



Advances on Societal Digital Transformation

DIGITAL 2021

November 14, 2021 to November 18, 2021 - Athens, Greece

[Keynote Speech]

Streamed Media and Quality of Experience (QoE)

Ph.D. Jounsup Park

Assistant Professor
The University of Texas at Tyler
jpark@uttyler.edu

Outline

- **Streamed Media**

- Video Streaming Services
- VR Video Streaming
- Hologram Streaming

- **Quality of Experience (QoE) Based Streaming Algorithms**

- Tiled Media
- Navigation Graph
 - History-based Navigation Graph
 - Semantic-aware Navigation Graph

- **Conclusion**

Outline

- **Streamed Media**

- Video Streaming Services
- VR Video Streaming
- Hologram Streaming

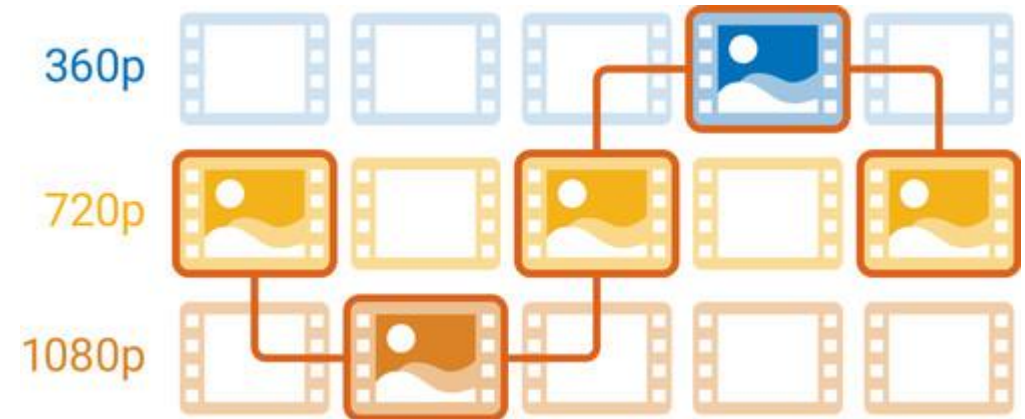
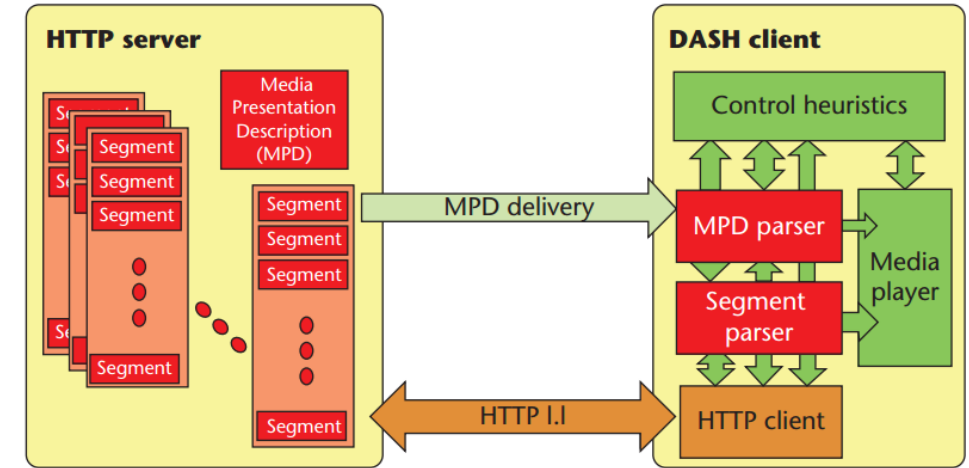
- **Quality of Experience (QoE) Based Streaming Algorithms**

- Tiled Media
- Navigation Graph
 - History-based Navigation Graph
 - Semantic-aware Navigation Graph

- **Conclusion**

Dynamic Adaptive Streaming on HTTP (DASH)

- Adaptive Bitrate Streaming platform
 - **Small video chunks (Segment)** are stored in the server
 - Server provides Media Presentation Description (MPD)
 - **Client-centric** rate adaptation
- Rate Adaptation
 - Network condition changes
 - Based on network condition and buffer status
 - Choose segment quality to request



VR/AR Streaming

- Virtual Reality (VR) and Augmented Reality (AR) traffic will increase **12-fold** between 2017 and 2022 globally

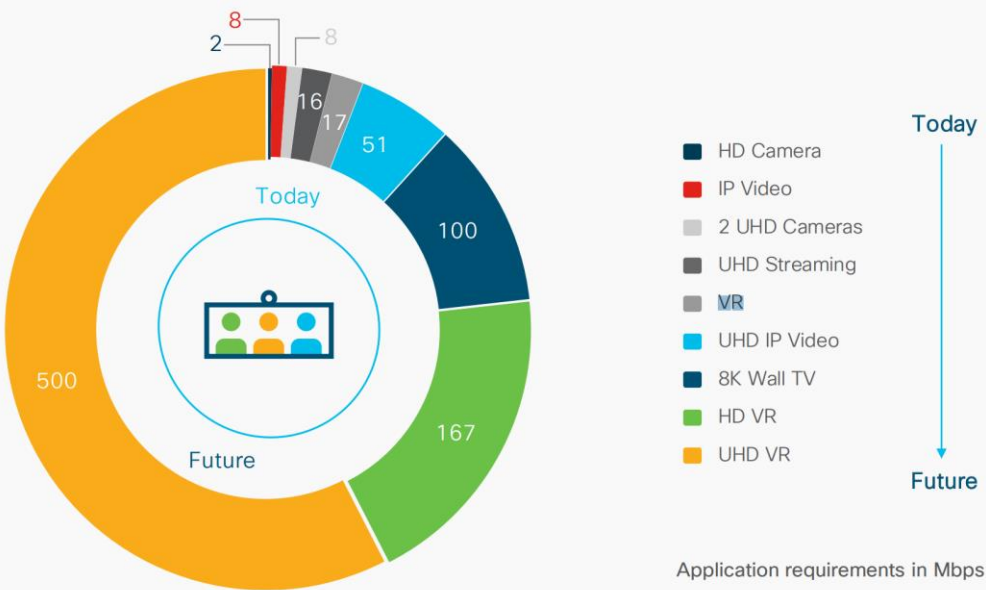


Table 6. Projected average mobile network connection speeds (in Mbps) by region and country

	2017	2018	2019	2020	2021	2022
Global						
Global speed: All handsets	8.7	13.2	17.7	21.0	24.8	28.5
Western Europe	16.0	23.6	31.2	37.2	43.8	50.5
Central and Eastern Europe	10.1	12.9	15.7	19.5	22.8	26.2
Middle East and Africa	4.4	6.9	9.4	11.2	13.2	15.3
North America	16.3	21.6	27.0	31.9	36.9	42.0
Asia Pacific	10.6	14.3	18.0	21.7	25.3	28.8
Latin America	4.9	8.0	11.2	13.0	15.3	17.7

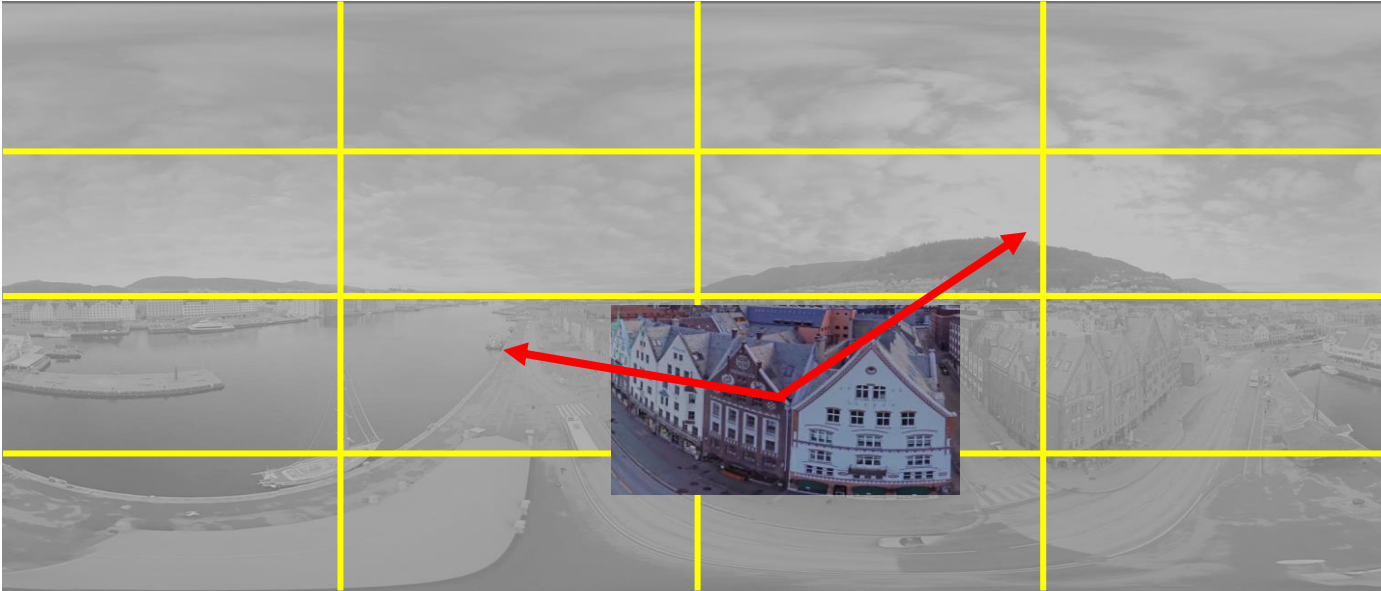
Table 7. Projected average Wi-Fi network connection speeds (in Mbps) by region and country

Region	2017	2018	2019	2020	2021	2022
Global	24.4	30.3	36.3	42.2	48.2	54.2
Asia Pacific	26.7	34.5	42.2	47.6	56.0	63.3
Latin America	9.0	10.6	12.1	13.8	15.2	16.8
North America	37.1	46.9	56.8	63.6	74.4	83.8
Western Europe	25.0	30.8	36.3	37.7	44.6	49.5
Central and Eastern Europe	19.5	22.6	24.1	27.4	30.1	32.8
Middle East and Africa	6.2	7.0	7.9	9.6	10.2	11.2

Source: Cisco VNI, 2018.

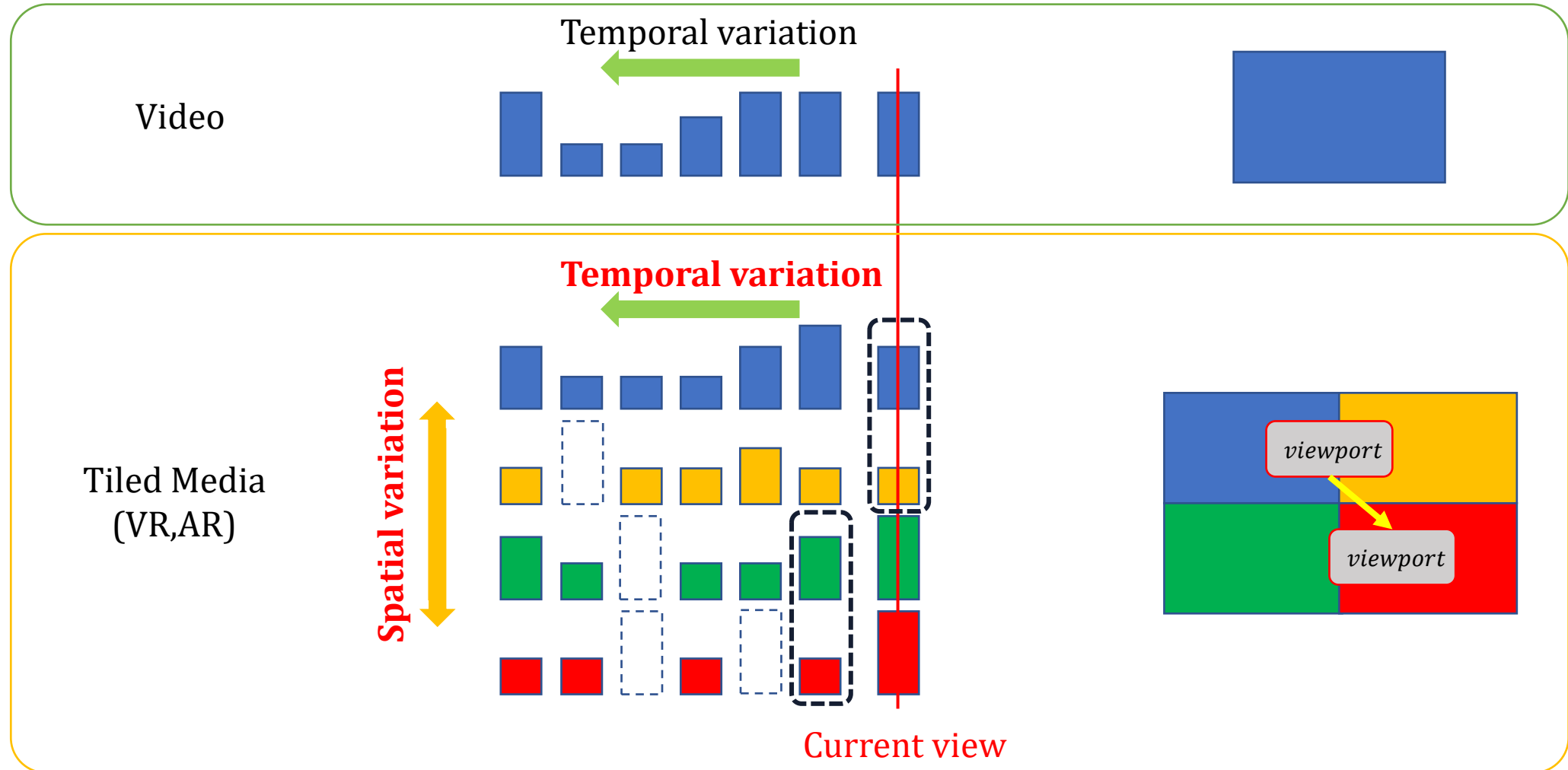
Challenges in 360-degree video streaming

