



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



北京大学
PEKING UNIVERSITY

超高清全景视频处理和传输

张行功
北京大学
2019. 8. 23



出品: LiveVideoStack
—— 音视频技术社区 ——

CSDN



深圳
2019

遨游“视”界 做你所想
Explore World, Do What You Want

LiveVideoStackCon 2019 深圳

2019.12.13-14



出品:  LiveVideoStack
—— 音视频技术社区 ——

成为讲师: speaker@livevideostack.com

成为志愿者: volunteer@livevideostack.com

赞助、商务合作: kathy@livevideostack.com

网络视频组 | NET VIDEO GROUP

<http://www.icst.pku.edu.cn/netvideo>



■ 北京大学NetVideo组

- 专注网络视频传输、未来网络基础研究和应用。
- 包括实时视频通信、自适应流媒体、ICN&CDN等

■ 成果

- SIGCOMM等论文50多篇
- 专利30余项
- 国际和国内标准3项
- 2次获得ICME DASH Challenge奖



目录



背景



基于用户行为的QoE分析



虚拟现实视频视点预测



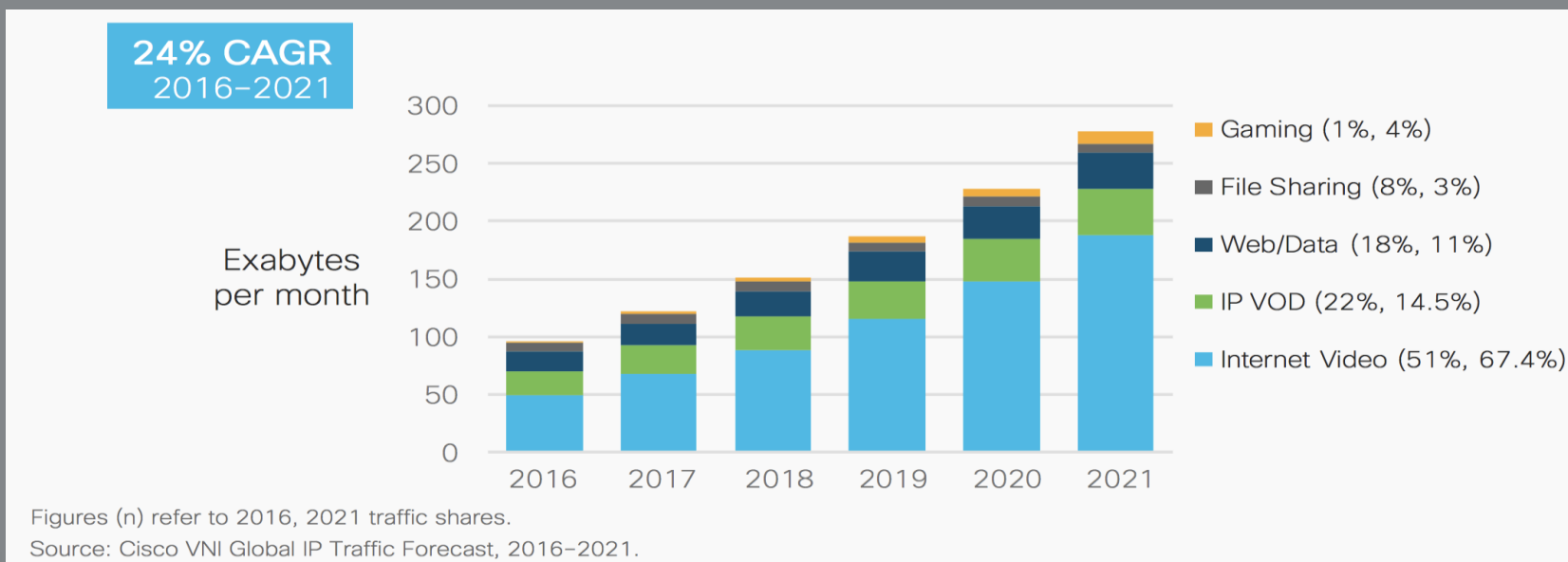
QoE驱动的视点自适应传输



总结

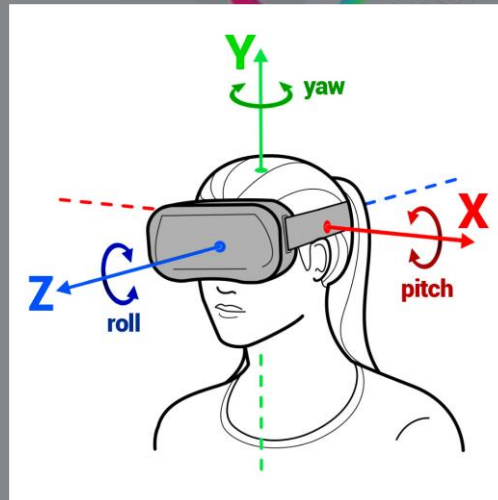
高清/超高清/VR视频应用

- 网络视频流量24%复合年增长率，将于2021年达到82%
- 超高清视频/ 4K 18Mbps/高清（High-Definition）7.2Mbps/标清（Standard-Definition）2Mbps
- VR视频（Immersive Video）预计2021年将增长20倍



全景视频传输的挑战

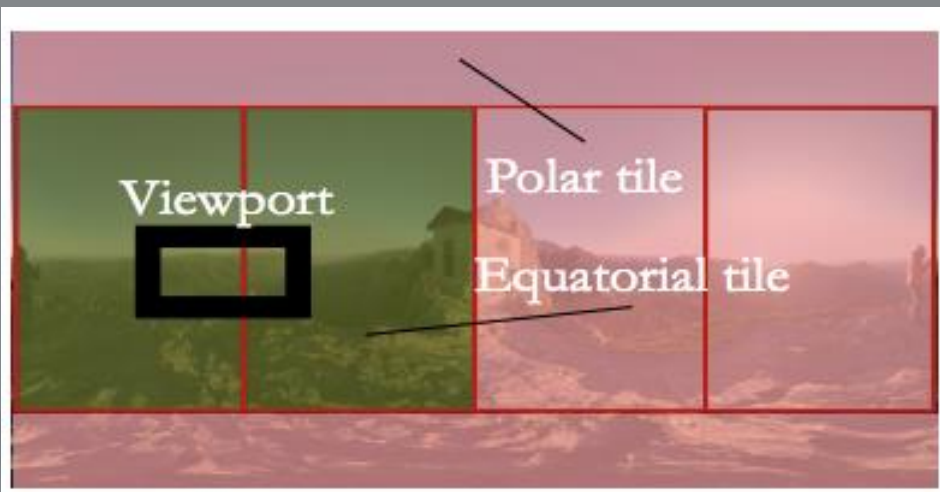
- 码率高
 - 4k (4096x2048) 15-20Mbps
 - 8k (8192x4096) 70-100Mbps
- Motion-photon延迟(20ms)
 - 解码延迟
 - 渲染延迟
- 网络Streaming QoE
 - 卡顿、启动延迟



