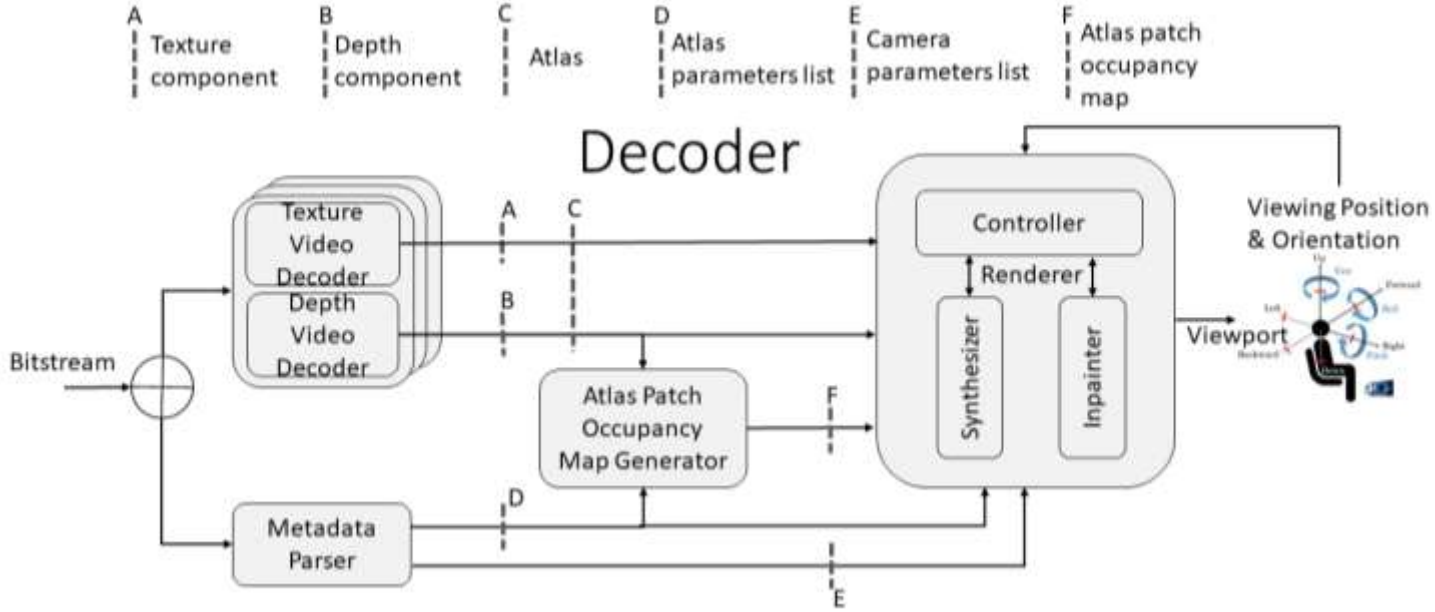
Example of clusters represented in false color on a pruned view



Example of 8-pixel neighborhood region growing

Test Model for Immersive Video (TMIV) (-Cont'd)

- Decoder
  - Decoder receives HEVC decoded atlases and the following metadata
  - TMIV decoder consists of three main steps:
  - (1) Video decoder and metadata parser
  - (2) Atlas patch occupancy map generator
  - (3) Renderer