- Much **higher video rate** (4~6 times more than conventional videos): at least 4K
- Can save up to 80% of bandwidth by removing **non-visible part** of the video (typically viewport is only 20% of whole 360-degree video)
- 360-degree Video is divided into tiles representing **tiled media**
 - **Spatial Relationship Descriptor** (SRD) describes tile configuration



Challenges in 360-degree video streaming

- View prediction is important to request future segments
 - Information of **Spatial and Temporal variation of view**



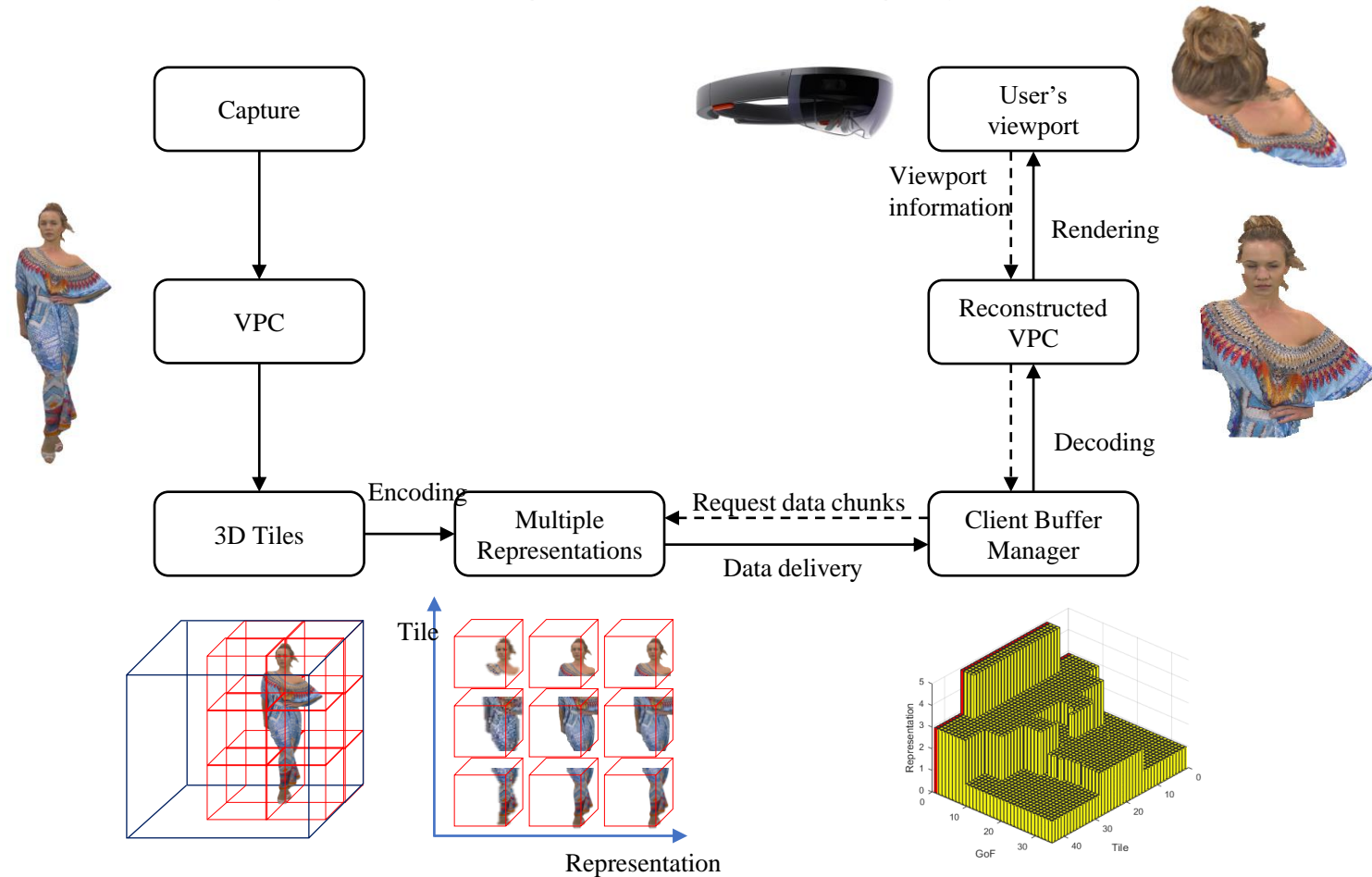
3D Point Cloud

- Color and Geometry
- Hologram



3D tile streaming system

- Network and User Adaptive Hologram Streaming System



Outline

- **Streamed Media**

- Video Streaming Services
- VR Video Streaming
- Hologram Streaming

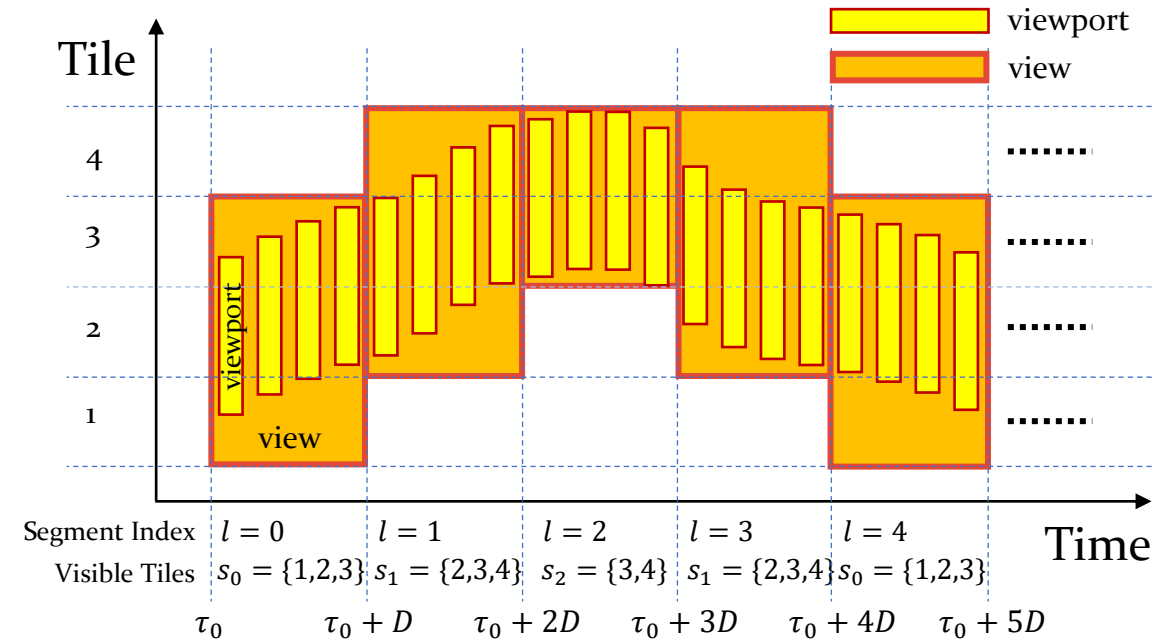
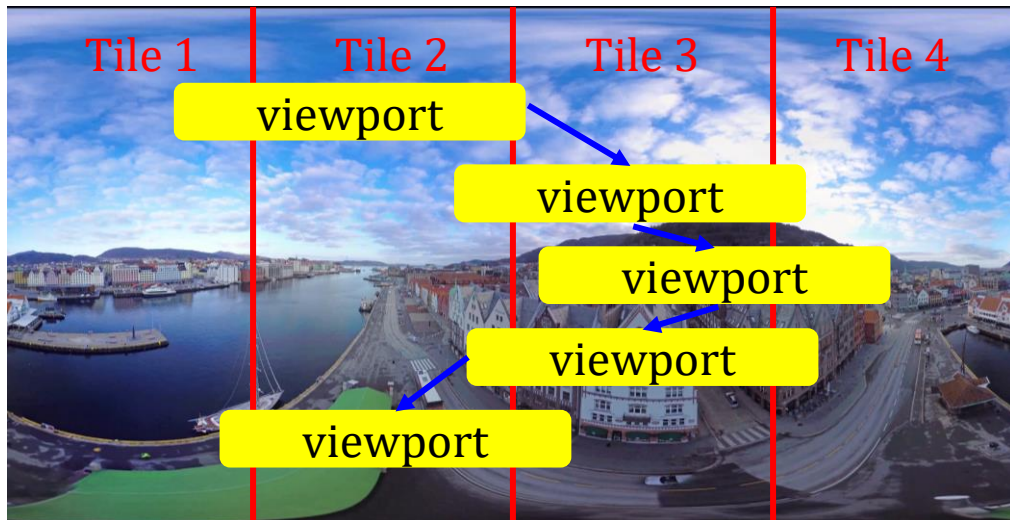
- **Quality of Experience (QoE) Based Streaming Algorithms**

- Tiled Media
- Navigation Graph
 - History-based Navigation Graph
 - Semantic-aware Navigation Graph