全景ERP传输

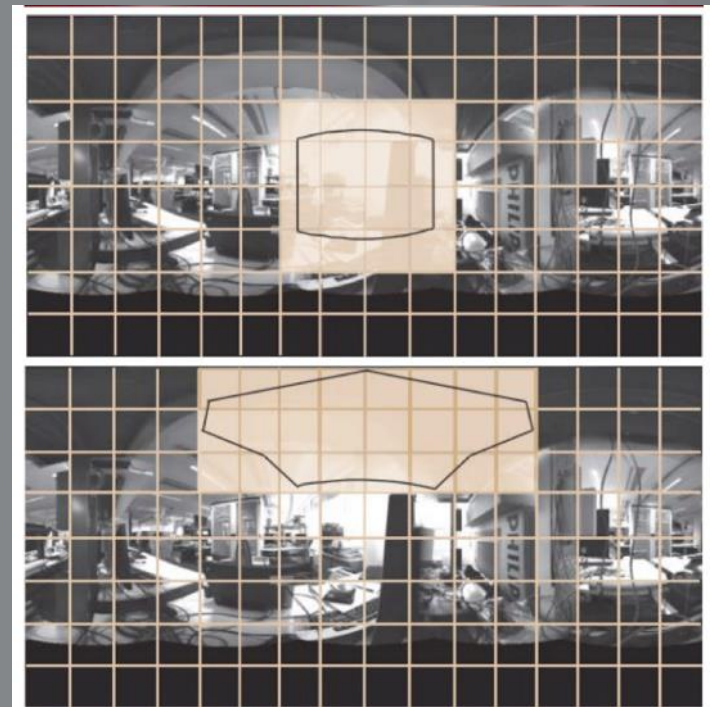
全景视频自适应传输方法

- 局部区域传输，减少数据量
- FOV编码。视窗区域高质量，其他区域低质量
- 分块传输。仅传输FOV区域



Tile-based Viewport Streaming

- 分块传输
 - 视频分为空间上的小分块 (tile), 每块拥有不同的码率
 - 请求主视窗覆盖范围的高码率分块内容
- 优缺点
 - 分块自由组合, 更灵活
 - 编解码复杂 (HEVC)

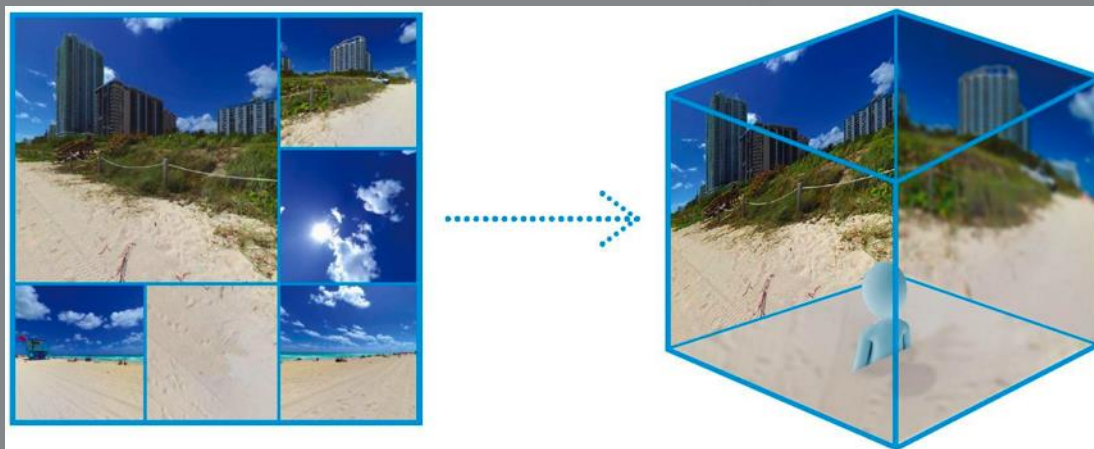


[1] S. Petrangeli, etc, “An http/2-based adaptive streaming framework for 360 virtual reality videos”, ACM MM’17

[2] M. Xiao, etc, “BAS-360: Exploring Spatial and Temporal Adaptability in 360-degree Videos over HTTP/2”, INFOCOM’18

Asymmetric Mapping and Streaming

- 不同区域质量不同(类似ROI编码)
- 传输360完整视频
- 优点
 - 编码效率高
 - 支持H.264
- 缺点
 - 存储空间高



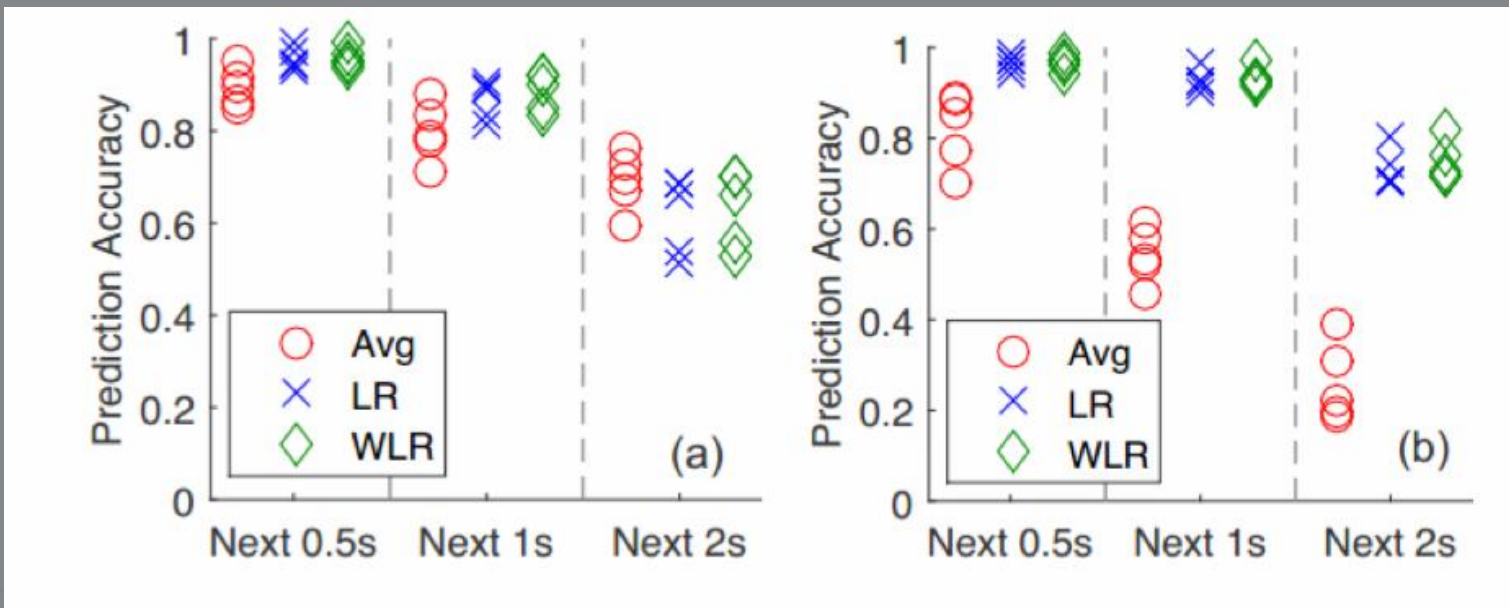
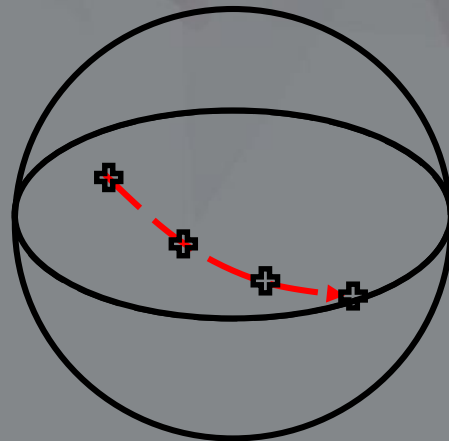
Google EAC

[1] Facebook, transform360[Z]. <https://github.com/facebook/transform360>, 2016.