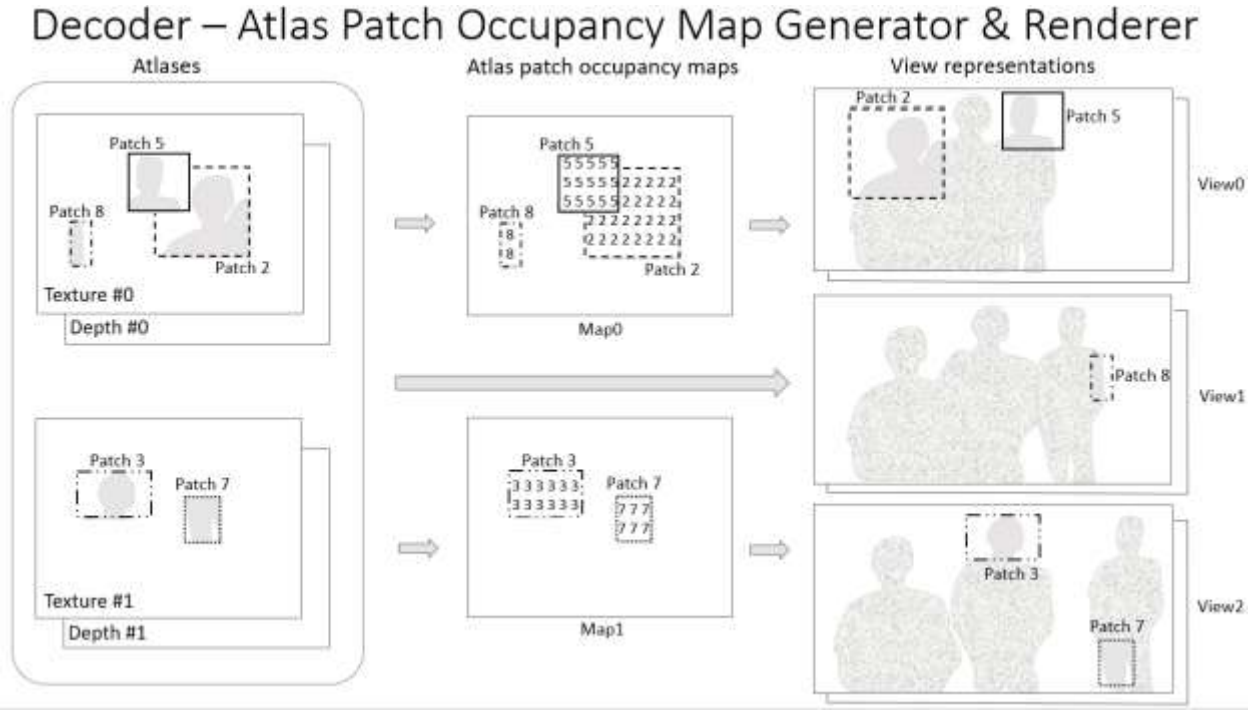


Architecture for TMIV decoder[1]

[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Test Model for Immersive Video (TMIV) (-Cont'd)

- Atlas Patch Occupancy Map Generator
  - An occupancy map is generated for each atlas
  - Occupancy map contains the locations of the patches
  - Based on the occupancy map, the source views are reconstructed