- **Conclusion**

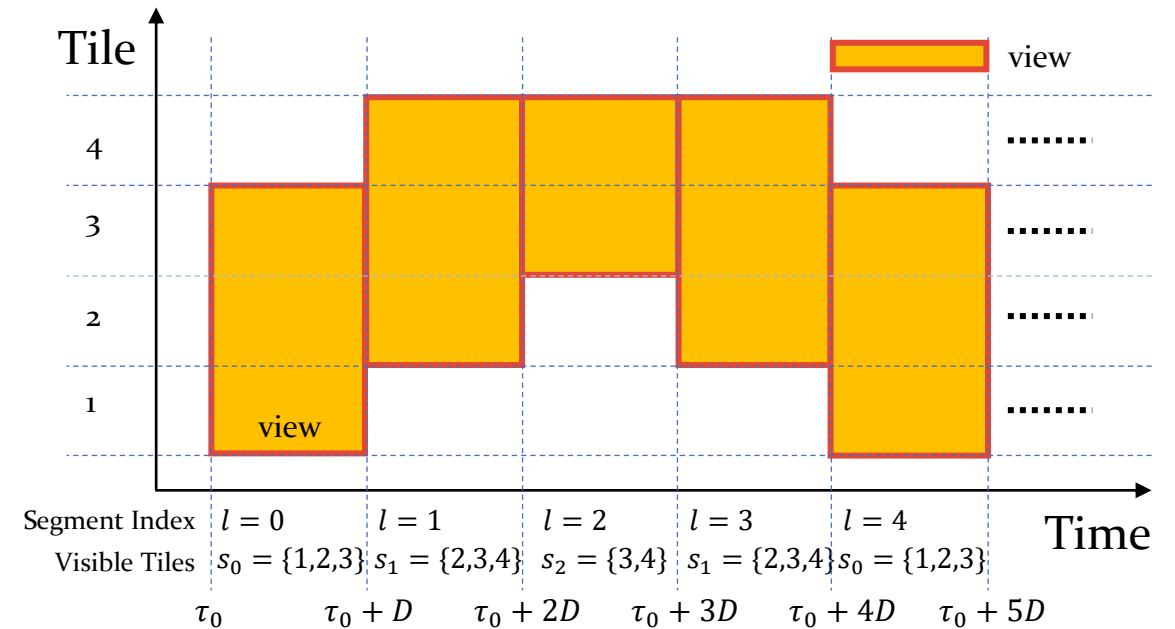
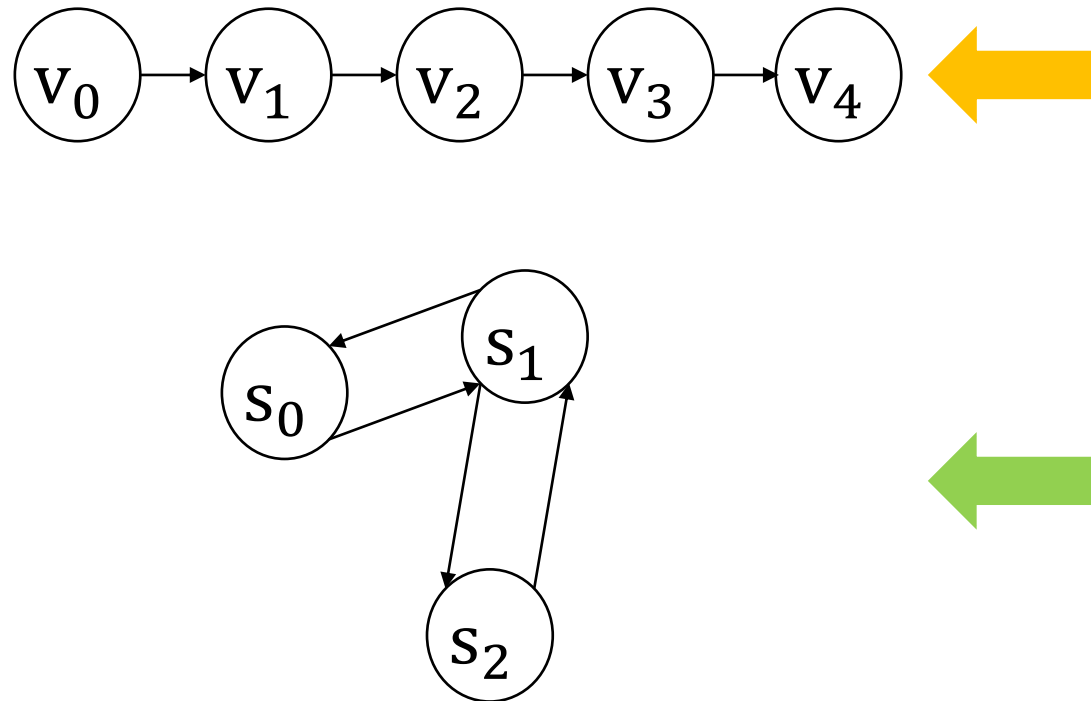
Segments, Tiles, and Viewports

- Segment: 1~15sec duration (D) video chunk
- Tile: Spatially divided video
- Viewport: Pixels that users watch



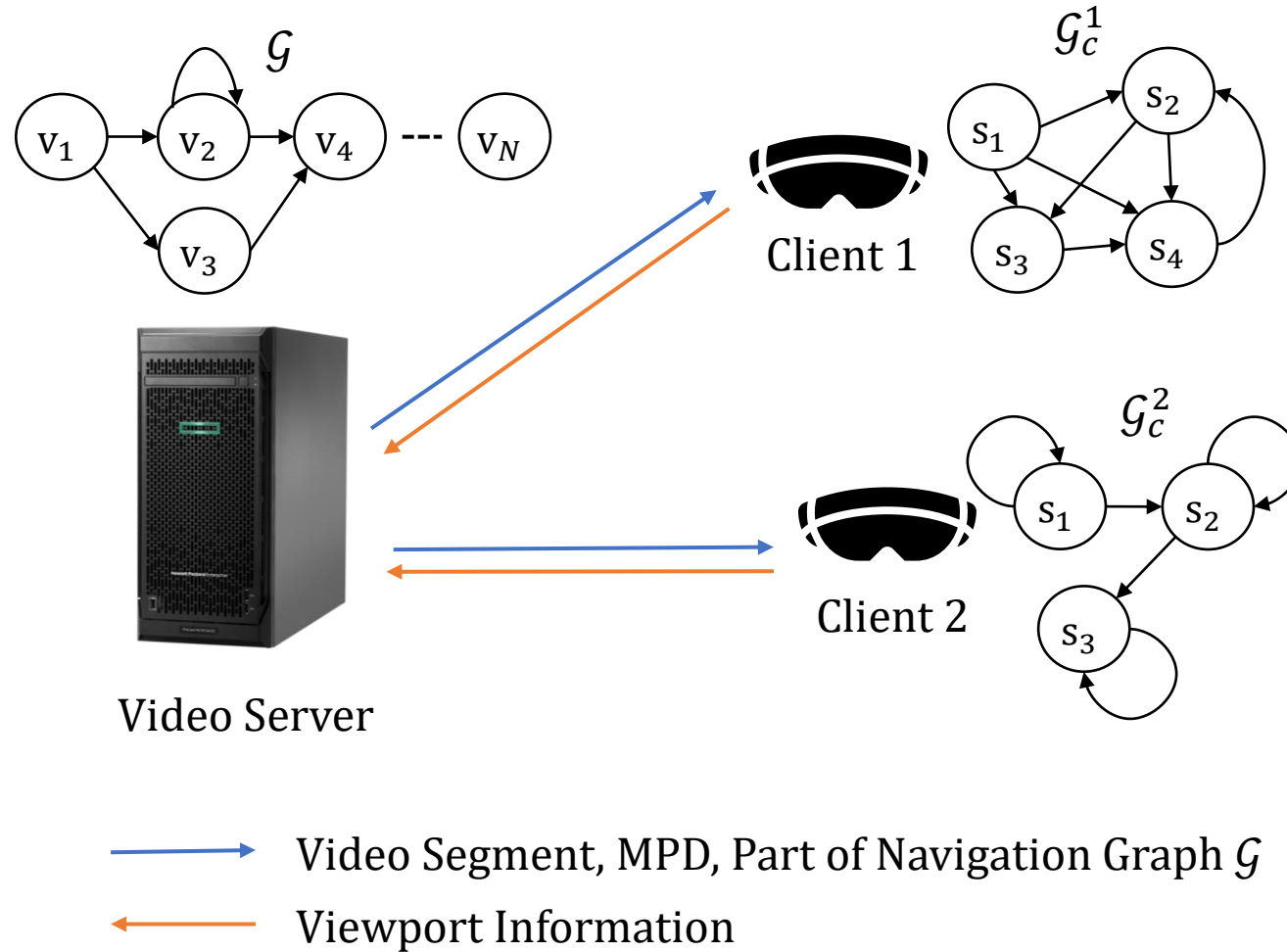
Navigation Graph

- Two different ways to define vertex
 - v : {segment index, set of visible tiles} $\rightarrow v_0 = \{l = 0, s_0 = (1,2,3)\}$
 - s : (set of visible tiles) $\rightarrow \{s_0 = (1,2,3)\}$
- Directed edges to connect vertices



System Overview

- Video server collects viewing data from clients
- Clients collect their past viewing data



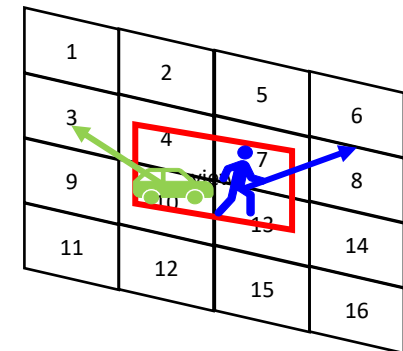
Applications of Navigation Graph

- **View History**

- Single-user view prediction
 - Navigation Graph in the client-end captures the user's viewing pattern
 - Spatial location of the current view affect future view change
- Cross-user view prediction
 - View transition information of prior users is recorded in the server as a Navigation Graph
 - It captures segment-by-segment view transition probabilities

- **Semantic Information**

- Object-level view prediction
 - Video analysis is performed in the server and server stores objects'(or saliency part) trajectories as a Navigation Graph
 - It can be used to categorize users by the objects they are tracking
- Story telling based streaming
 - Director's story line can be encoded as a Navigation Graph
 - Clients can follow the story line without moving their head



Outline

- **Streamed Media**

- Video Streaming Services
- VR Video Streaming
- Hologram Streaming

- **Quality of Experience (QoE) Based Streaming Algorithms**

- Tiled Media
- Navigation Graph
 - **History-based Navigation Graph**
 - Semantic-aware Navigation Graph

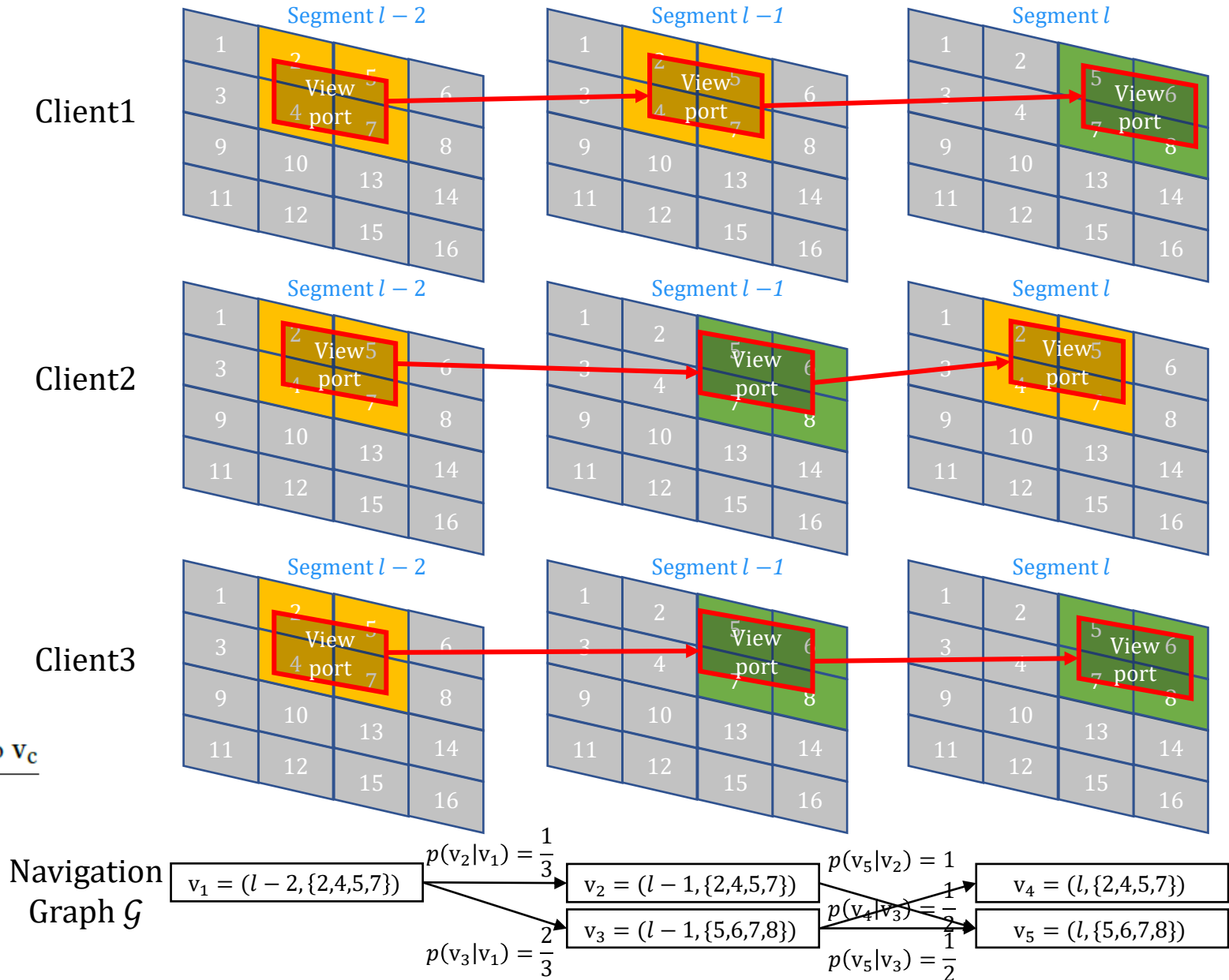
- **Conclusion**

Navigation Graph at the Server

- Collects all users' view data
- Vertex consists of
 - Segment index
 - Set of visible tiles

$$v_1 = (l - 2, \{2, 4, 5, 7\})$$

$$p(v_c | v_p) = \frac{\text{number of clients moving their view from } v_p \text{ to } v_c}{\text{number of clients visiting } v_p}$$

