Viewport-based 360 Video Streaming:

Motion-Constrained Tile and Viewport Extraction

Presenter: Eun-Seok Ryu (esryu@gachon.ac.kr)

Dongmin Jang, Jong-Beom Jeong, Eun-Seok Ryu

Multimedia Communications and Systems Lab. (MCSL)

http://mcsl.gachon.ac.kr

Department of Computer Engineering

Gachon University

Introduction

❖High Bandwidth Requirement of VR

- Recently, various HMD devices are on the market
- Recommend 12K resolution for reducing nausea with high quality VR
- High Bandwidth and high computational complexity are huddle

↳ Need to reduce the required bandwidth!

The emergence of various HMD
(Gear VR, Oculus Rift, Daydream, PlayStation VR)

Requirement	details
pixels/degree	40 pix/deg
video resolution	11520x6480
framerate	90 fps

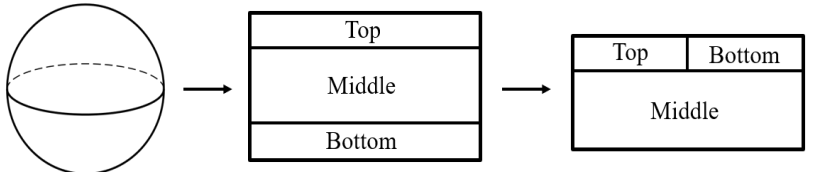
Requirements for high quality VR
Source:Technicolor, Oct. 2016 (m39532, MPEG 116th Meeting)

Viewport Independent vs Viewport Dependent

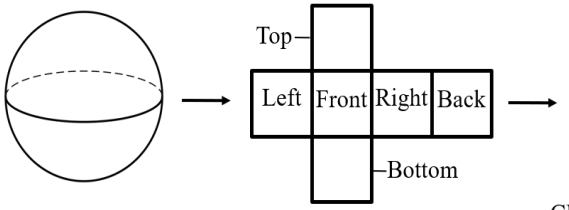
Viewport Independent

- ❖Transmit whole picture
- ❖Projection and packing
- ❖Downsampling / adjusting QP

ERP – Adaptive QP



ERP – Region Wise Packing

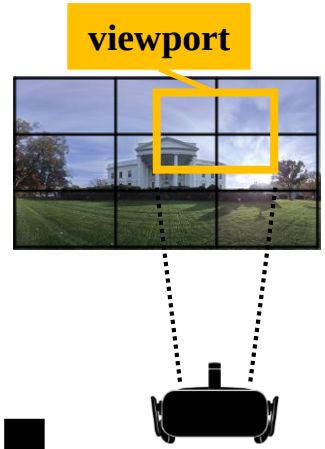


CMP – Region Wise Packing

Viewport Dependent (Proposed Method)

- ❖Transmit viewport only
- ❖Bitrate saving
- ❖ecoding complexity reducing
- ❖But, delay

Select a tile set corresponding to the viewport

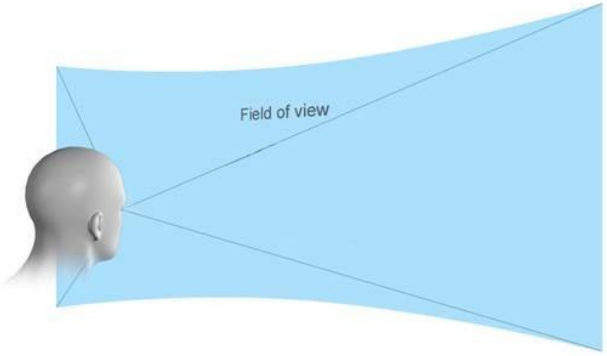


Keypoint for Viewport Dependent Streaming

IEEE 3079

- Field of View(FOV)
  - The field of view (FOV) in the HMDs : 96˚ to 110˚
  - Part of the 360˚ picture