[2] M. Coban etc. “360 VR Video Truncated Square Pyramid Geometry for OMAF[R]”, Tech. rep., 2016

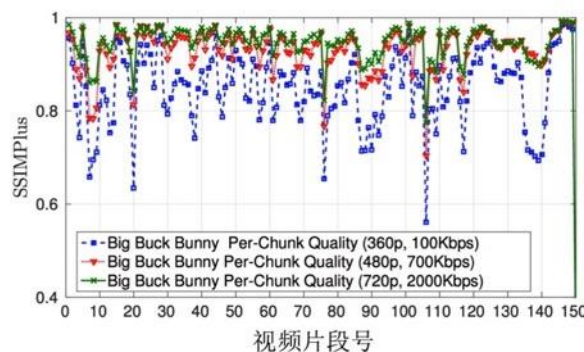
挑战1 - 视点预测

- 头部运动存在随机特性
- 预测误差大
 - 短期预测 (1s) 的准确率较高
 - 长期预测 (>2s) 准确率较低, 仅70%

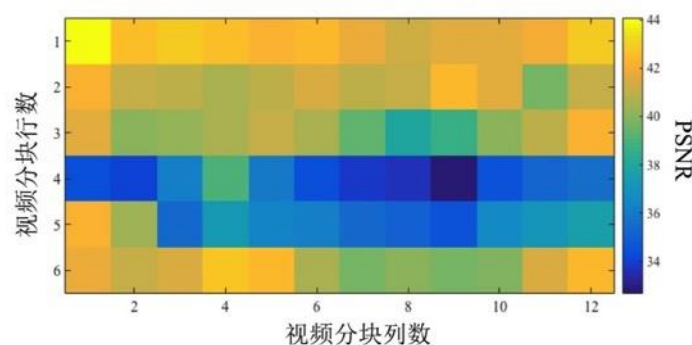


挑战2 – 客观QoE评价方法

- 视频质量 (BitRate, PSNR)
- 卡顿 (Rebuffering/Stalling)
- 边界 (Spatial content continuity)
- 抖动 (Switch frequency)
- 启动延迟 (Startup delay)



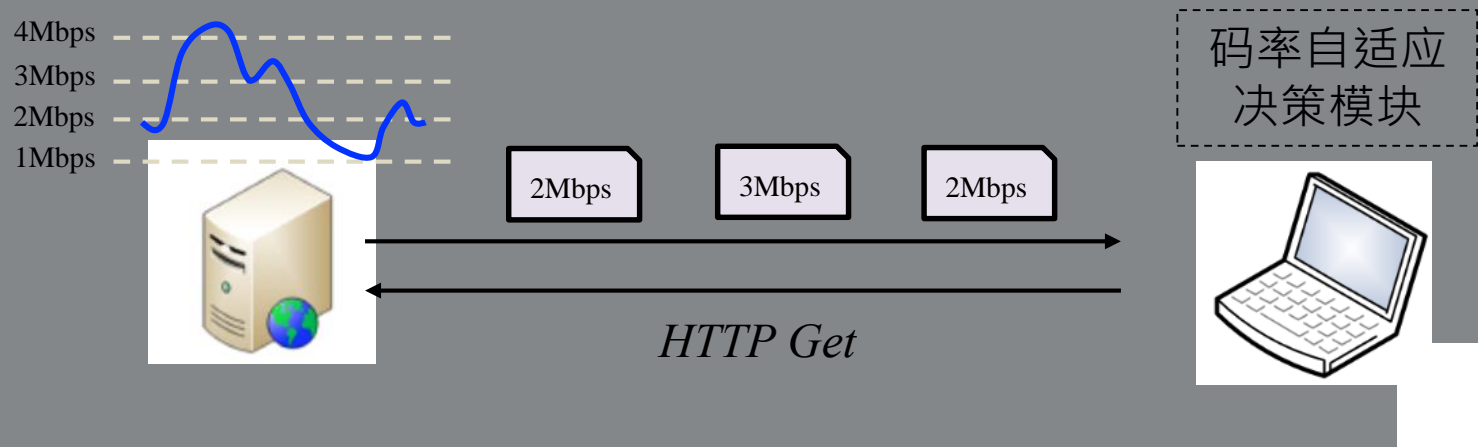
(1) 时间上的质量波动



(2) 空间上的质量波动

挑战3 – 带宽波动

- 互联网带宽波动
- 码率自适应算法



■ 研究内容

- 网络视频质量评价方法
 - 客观评价网络传输对QoE的影响
 - 人眼视觉与传输优化 [SIGCOMM 2019]
- VR视频视点预测
 - Learning-based视点预测 [ACM MM'2019]
- QoE驱动的自适应传输方法
 - 视点自适应传输优化方法 [ACM MM'2018]

基于用户行为的QoE分析

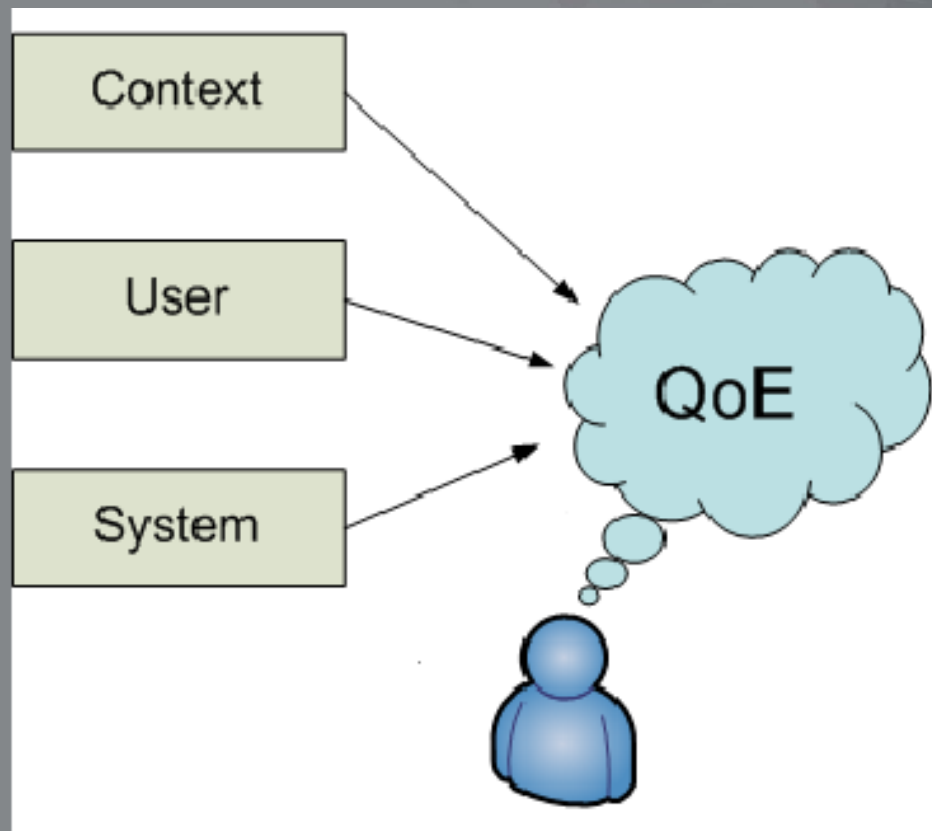
用户体验质量(Quality of Experience, QoE)

- What?
 - Quality of Experience
- How to quantity?
 - 主观
 - Mean Opinion Score (MOS)
 - 客观
 - PSNR
 - Bitrate
 - Play time
 - ...