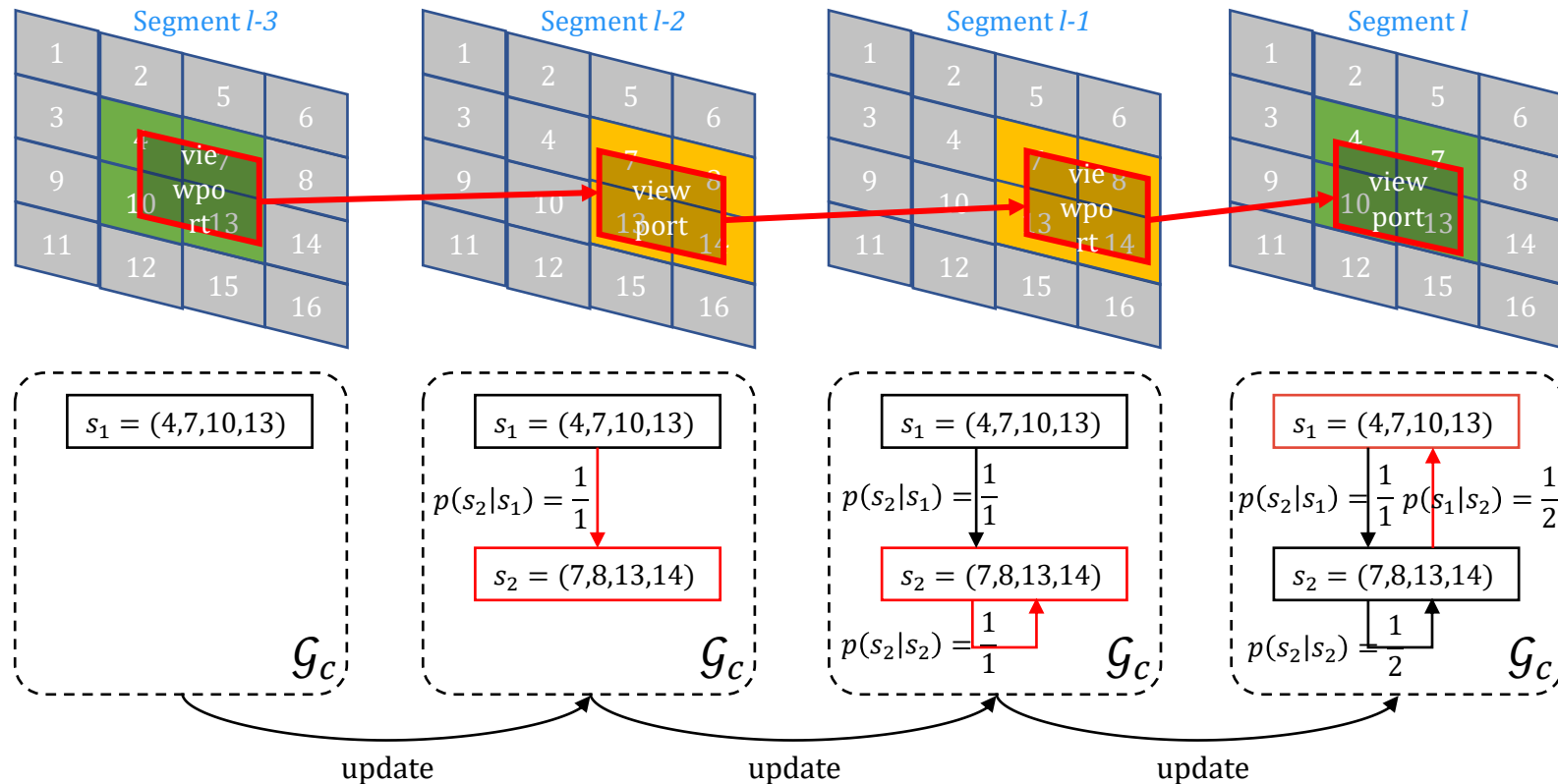


Navigation Graph at the Clients

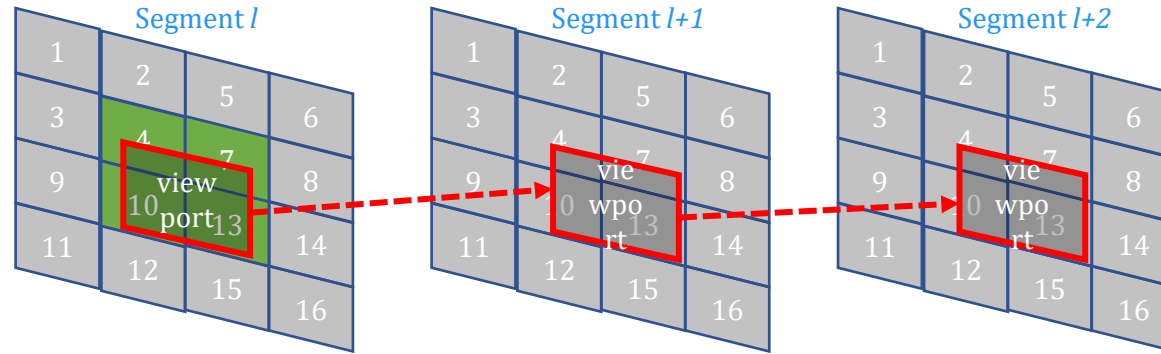
- Records past viewing behavior of the client itself
- Vertex consists of a set of visible tiles

$$s_1 = (4,7,10,13)$$

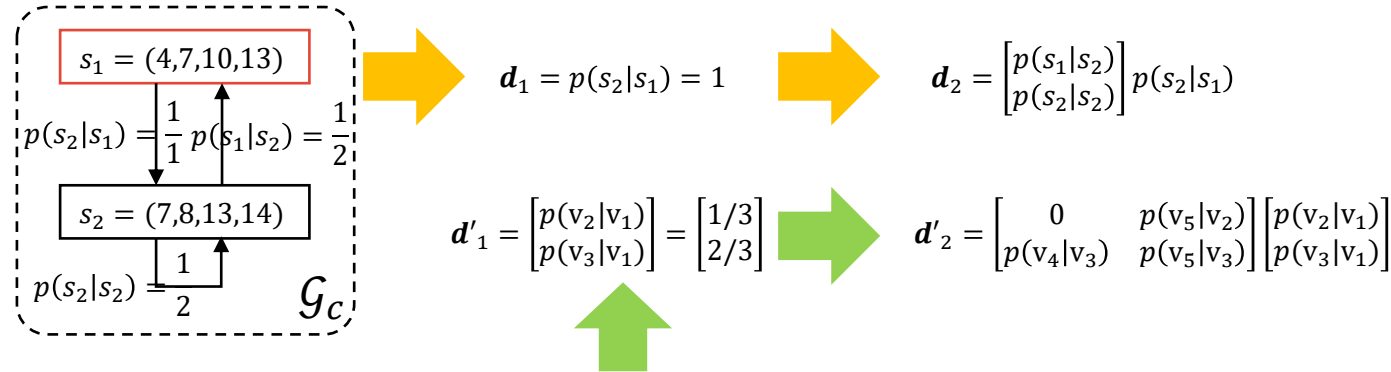
$$p(s_c|s_p) = \frac{\text{number of transitions from } s_p \text{ to } s_c}{\text{number of times visiting the set of visible tiles } s_p}$$



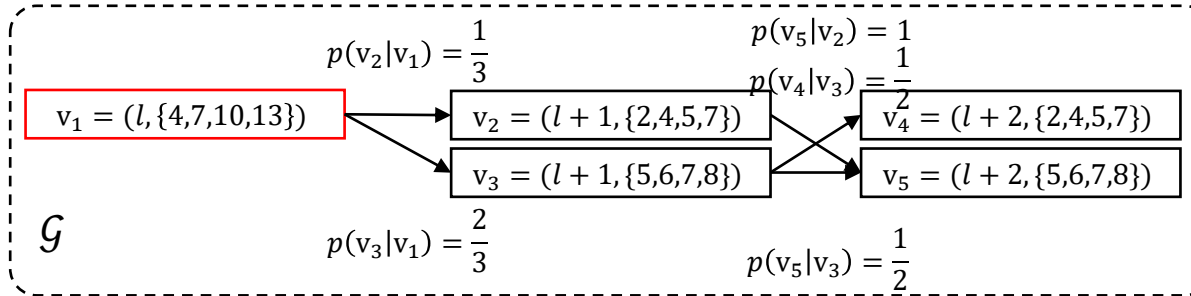
View Prediction with Navigation Graph



Single-User Prediction



Cross-User Prediction



Prediction Matrix

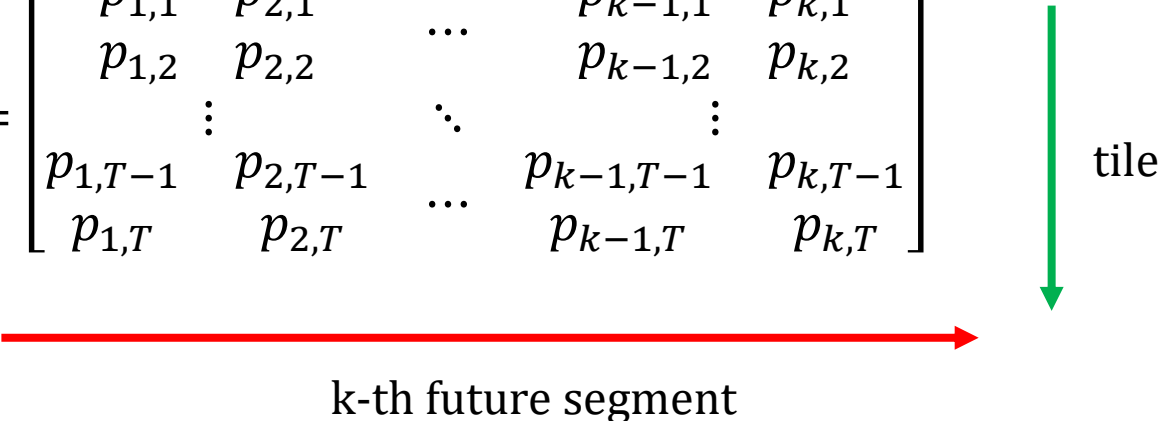
- Navigation Graph to Tile-level
 - Prediction vector for k th future segment, where m is the number of states

$$\mathbf{d}_k = [d_{k,1} \ d_{k,2} \ \dots \ d_{k,m}]^T$$

- Tile prediction from prediction vector

$$p_{k,t} = \sum_{\forall m, t \in \mathbf{v}_m} d_{k,m}$$

- Prediction matrix \mathbf{P}

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{2,1} & \dots & p_{k-1,1} & p_{k,1} \\ p_{1,2} & p_{2,2} & & p_{k-1,2} & p_{k,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{1,T-1} & p_{2,T-1} & \dots & p_{k-1,T-1} & p_{k,T-1} \\ p_{1,T} & p_{2,T} & \dots & p_{k-1,T} & p_{k,T} \end{bmatrix}$$


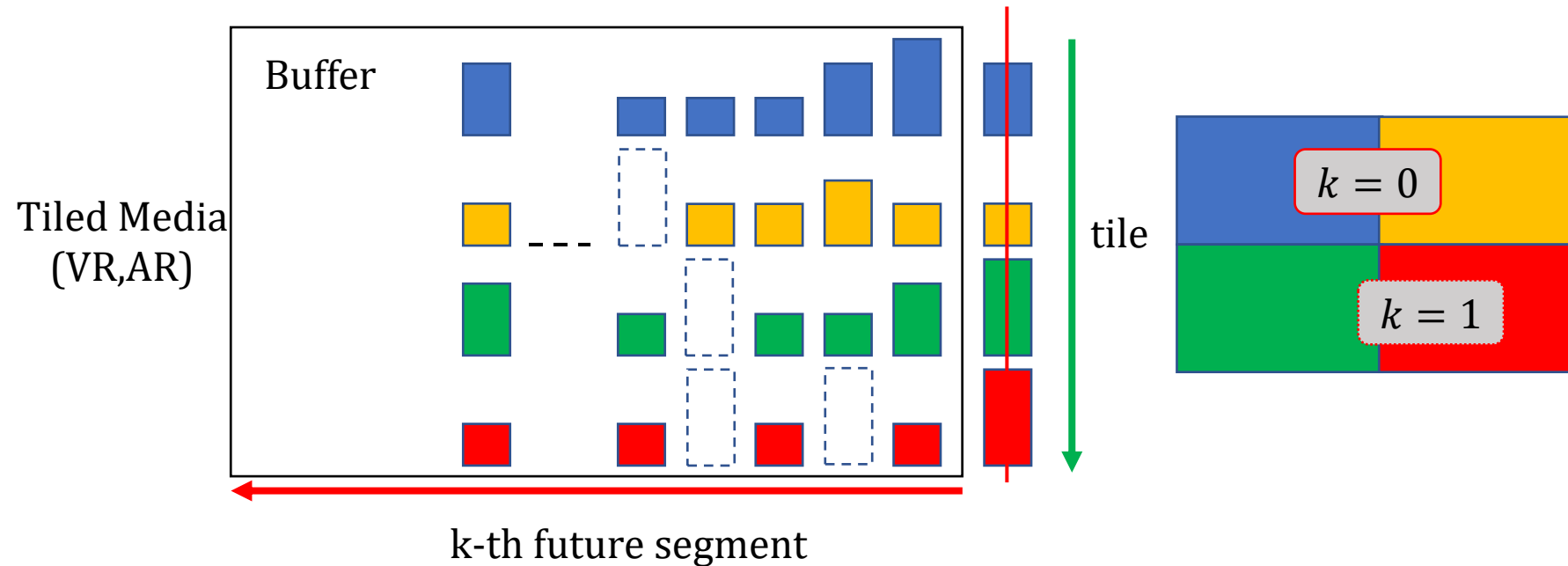
Rate Selection for Tiled-Media Streaming