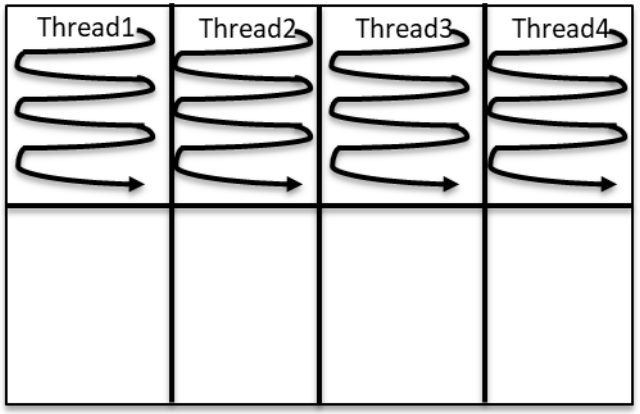
  - The user's current viewport : high resolution
  - Remaining part : low resolution



Field of View (FOV)

- Tiles
  - Parallelization Tools
  - Divide into rectangular regions
  - Flexible horizontal and vertical boundaries

  - Spatially refers to only its own tile, but temporally refers to other tiles
  - Decoding problems occurs when transmitting only some tiles



A frame divided into 8 Tiles

Proposed Motion Constraint Tile Sets

Reference Picture

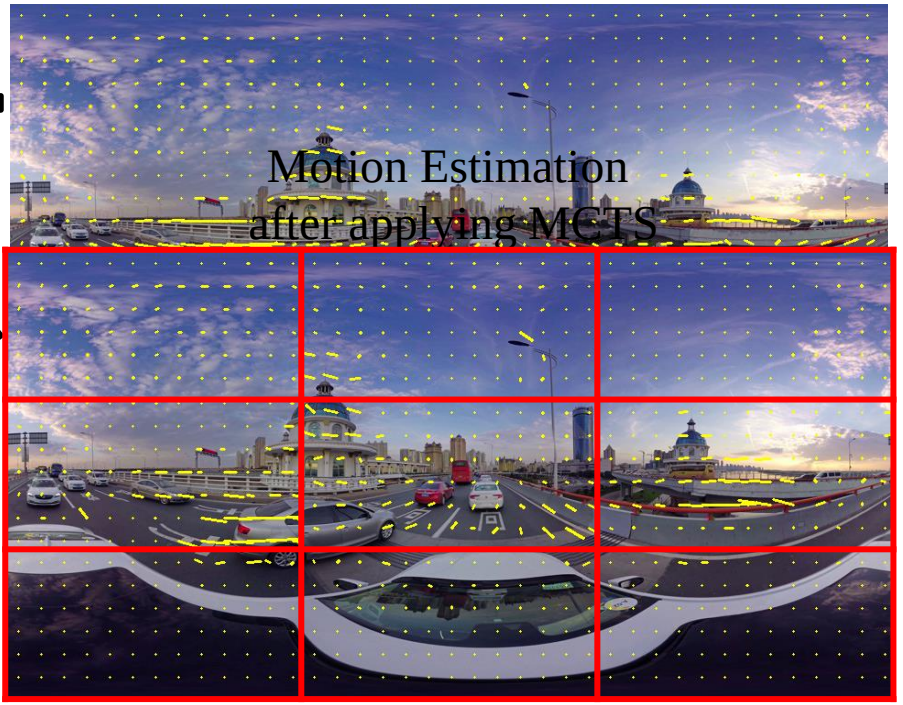
Reference

Current Picture

Modify motion vector range

Motion vectors can refer to anyware

Original Motion Estimation before applying MCTS

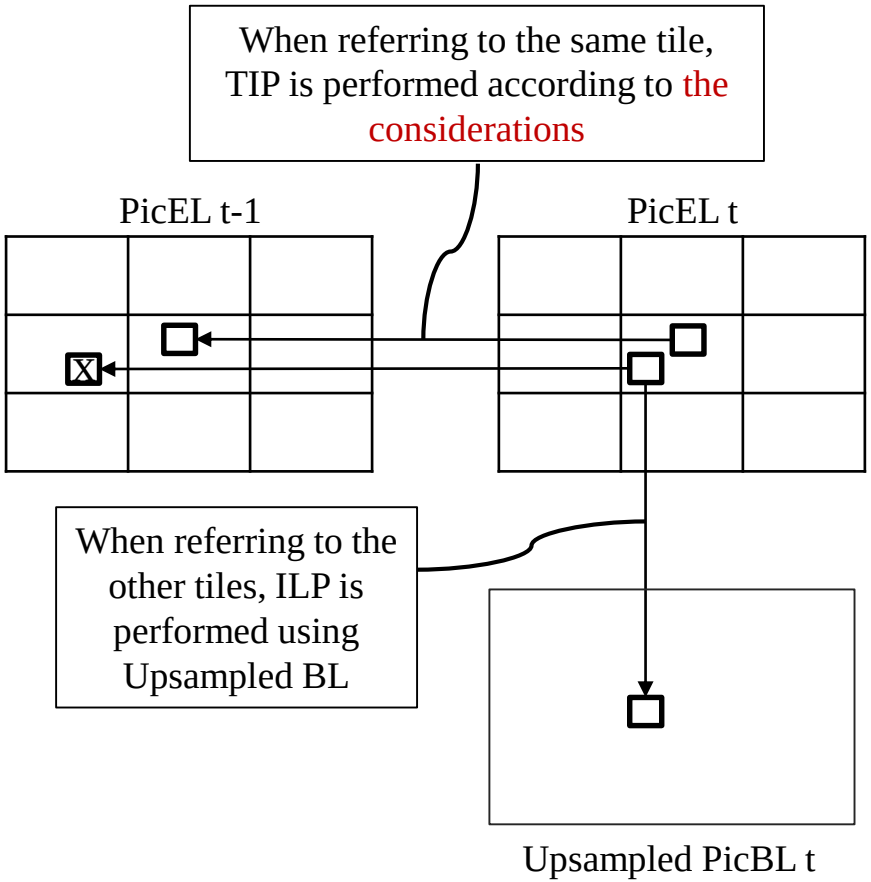


Motion vectors only refer to same position Tile

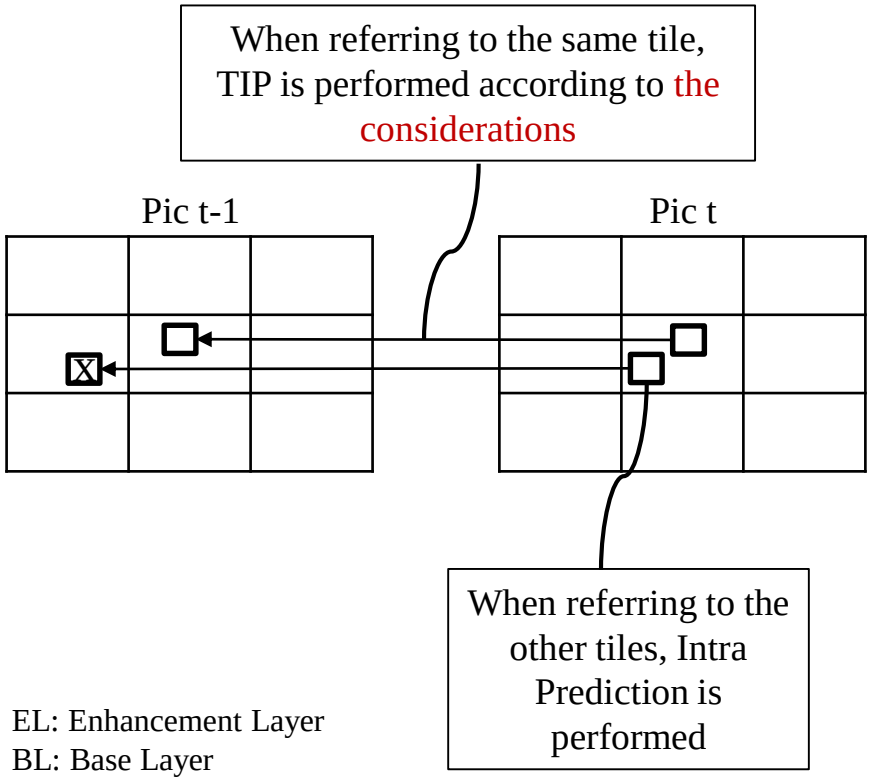
Adopt our MCTS in MPEG (HM ver. 16.18)