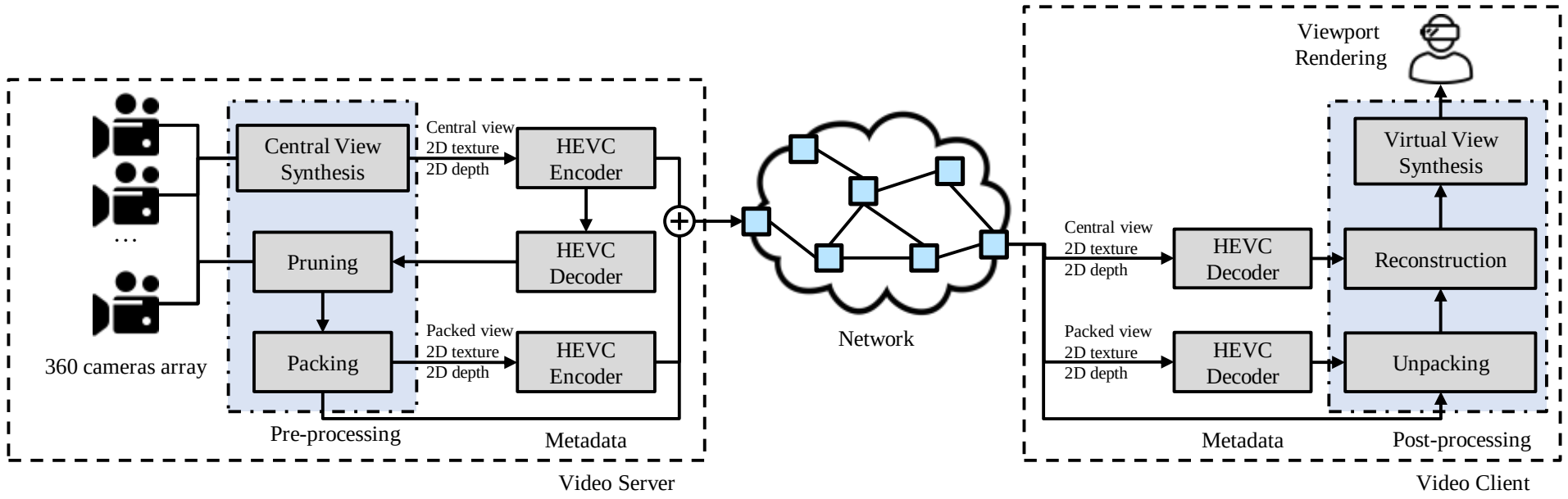


Data flow of TMIV decoder[1]

[1] Basel Salahieh, et al. Test Model for Immersive Video. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, w18470.

Experimental Results

- HEVC anchor, TMIV anchor, and MCSL's method are evaluated
- MCSL's method represents TMIV-like encoder and decoder
- BD-rate, required number of decoders, encoding time are measured



Block diagram for MCSL 3DoF+ S/W platform[1]

[1] Jong-Beom Jeong, et al. Towards 3DoF+ 360 Video Streaming System for Immersive Media. IEEE Access, vol. 7, pp. 136399-136408, 2019.

Experimental Results (-Cont'd)

- Server
  - 2 linux servers (ubuntu 16.04) were used in experiment
  - One has 2 intel xeon E5-2687w (12 cores, 24 threads) CPUs and 128GB memory
  - Another has 2 intel xeon E5-2620 (6 cores, 12 threads) CPUs and 128GB memory
- Software
  - Versions of softwares meet the requirements of CfP
  - OpenCV 3.4.2 was used in source view pruning, reconstruction, and RVS
  - OpenGL was not used because of the building issue in linux

Experimental Results (-Cont'd)