- Prediction matrix is derived from the Navigation Graph
- Utility maximization problem

$$\underset{q}{\text{maximize}} \sum_{\forall k} \sum_{\forall t} u(q_{t,k}) p_{t,k} \quad \text{subject to} \quad \sum_{\forall k} \sum_{\forall t} r(q_{t,k}) \leq BW$$

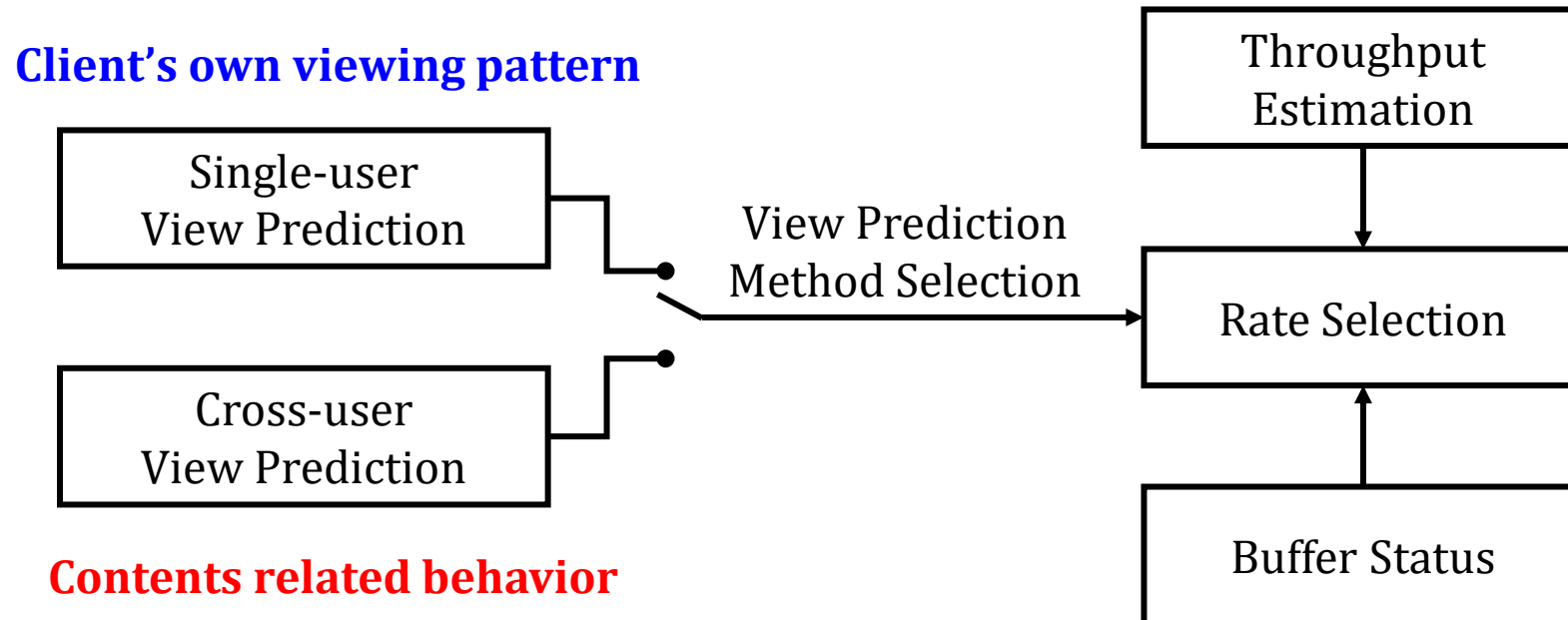
• Greedy algorithm allocates $q_{t,k}$ that has the largest $\frac{u(q_{t,k}) p_{t,k}}{r(q_{t,k})}$

q : quality matrix size of $T \times K$
 $q_{t,k}$: quality index of tile t of segment k
 $p_{t,k}$: prediction value
 $u(q)$: utility function
 $r(q)$: required bitrate for q



View Prediction Method Selection

- Single-user prediction (SU) or Cross-user prediction (CU) is used to decide the Prediction Matrix
- Switching based on precision values of prior decision



Evaluation Method

- Videos and Head motion data
 - 9 videos tested by 48 users
 - Reference: Chenglei Wu, Zhihao Tan, Zhi Wang, Shiqiang Yang. "A Dataset for Exploring User Behaviors in VR Spherical Video Streaming," In Proceedings of ACM Multimedia Systems (MMSys) 2017, Taipei, Taiwan, June 20-23, 2017
- Network Traces
 - HSDPA-bandwidth logs for mobile HTTP streaming scenarios
 - Reference: Haakon Riiser, Paul Vigmostad, Carsten Griwodz, Pål Halvorsen, "Commute Path Bandwidth Traces from 3G Networks: Analysis and Applications", Proceedings of the International Conference on Multimedia Systems (MMSys), Oslo, Norway, February/March 2013, pp. 114-118
- Measurements
 - Precision
 - Prediction Error
 - V-PSNR
 - Effective Rate

Prediction Performance

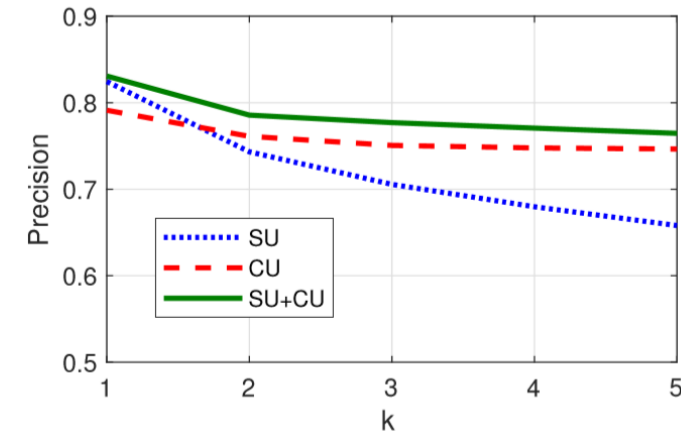
- Prediction Precision for k -th future segment

$$\sum_{t=1}^T \min\{p_{t,k}, g_{t,k}\}$$

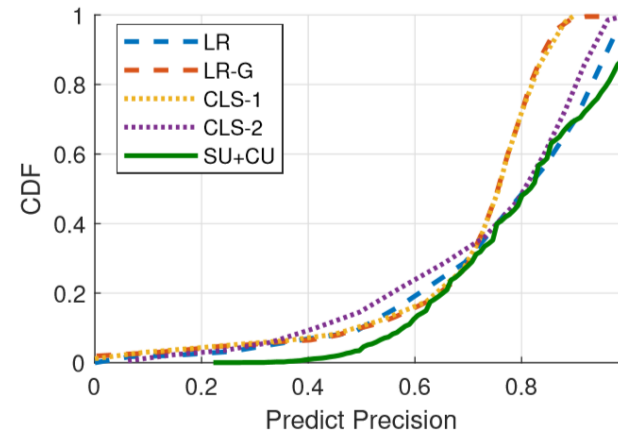
where $g_{t,k}$ is normalized ground truth

- Ground truth: $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$
- Normalized ground truth: $\begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}$

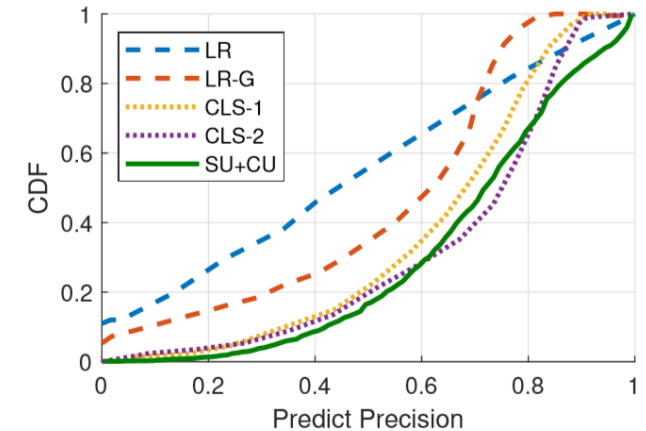
- Compared with
 - LR: Linear Regression
 - LR-G: Linear Regression with Gaussian assumption
 - CLS-1: without user classification
 - CLS-2: with user classification
 - SU: Single-User Prediction
 - CU: Cross-User Prediction
 - SU+CU: switch between SU and CU**



(a) Average Precision



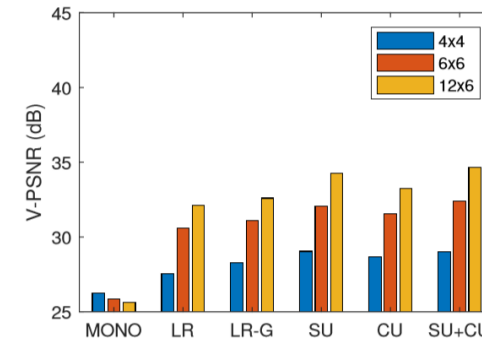
(b) Prediction of 1-sec after current segment



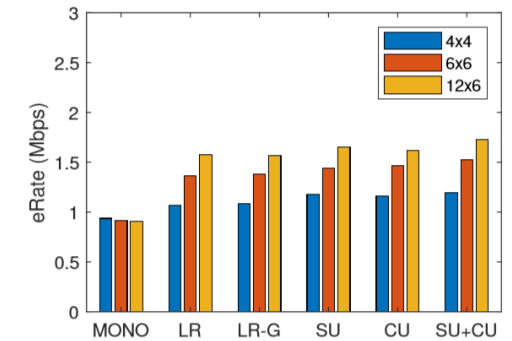
(c) Prediction of 5-sec after current segment

V-PSNR and Effective Rates

- Viewport PSNR (V-PSNR)
 - PSNR in the viewport
- Effective Rate
 - Actual rate for visible tiles
- Tile configuration
 - 12x6, 6x6, 4x4
- Segment Durations
 - 1.0sec, 1.5sec, 2.0sec

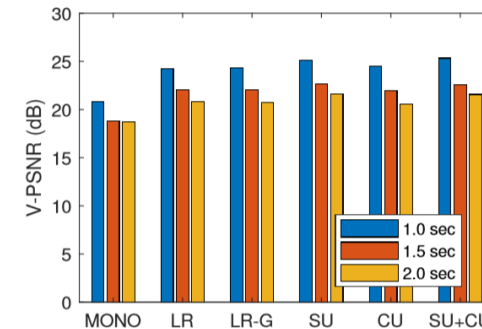


(a) V-PSNR

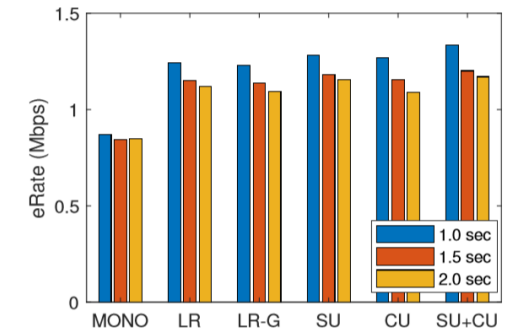


(b) Effective Rate

Figure 9: V-PSNR and eRate in Real Mobile Traces with Different Tile Configurations (Segment Duration = 1.0 sec)



(a) V-PSNR



(b) Effective Rate

Figure 10: V-PSNR and eRate in Real Mobile Traces with Different Segment Durations (Tile Configuration = 6×6)