MCTS with HEVC and SHVC - Structure

SHM Encoder



HM Encoder



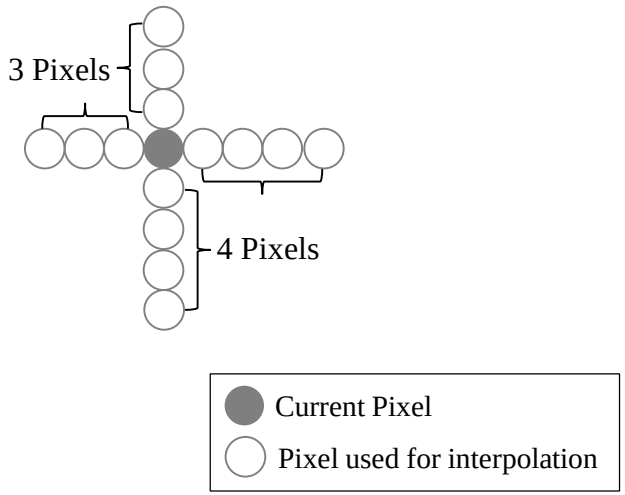
- EL: Enhancement Layer
- BL: Base Layer
- TIP: Temporal Inter Prediction
- ILP: Inter Layer Prediction

Considerations : Interpolation, Temporal Candidate of AMVP and MERGE

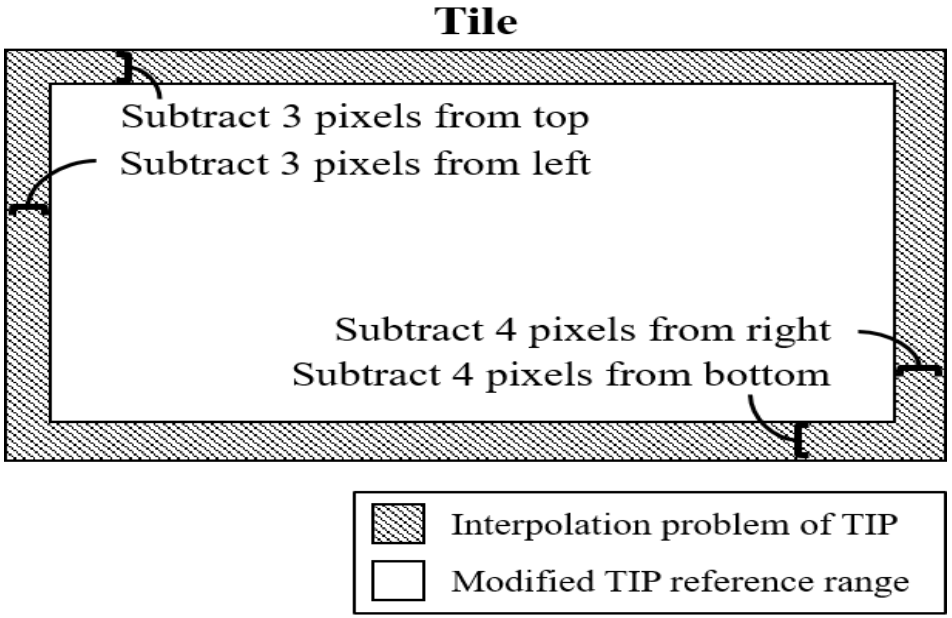
MCTS Considerations (1/2)