用户体验质量

- **User**
 - 年龄、心情、性别、学历...
- **System**
 - 网络层
 - 吞吐量、时延、丢包...
 - 应用层
 - 码率、帧率、卡顿...
- **Context**
 - 时间、地点、设备



基于用户行为的视频体验质量分析

➤ Problem Scope

- 基于大规模真实用户行为的QoE模型
- 分析App metrics与QoE的关系

User Factors		Age, education background, emotions, gender...
System Factors	Network	Throughput, delay, loss...
	Application	Bitrate, frame rate, rebuffer...
Context Factors		Device, access, time of day...

基于用户行为的视频体验质量分析

➤ User behaviors

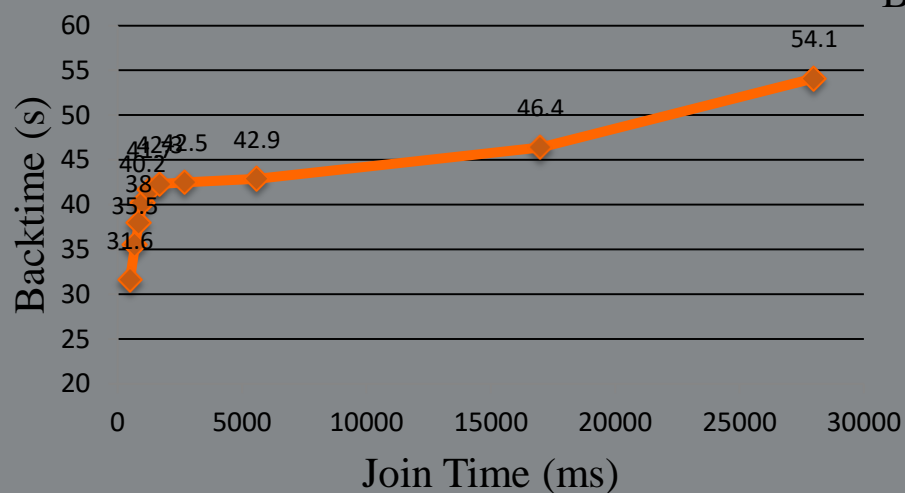
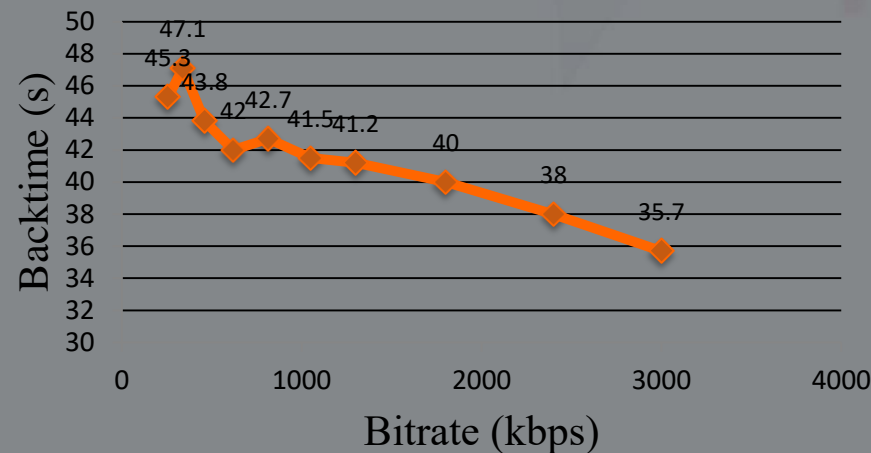
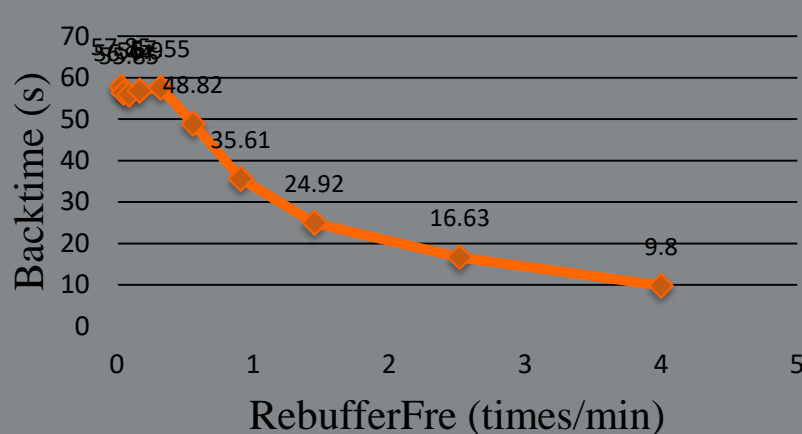
User Activity					
Exit			Seek	Switch	
Start-up	Playing	Rebuffering		Up	Down
QoE Metrics					
Start-up delay	Play time	Rebuffering time	Seek Times	Switch Times	

➤ Feature

Resolution, Bitrate, Rebuffer frequency, Rebuffer ratio, Join time, Seek Cost, Content, Device, Access...