Outline

- **Streamed Media**

- Video Streaming Services
- VR Video Streaming
- Hologram Streaming

- **Quality of Experience (QoE) Based Streaming Algorithms**

- Tiled Media
- Navigation Graph
 - History-based Navigation Graph
 - **Semantic-aware Navigation Graph**

- **Conclusion**

View Prediction and User Experience

- Navigation Graph (NG)
 - ➔ Multiple viewers' view traces

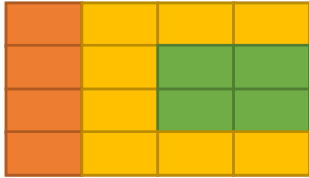


Server

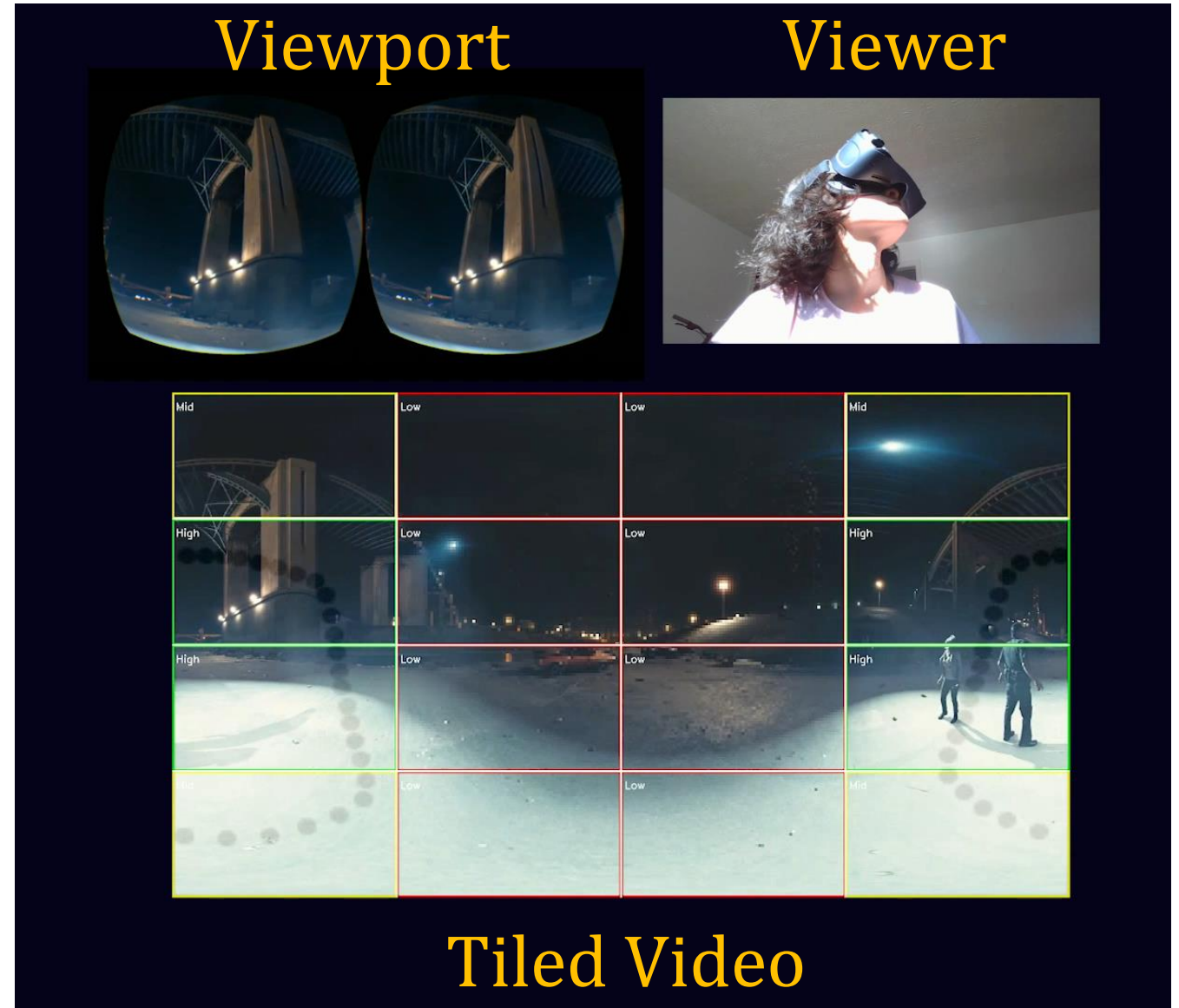
High Quality

Medium Quality

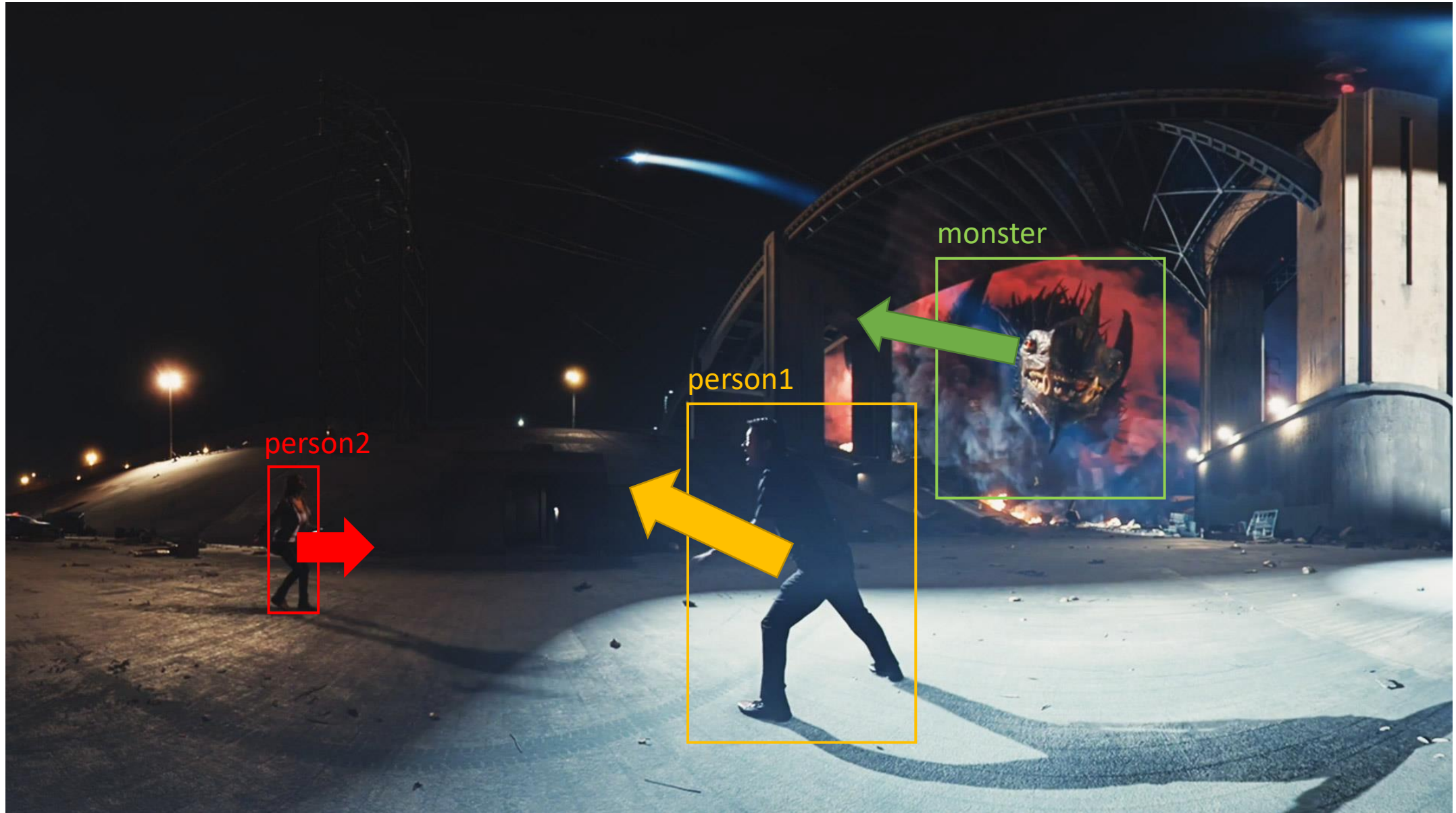
Low Quality



Client

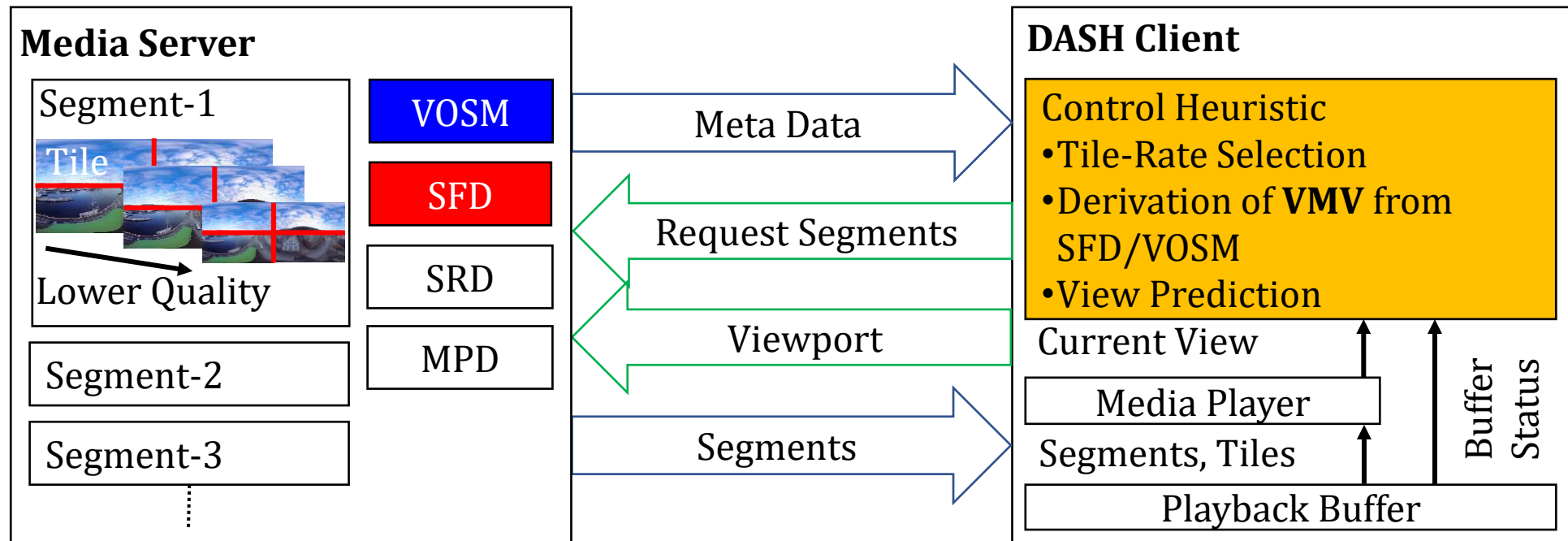


Motivation



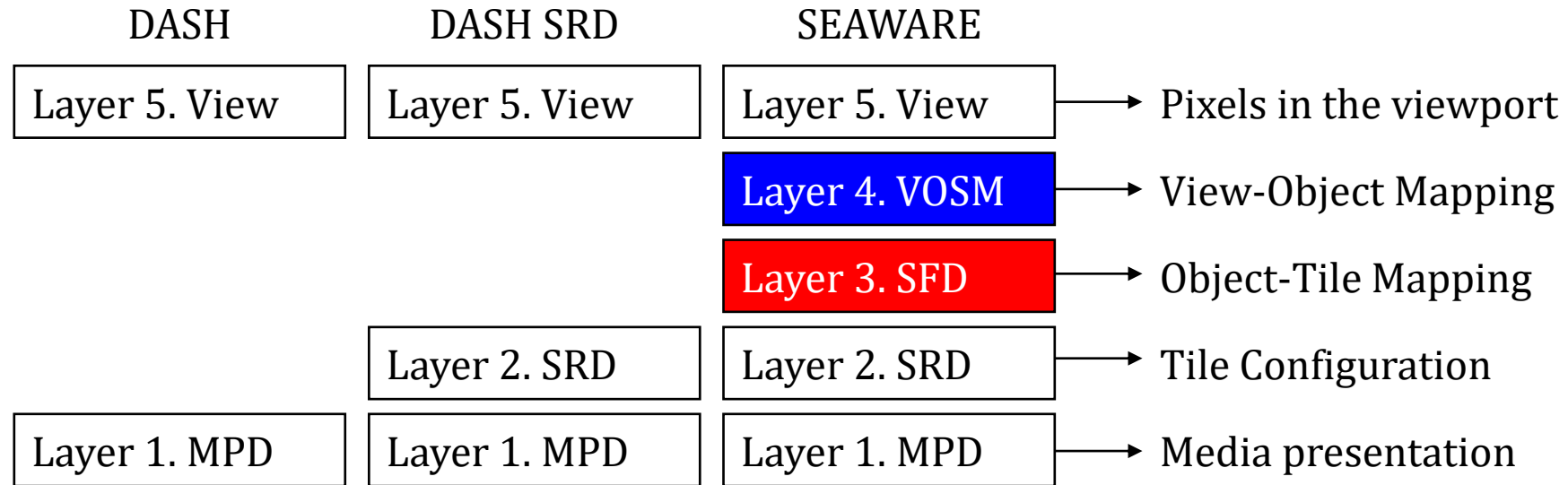
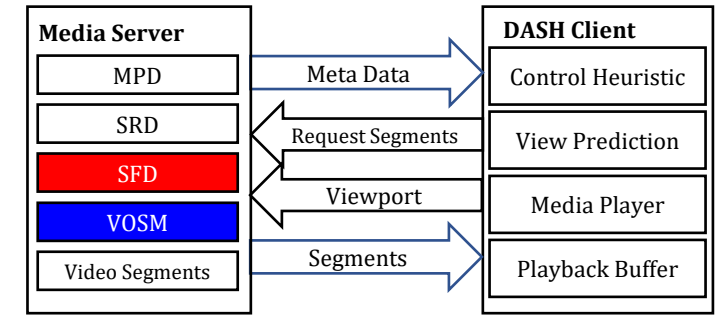
SEAWARE System