❖Interpolation

- Use an eight-tap filter to interpolate luma prediction
- Use 3 pixels of left and top, 4 pixels of right and bottom for Interpolation
  - → Modify reference range of motion vectors



The current pixel and
the pixels used for interpolation



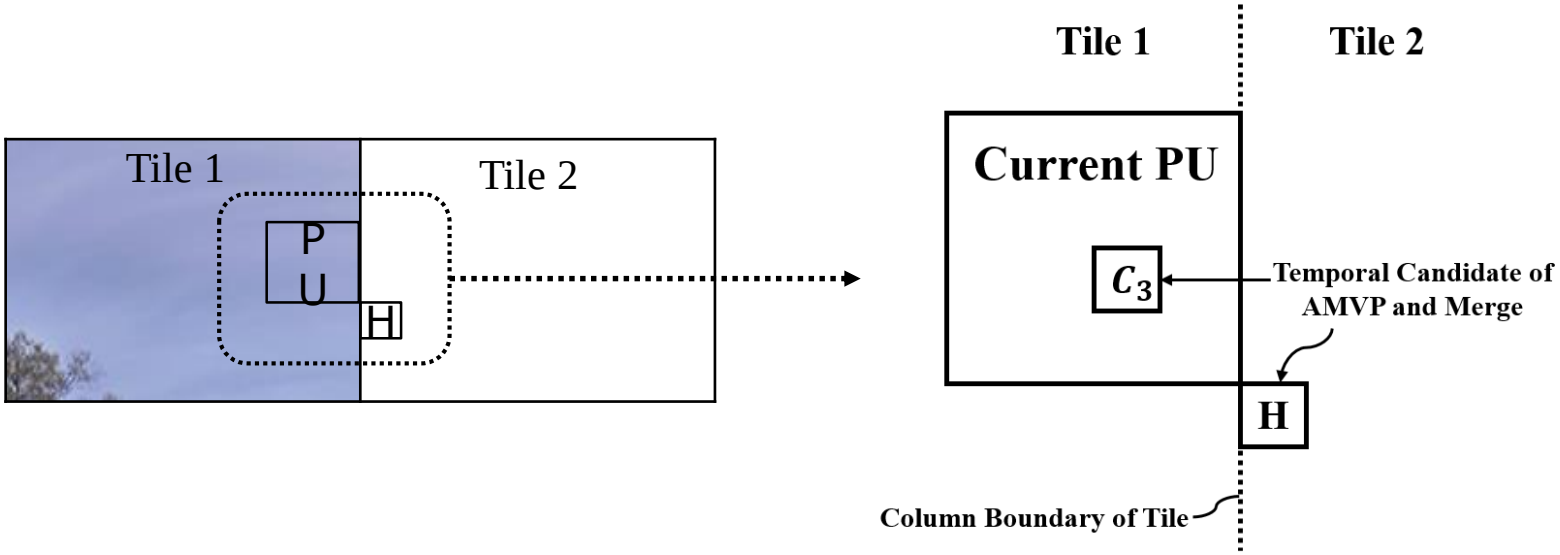
Interpolation problem of referring to a tile
at the same position in TIP

MCTS Considerations (2/2)

❖Temporal Candidate of AMVP and MERGE

- Temporal candidates : C3 and H block (right figure)
- Problem: cross the column boundary between tiles