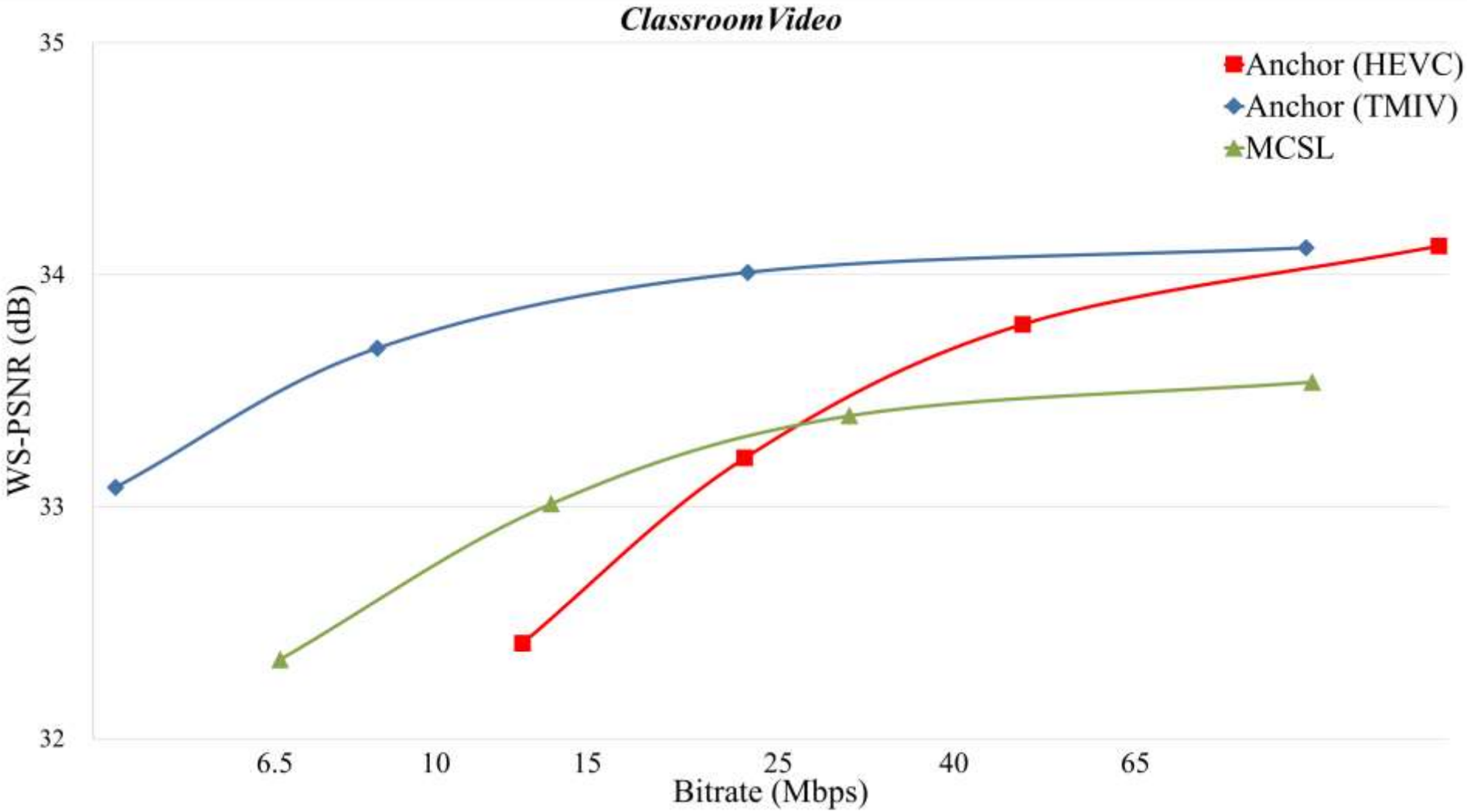
- Experimental results for *ClassroomVideo* and *TechnicolorMuseum*
  - For the objective evaluation, WS-PSNR[1] is used
  - TMIV showed the best results on BD-rate and encoding time saving
  - MCSL's method requires the least number of decoders among the introduced methods

Sequence	Method	BD-rate	No.of decoders required	Encoding time saving
ClassroomVideo	HEVC-anchor	0.00%	30	0.00%
	TMIV-anchor	-80.59%	4	-86.28%
	MCSL	-37.14%	4	-81.68%
TechnicolorMuesum	HEVC-anchor	0.00%	48	0.00%
	TMIV-anchor	-68.26%	8	-86.72%
	MCSL	-18.90%	4	-95.21%