- Advanced MPD: Semantic Flow Descriptor (SFD) and View-Object State Machine (VOSM)
- Control Heuristic: Semantic-Aware View Prediction



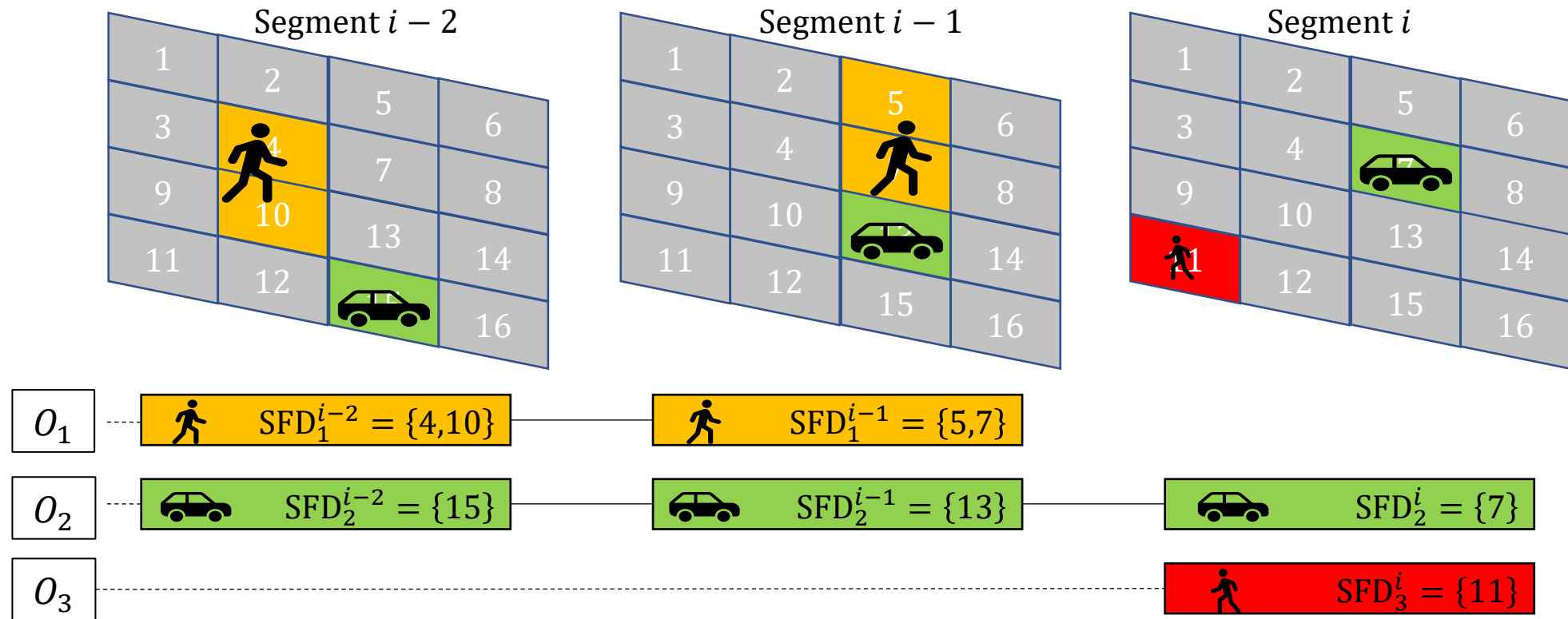
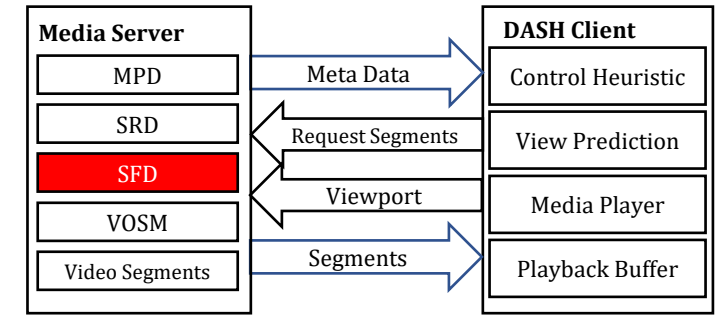
Layered Architecture of MPD

- Media Presentation Description (MPD) for DASH
- Spatial Relationship Description (SRD) for Tiled Media



Semantic Flow Descriptor (SFD)

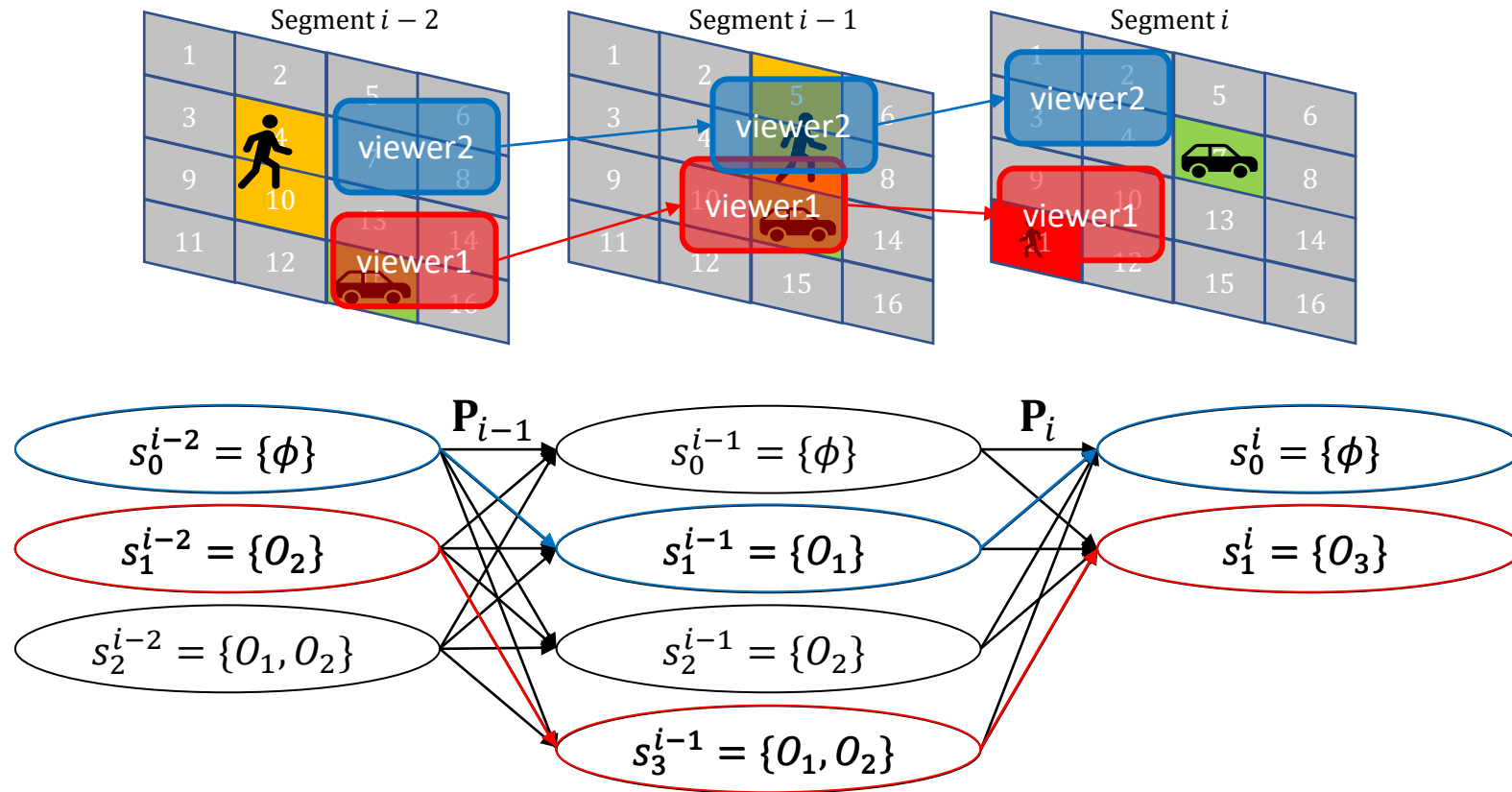
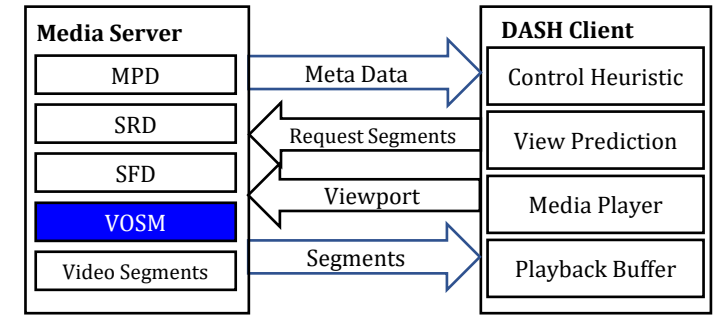
- SFD stores location information of objects



View-Object State Machine (VOSM)