↳ Exclude H block at the column boundary between tiles



Temporal candidate problem at column boundary between Tiles

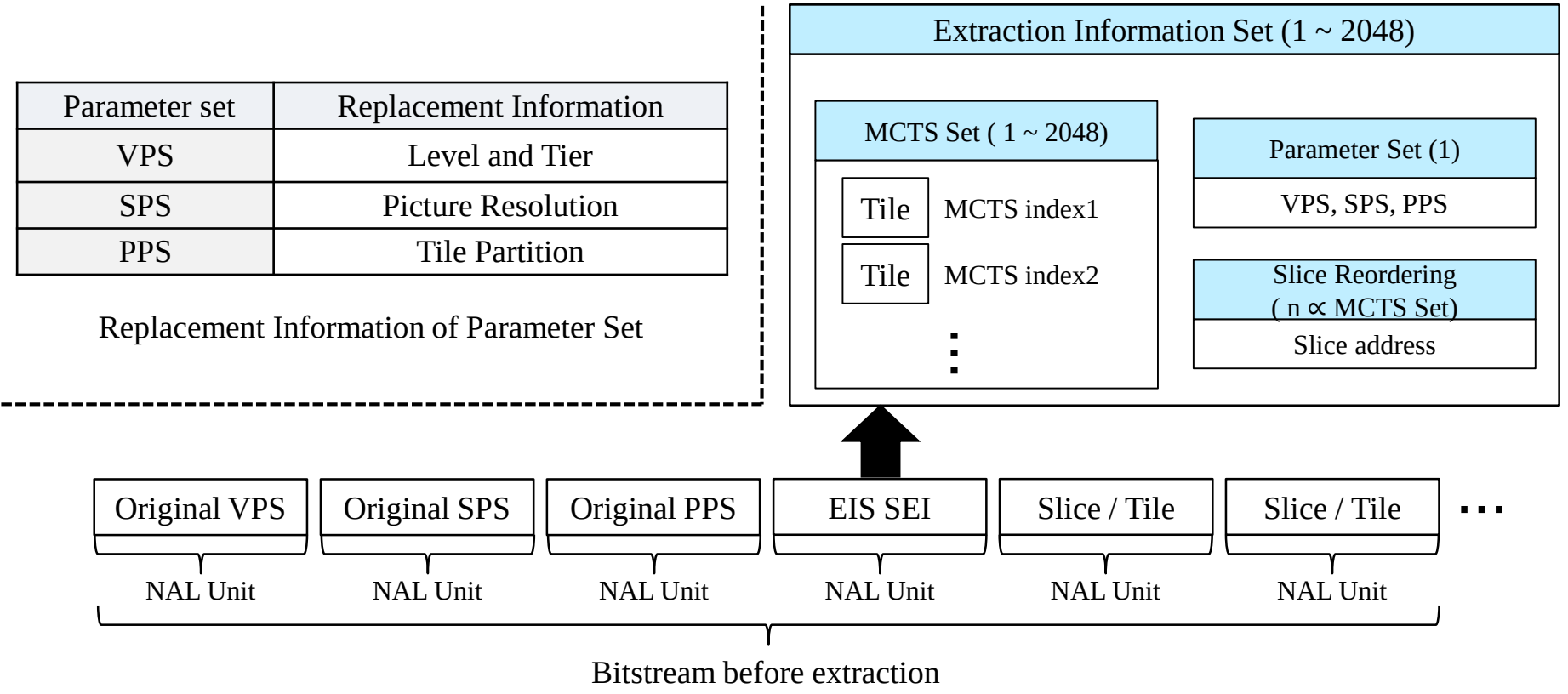
Extraction Information Sets (EIS) SEI Message