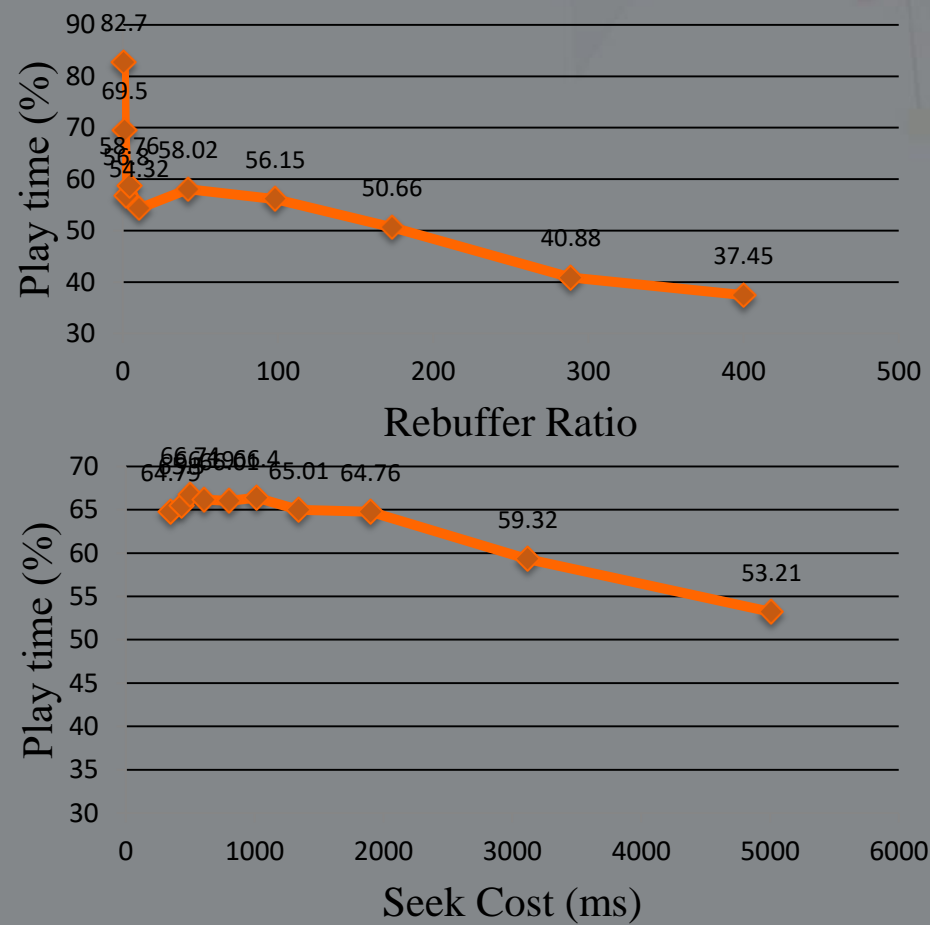
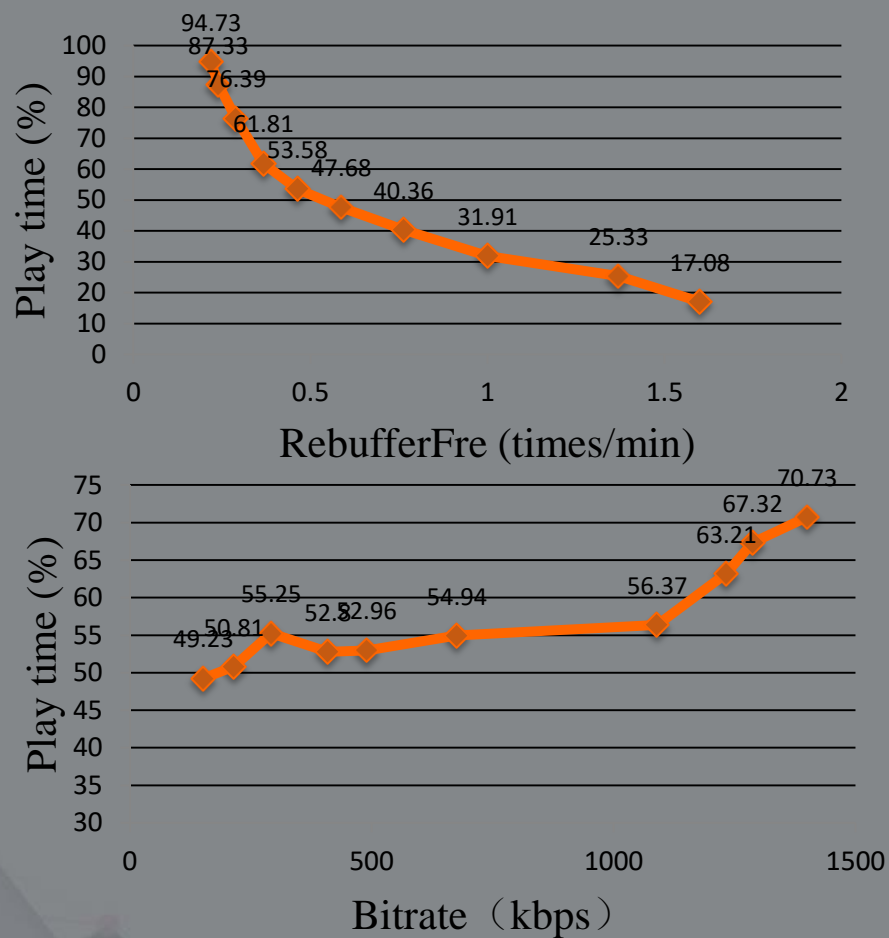
基于用户行为的视频体验质量分析

➤ Rebuffering Time



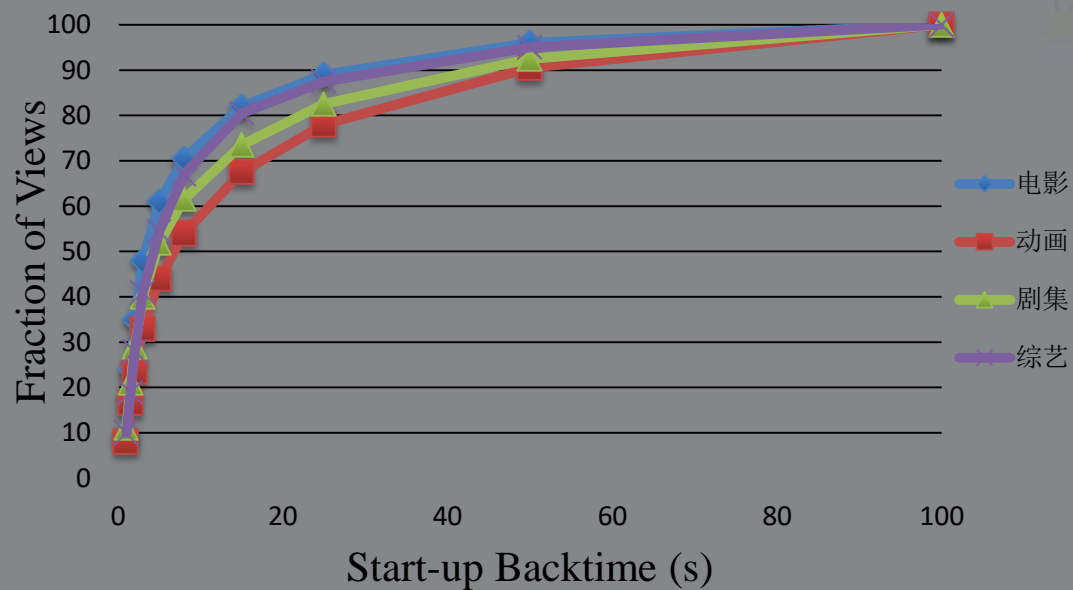
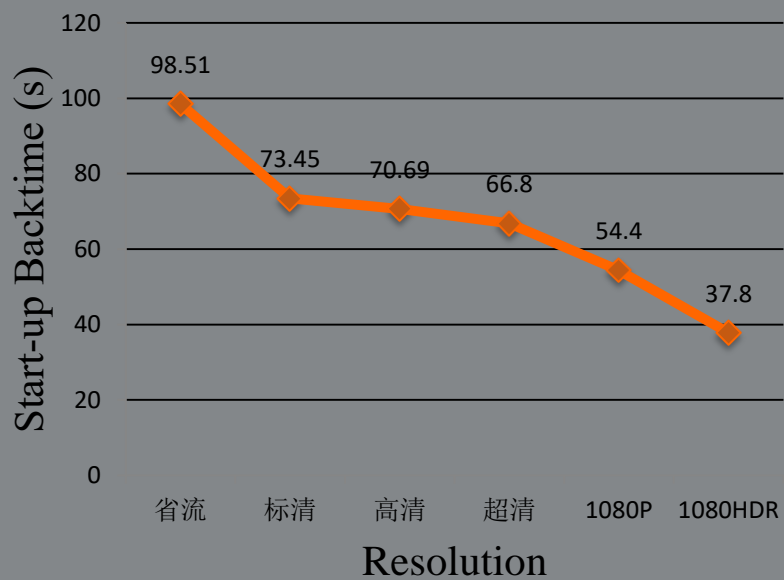
基于用户行为的视频体验质量分析