Results of HEVC anchor, TMIV anchor, and MCSL's method

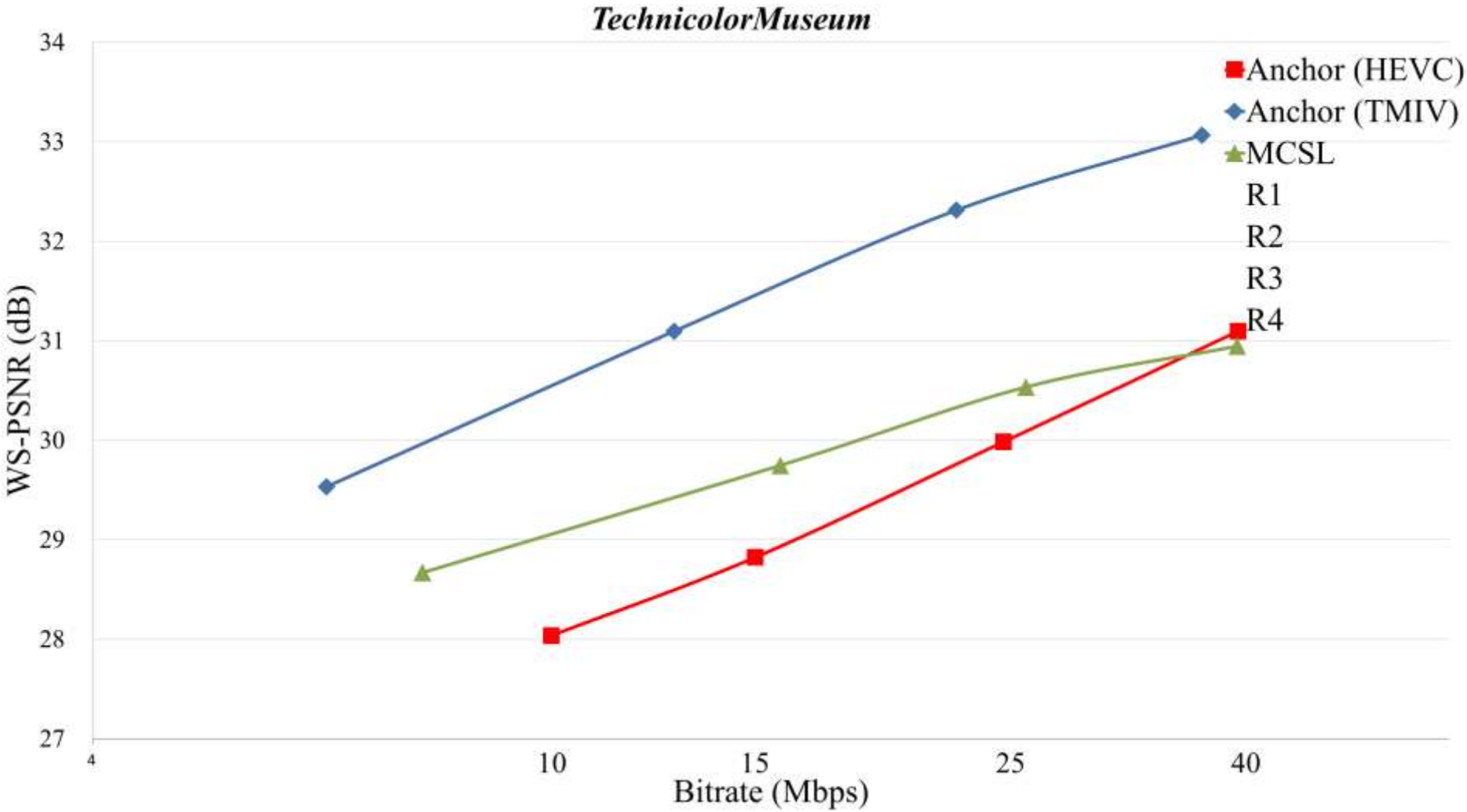
[1] Yule Sun, et al. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. IEEE signal processing letters, 24(9), 1408-1412. 2018.

Experimental Results (-Cont'd)



RD-curve of *ClassroomVideo*

Experimental Results (-Cont'd)