❖Extraction Information Sets (EIS) SEI Message

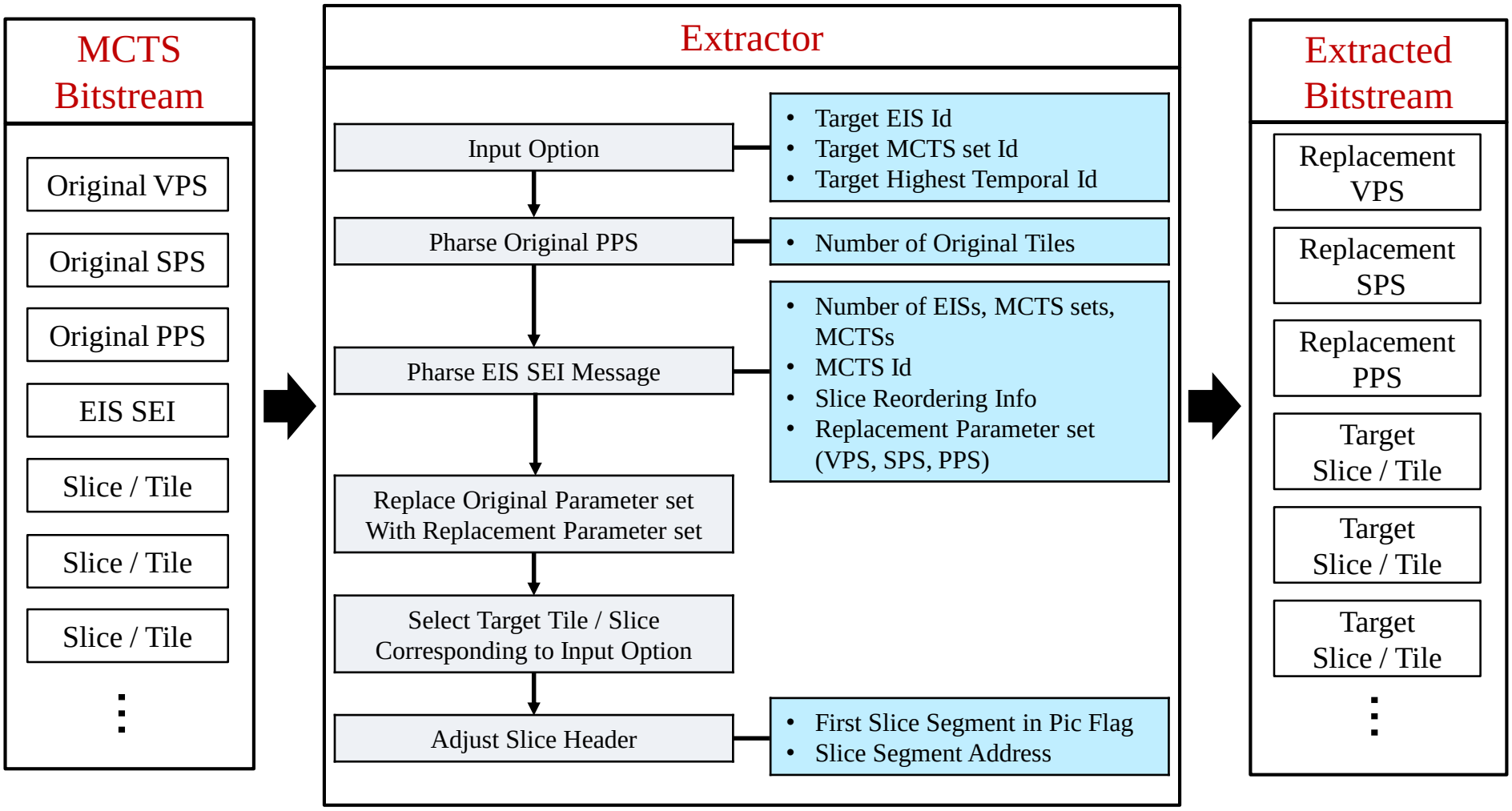
- Contains replacement parameter set (max: around 2000 MCTS sets)
- MCTS set contains a set of tiles to be extracted

Parameter set	Replacement Information
VPS	Level and Tier
SPS	Picture Resolution
PPS	Tile Partition

Replacement Information of Parameter Set



NAL Bitstream Extractor: Fucntional Flow



Implemented Renderer and Player

Experimental Results (1/2)

❖Experimental Setup

- 8K test sequences defined in JVET CTC (common test condition)
- Random Access (RA) coding structure
- Uniform 3x3, 9 Tiles
- HM ver. 16.16 encoder / SHM ver. 12.3 encoder

Name	Resolution	Frame Length	Frame Rate
KiteFlite	8192×4096	300	30 fps
Harbor	8192×4096	300	30 fps
Trolley	8192×4096	300	30 fps
GasLamp	8192×4096	300	30 fps