► Playback Time



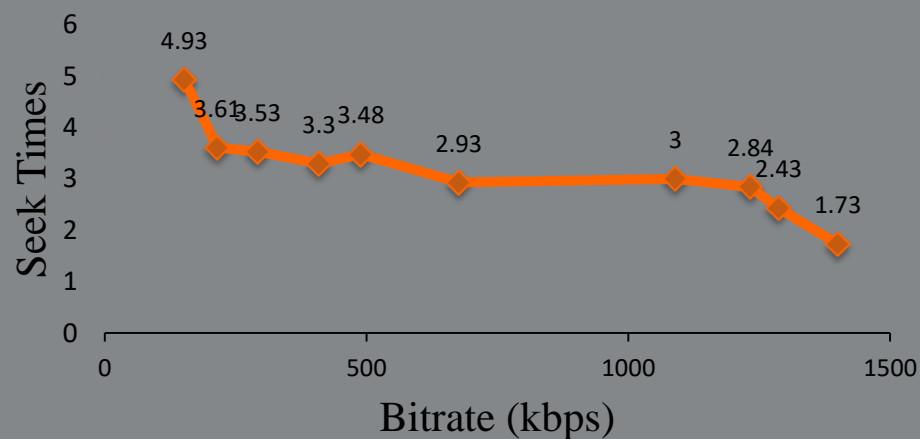
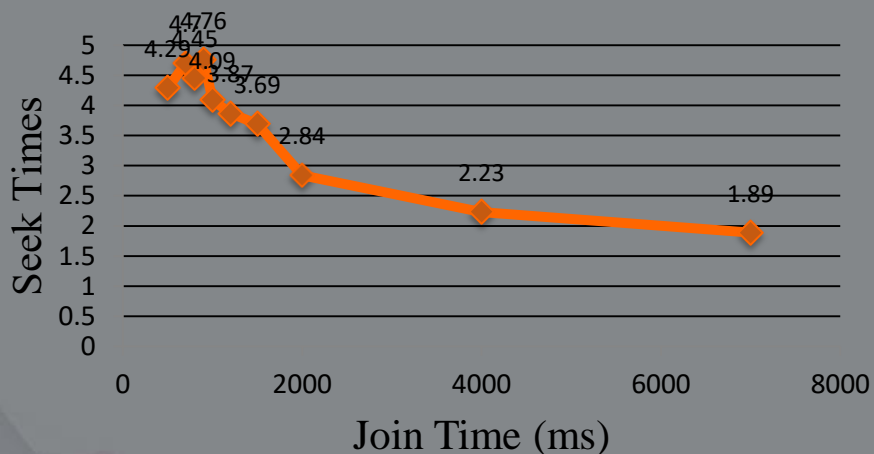
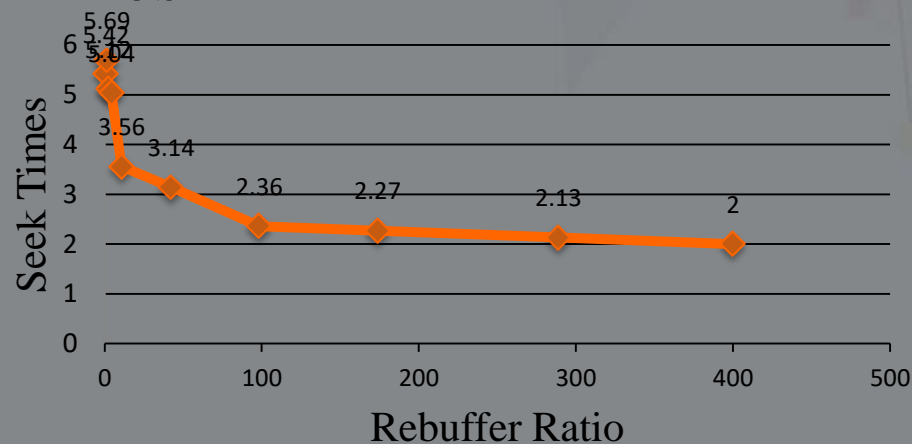
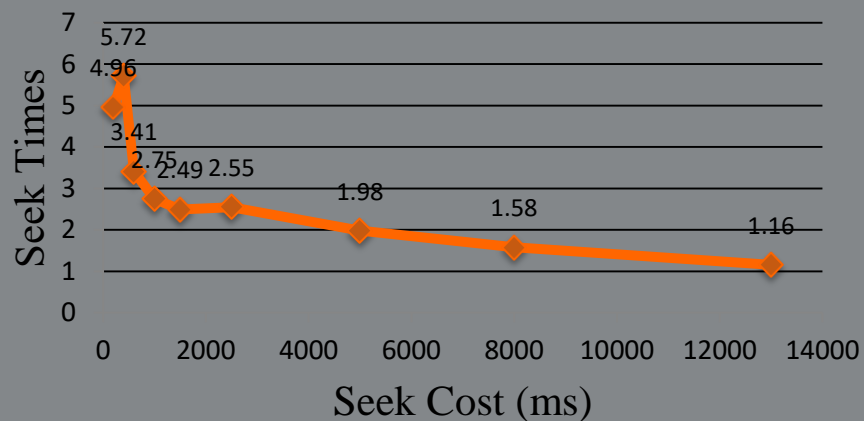
基于用户行为的视频体验质量分析

➤ Start-up Delay



基于用户行为的视频体验质量分析

► Analysis Results of Seek Times



小结

- 如何评价视频传输质量
 - 编码、终端、网络、用户等复杂因素
- 基于用户行为的QoE建模
 - 播放时长与QoE正相关
 - 卡顿时长、启动时延与用户期望相关

[1] Yu Guan, Chengyuan Zheng, Zongming Guo, Xinggong Zhang*, Junchen Jiang, "Pano: Optimizing 360° Video Streaming with a Better Understanding of Quality Perception", To appear in SIGCOMM'2019.