RD-curve of *TechnicolorMuseum*

Appendix : Atlas generated by TMIV

- Class A1(ClassroomVideo), atlas 0 (basic view), texture

Resolution = 4096x2048

Appendix : Atlas generated by TMIV

- Class A1(ClassroomVideo), atlas 0 (basic view), depth

Resolution = 4096x2048

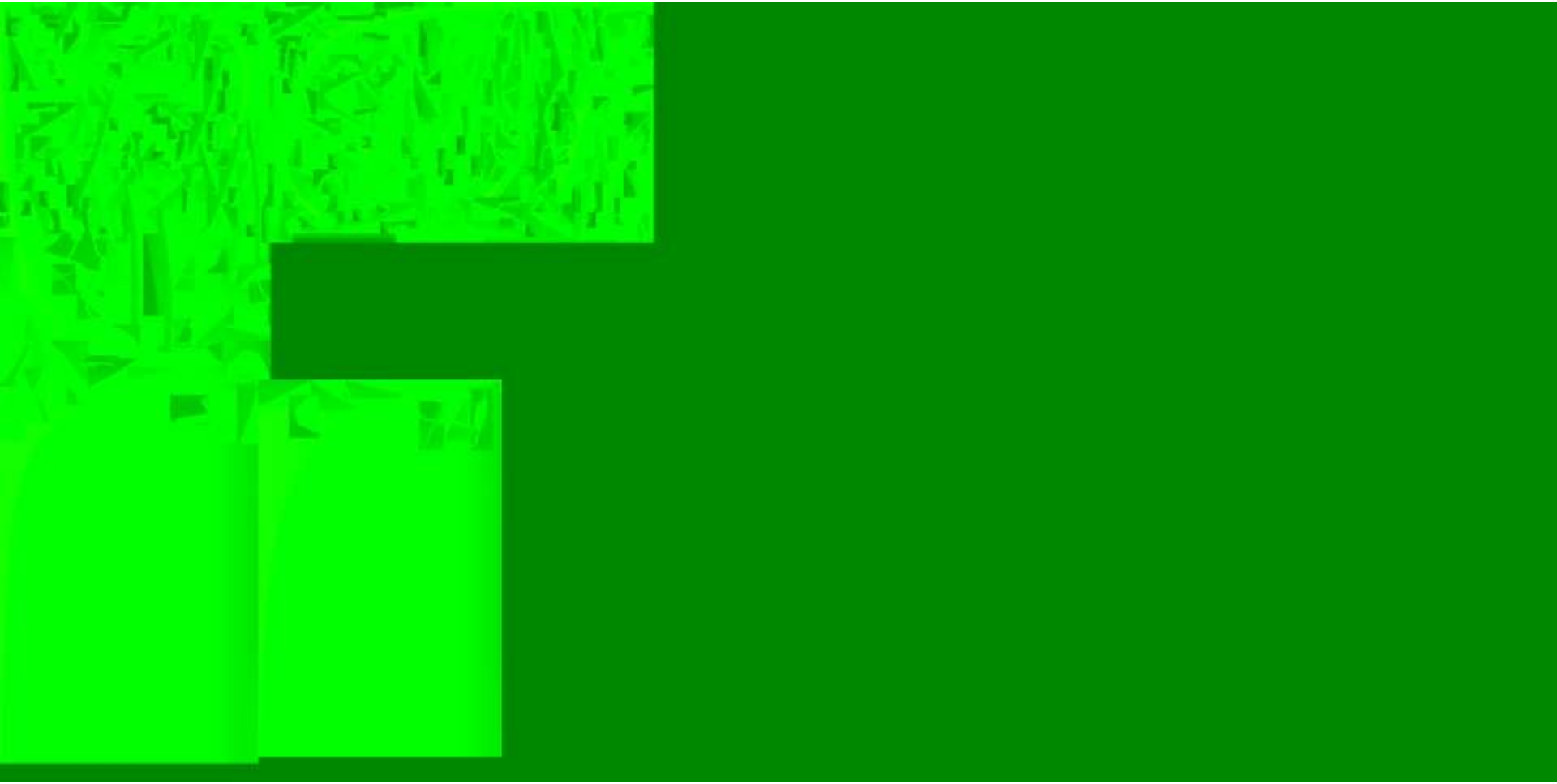
Appendix : Atlas generated by TMIV

- Class A1(ClassroomVideo), atlas 1 (additional view), texture

Resolution = 4096x2048

Appendix : Atlas generated by TMIV

- Class A1(ClassroomVideo), atlas 1 (additional view), depth



Resolution = 4096x2048

Appendix : Packed View generated by MCSL

- Class A1(ClassroomVideo), texture

Resolution = 4096x4480