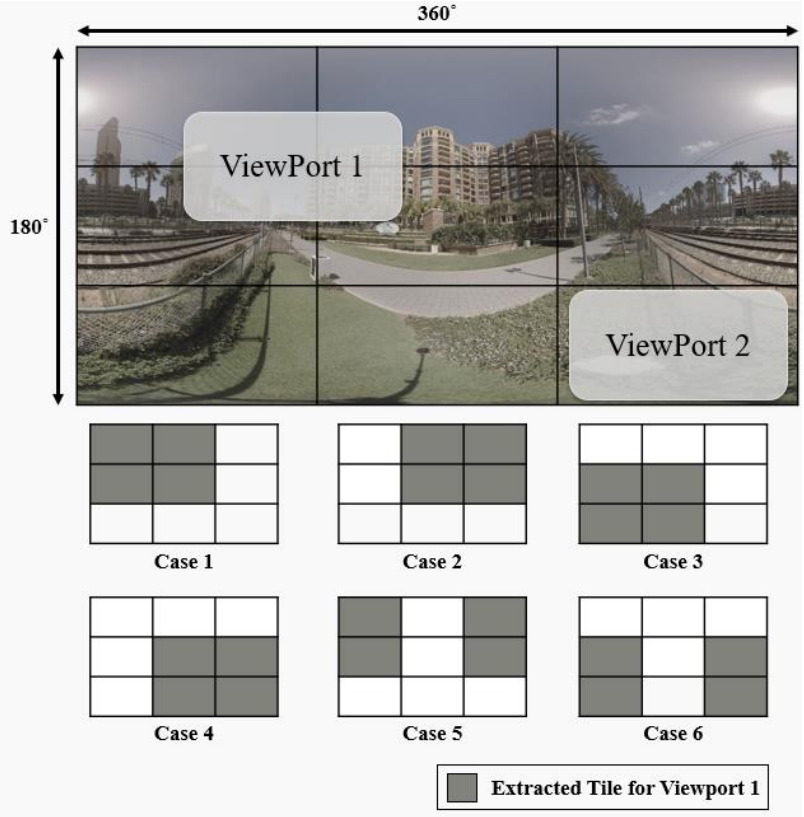
8K Test sequences

Coding Option	SHM Parameter	HM parameter
Version	12.3	16.16
CTU size	64×64	
Coding structure	RA	
QP	-	
Base Layer QP	22	
Enhancement Layer QP	22	
Tile	Uniformly 3x3 = 9 tiles	
Slice mode	Disable all slice options	
WPP mode	Disable all wpp options	

Coding options

Experimental Results (2/2)

- Bitrate savings in case of transmiting some Tiles only using proposed MCTS (among 9 Tiles)
- Various number of Tiles and eye-tracking with DL are under researching now; expecting NOSSDAV 2019. :)



	Proposed SHM		Proposed HM	
Name	4 tiles bitrate saving	1 tile bitrate saving	4 tiles bitrate saving	1 tile bitrate saving
KiteFlite	52%	88%	51%	87%
Harbor	53%	88%	51%	87%
Trolley	50%	87%	49%	87%
GasLamp	49%	87%	47%	86%
Average bitrate saving	51%	87%	49%	86%

Conclusion

- Motivation
  - High Quality VR >= 12k resolution
  - High BW, High computational complexity
  - Viewport tile streaming for 360 VR
- Proposed method
  - Motion Constrained Tile Sets (MCTS)
  - Extraction Information Set SEI Message (EIS SEI)
  - NAL Packet Extractor for Selected (ROI) Tiles
- Results
  - Transmit Selected Tiles without Decoding Errors
  - Save Bitrate, Reduce Computational Complexity at Decoder Side
- Future work
  - Eye tracking for Accurate Viewport Extraction
  - Deep Learning for ROI Estimation and Prefetching