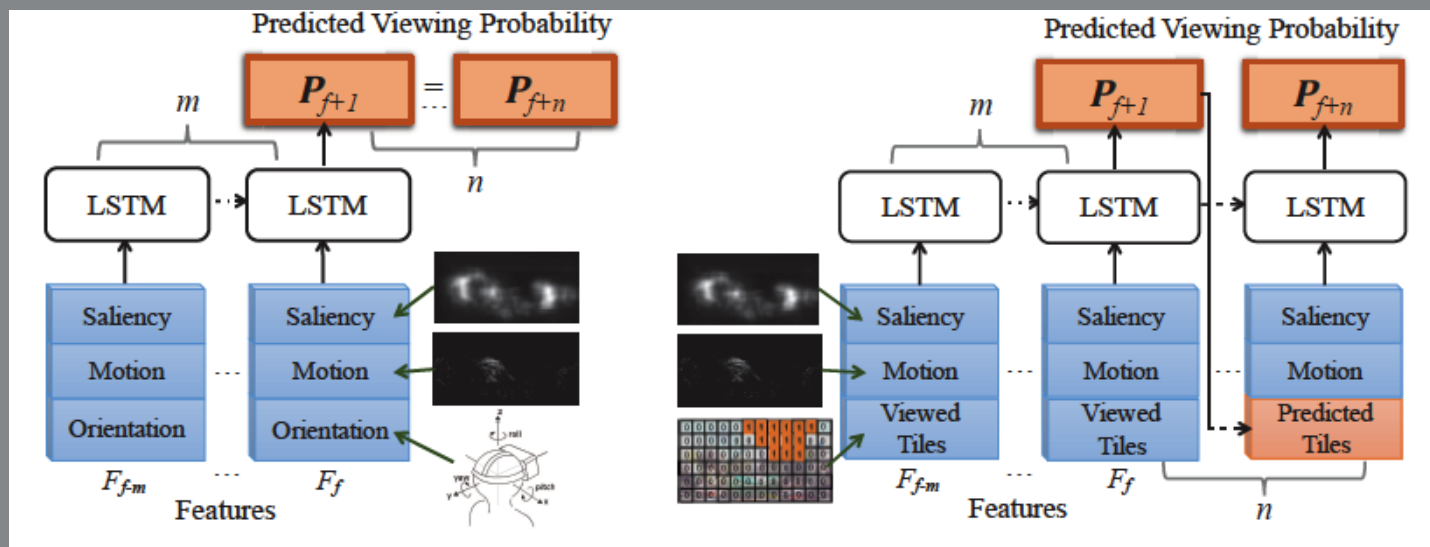
[2] Zhimin Xu, Xinggong Zhang*, Zongming Guo, "QoE-driven Adaptive K-Push for HTTP/2 Live Streaming", in IEEE Transactions on Circuits and Systems for Video Technology(CSVT), vol. 29, no. 6, pp. 1781-1794, June 2019.

跨用户行为学习的VR视频视点预测

- Cross-user Viewport Prediction

跨用户学习的视点预测算法

- 基于用户历史运动信息的视点预测算法
- 基于视频内容的视点预测算法[Fan17]



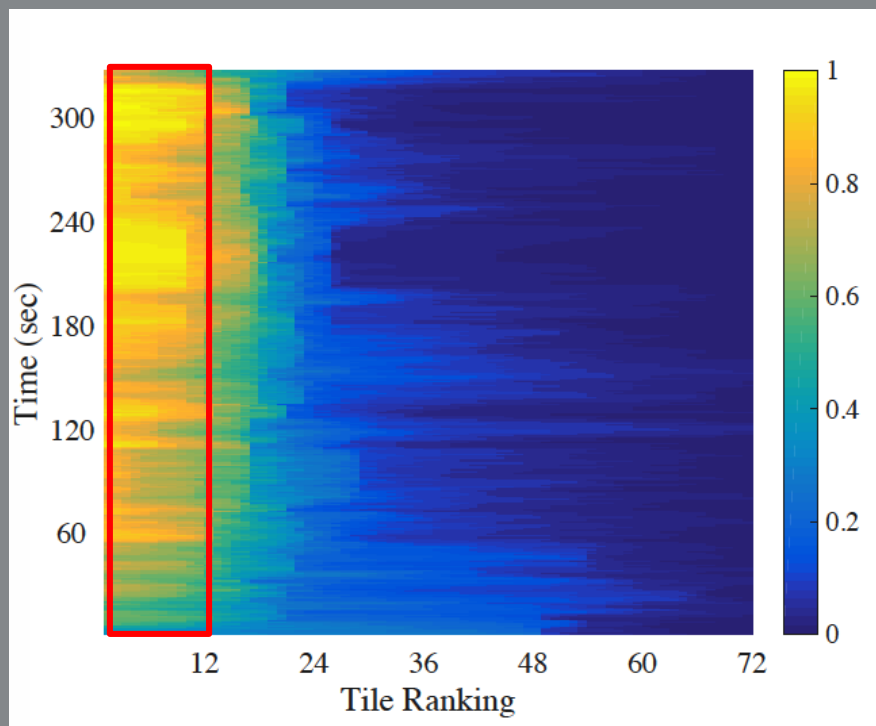
用户行为研究

■ VR用户行为研究

- 在同一帧视频片段，存在明显的感兴趣区域 (ROI)

视频10%的区域被80%
以上的用户观看

热度图：视频分块的用户观看人数比例

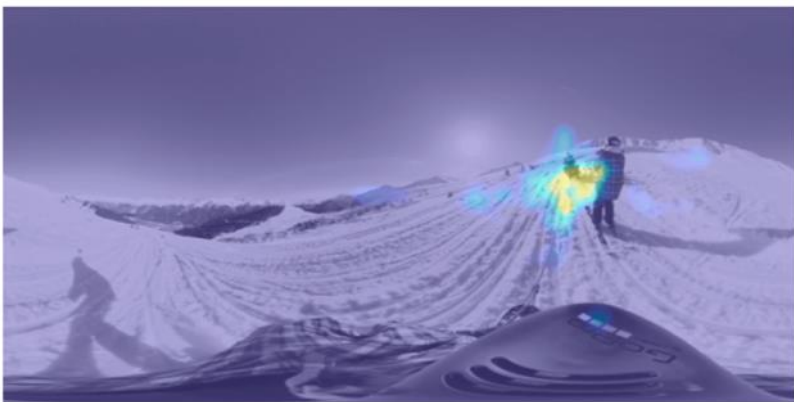


用户行为研究

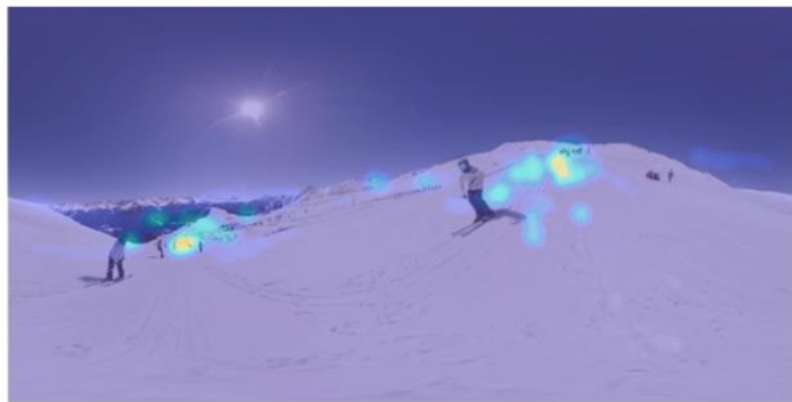
■ 基于跨用户学习的视点预测算法

■ 用户行为研究

- 在同一帧视频片段，存在明显的感兴趣区域（ROI）
- 用户的感兴趣区域可能有多个



包含一个感兴趣区域



包含两个感兴趣区域

跨用户视点学习

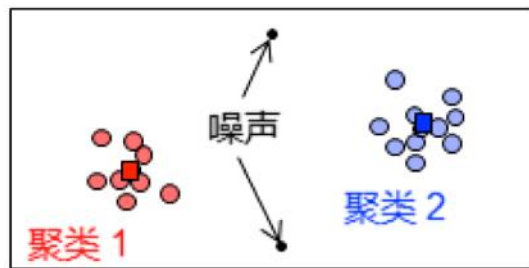
■ 用户聚类 (DBSCAN)