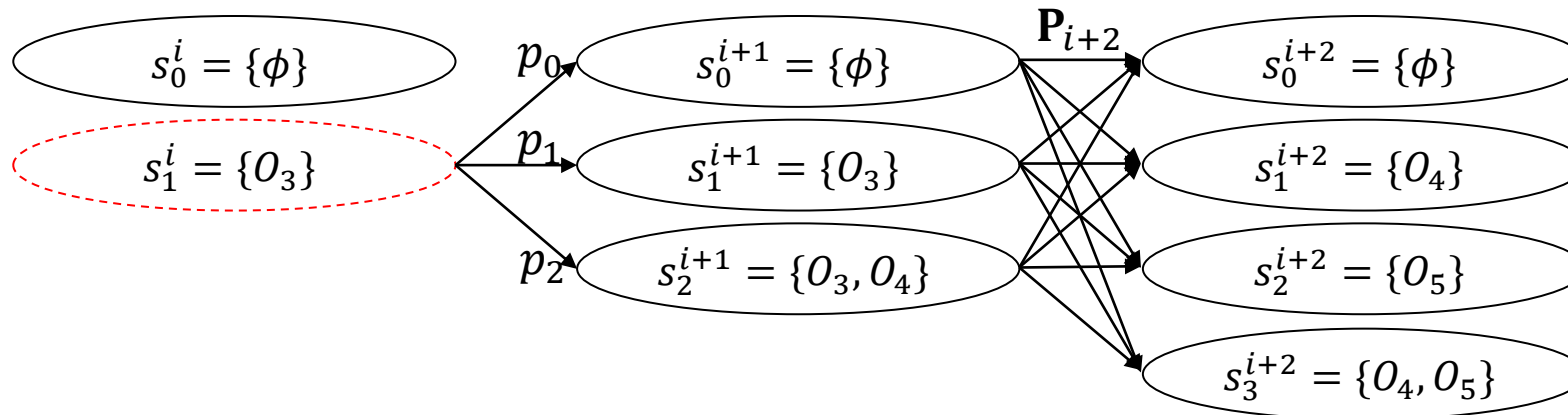
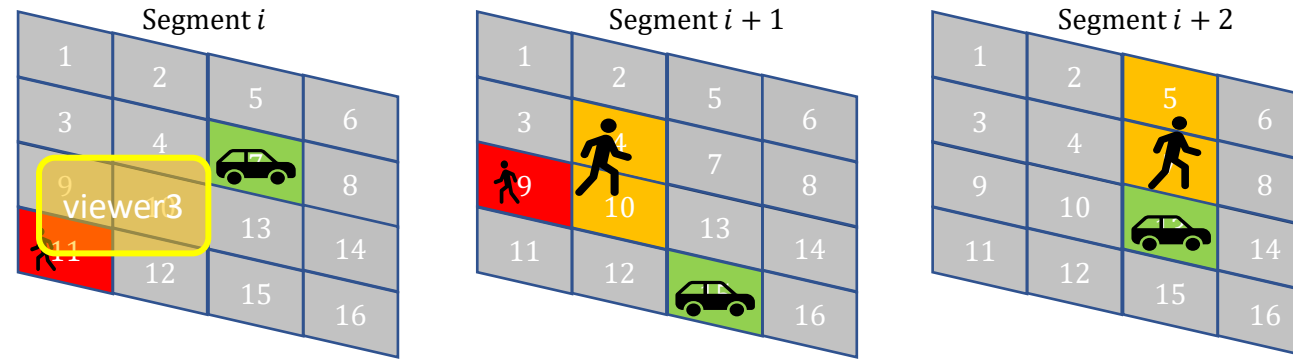
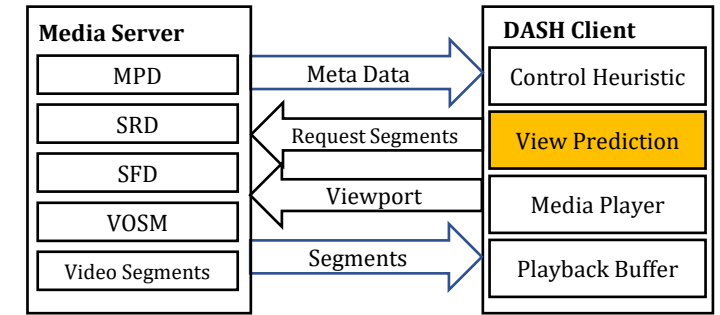
- States indicate the set of objects: $\{\{\emptyset\}, \{O_1\}, \{O_2\}, \{O_1, O_2\}, \dots\}$
- Transition Probability

$$p(s_m^i | s_c^{i-1}) = \frac{\text{number of clients change their state from } s_c^{i-1} \text{ to } s_m^i}{\text{number of clients visiting } s_c^{i-1}}$$



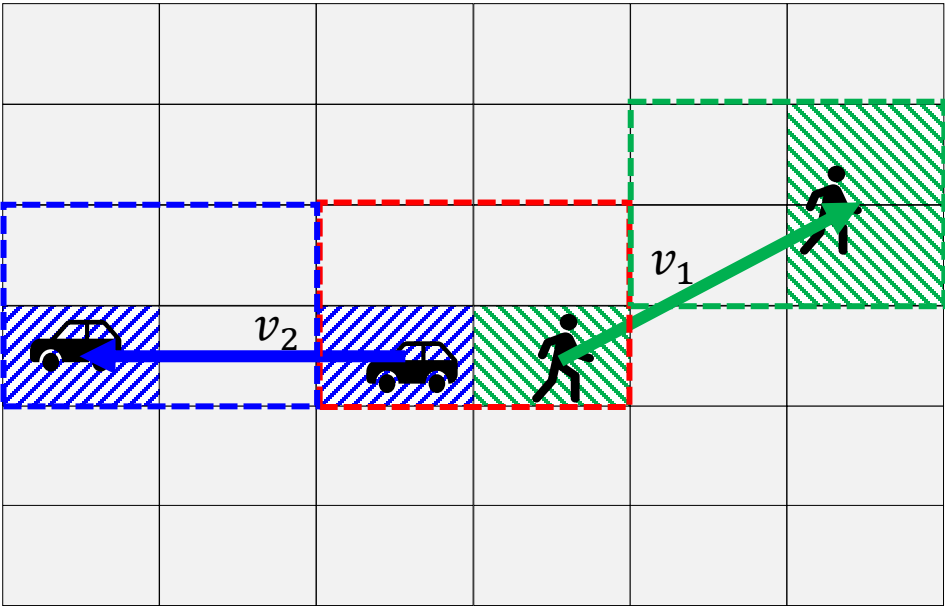
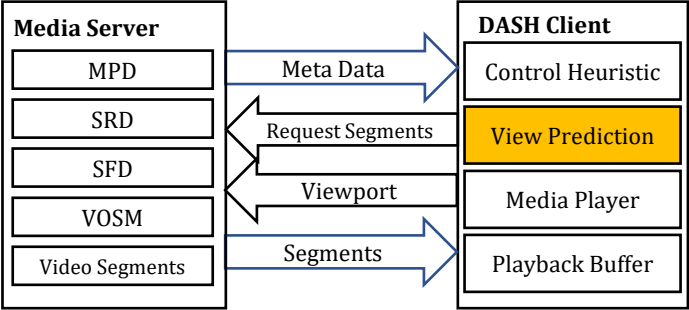
View Prediction

- Which object the viewer will watch in next segment?



View Motion Vector (VMV)

- Which tiles are required to show the objects?



- Current View at i
- Future View-1 at $i+1$
- Future View-2 at $i+1$
- VMV_1
- VMV_2
- Object-1
- Object-2
- SFD_1^i, SFD_1^{i+1}
- SFD_2^i, SFD_2^{i+1}

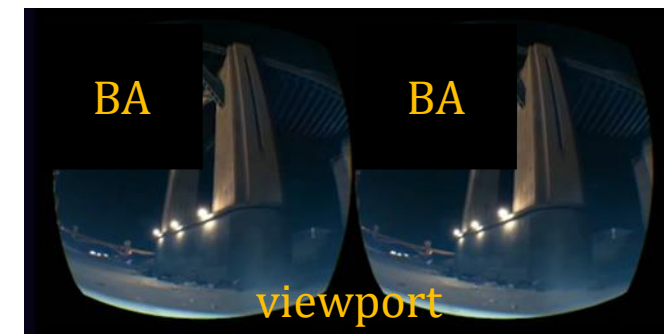
Experimental Setup

- Dataset
 - 9 360-degree videos, 48 viewers^[3] ACM MMsyst2017
 - 12x6 tiles with 5 representations
 - YOLOv3 based object detection and tracking
- Baselines
 - Navigation Graph (NG)^[1] ACM MM 2019: Multiple viewers' data
 - Linear Regression (LR)^[2] ACM MM 2017: Single viewer's data
 - MONO: Every tile has same probability
 - Ideal: Perfect view prediction
- Network condition
 - Wired network: constant throughput
 - Mobile network: variable throughput^[4] ACM TOMM 2012

No	Content	Length	Category
1	Conan	2'44"	Performance
2	Ski	3'21"	Sport
3	Help	4'53"	Film
4	Conan	2'52"	Performance
5	Tahiti Surf	3'25"	Sport
6	The fight for Falluja	10'55"	Documentary
7	Cooking Battle	7'31"	Performance
8	LOSC Football	2'44"	Sport
9	The Last of the Rhinos	4'53"	Documentary

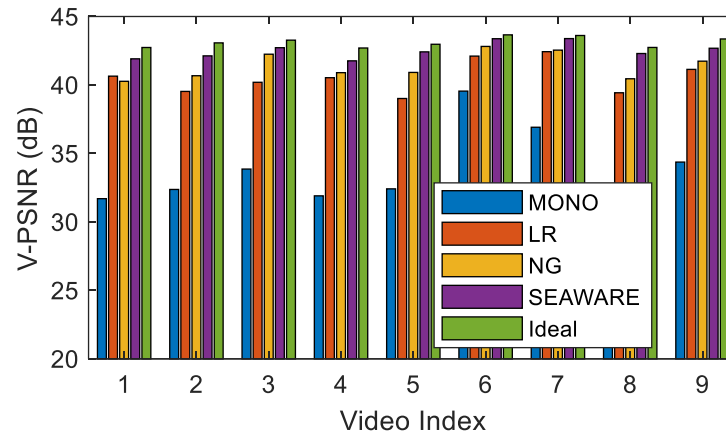
Streaming Performance

- Viewport-PSNR (V-PSNR) and Average Blank Area (BA)

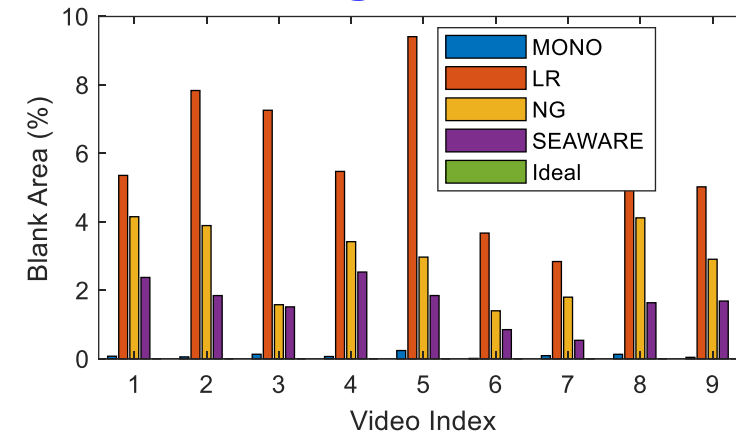


Wired Network

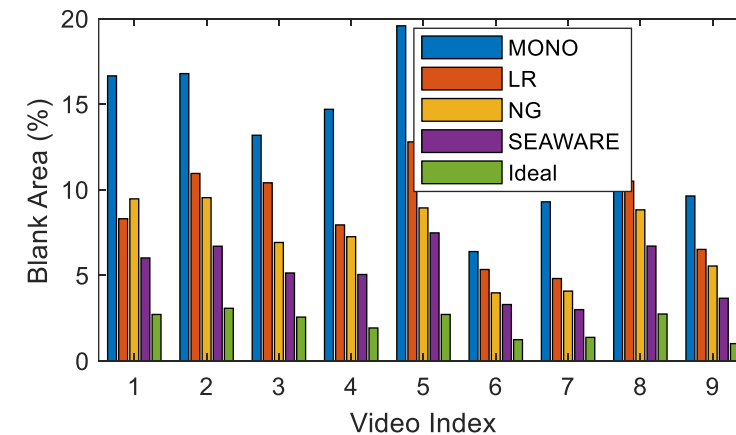
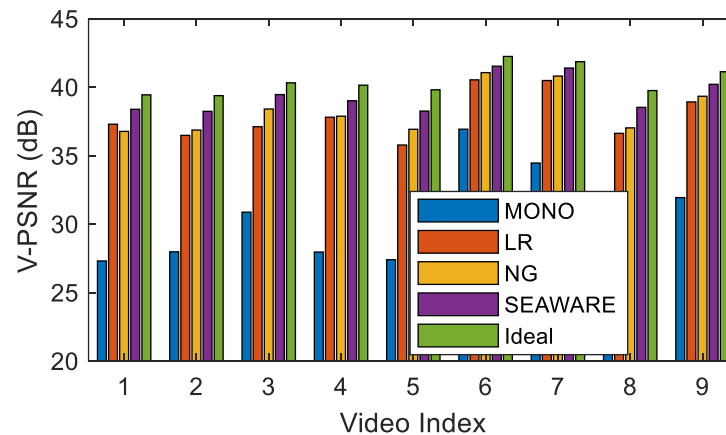
V-PSNR



Average Blank Area



Mobile Network



Overhead Analysis

- Server: Ubuntu machine that has CPU (Intel Xeon) and GPU (NVIDIA Quadro P4000) with 32GB memory
- Client device: Intel Core i7, 8GB memory not using GPU

	Storage	Average Processing Time	
	per video	Off-line (per video)	On-line (per segment)
SFD	100KB	4min	-
VOSM	10KB	25sec	-
View Prediction	-	-	2.2ms

- View prediction: NG-19ms, LR-1.9ms

Outline

- **Streamed Media**

- Video Streaming Services
- VR Video Streaming
- Hologram Streaming

- **Quality of Experience (QoE) Based Streaming Algorithms**

- Tiled Media
- Navigation Graph
 - History-based Navigation Graph
 - Semantic-aware Navigation Graph

- **Conclusion**

Conclusion

- Various Multimedia data (Video, VR and AR) are streamed over the internet
- Navigation Graph helps improving QoE of streamed media by predicting future behavior