得到每个视频片段中的感兴趣区域，并计算对应的视点概率分布

■ 用户分类 (SVM)

预测用户未来的类别，从而预测观看的ROI

特征：用户历史窗口的视点



在一个视频片段上用户的视点

0.0	0.0	0.0	0.0
0.1	0.1	0.1	0.0
0.1	0.2	0.1	0.0
0.1	0.1	0.1	0.0

通过聚类1中的用户计算出分块观看概率

0.0	0.1	0.1	0.1
0.0	0.1	0.2	0.1
0.0	0.1	0.1	0.1
0.0	0.0	0.0	0.0

通过聚类2中的用户计算出分块观看概率

系统框架

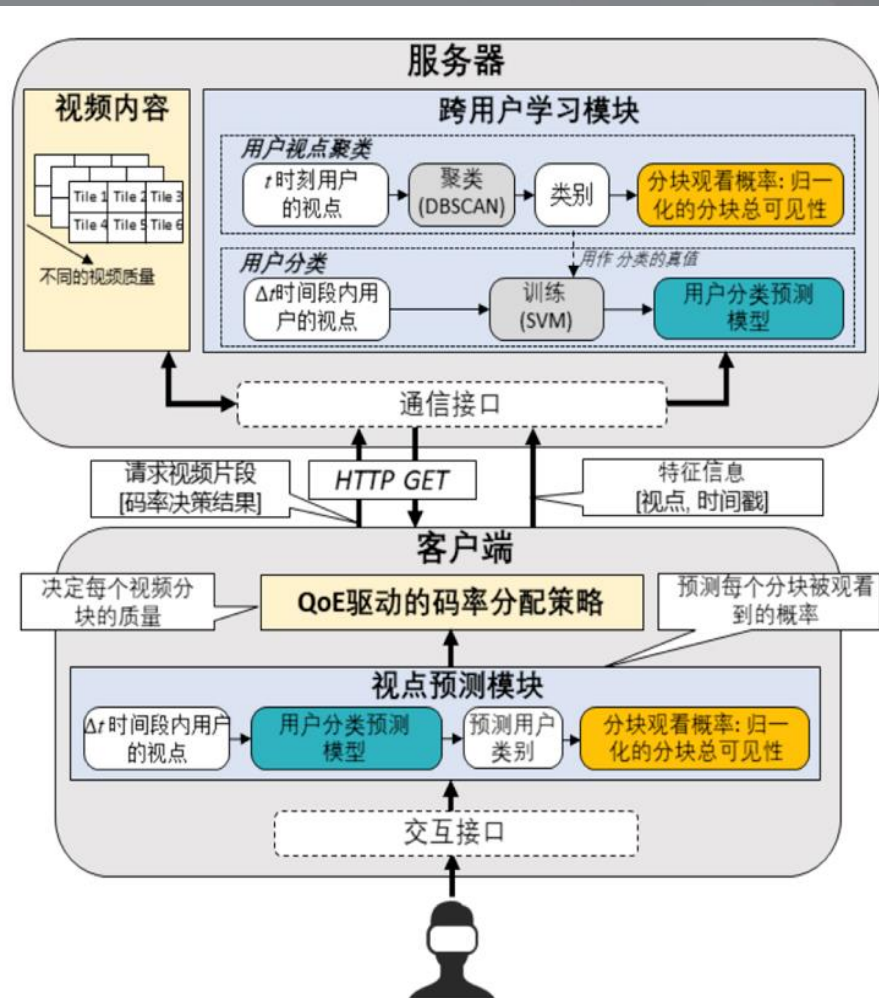
■ 基于跨用户学习的系统框架

■ 服务器

- 处理视频内容
- 跨用户学习模块

■ 客户端

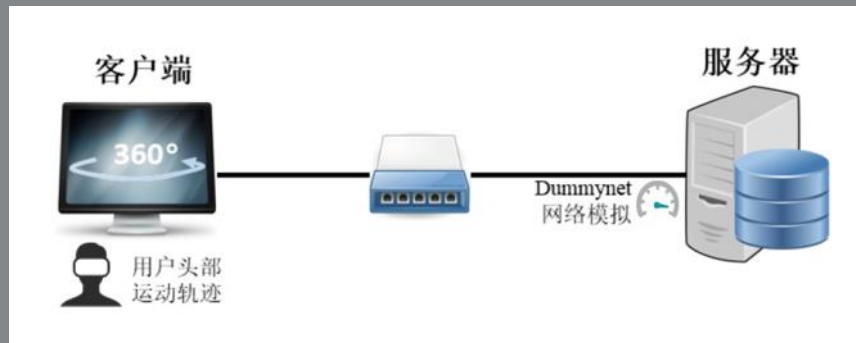
- 视点预测
- QoE驱动的码率分配策略



实验验证

■ 实验设置

- 数据集^[Wu17]: 3个全景视频, 每个视频有48个用户行为数据 (43/5)
- 视频分块: 6 x 12
- 编码码率: 使用CQP编码为5个版本
- 对比算法:
 - 基于线性回归的预测算法 (Tile-LR)
 - 基于视窗概率模型的预测算法 (360ProbDASH)
 - 本文算法, 无用户分类(CLS-1); 有用户分类 (CLS-2)
- 网络拓扑:



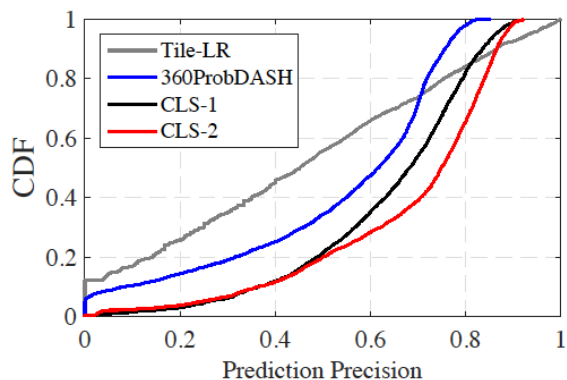
实验结果

■ 以分块为粒度的视点预测准确率

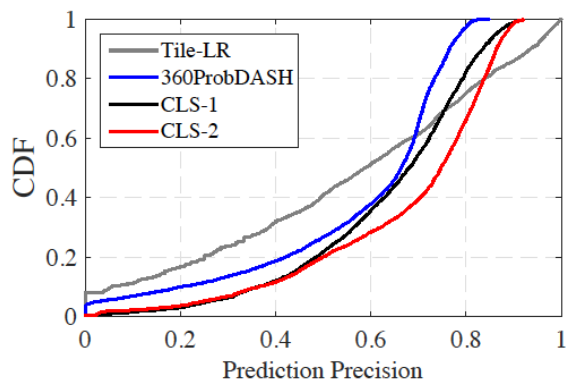
每个分块的预测率 = $\min\{\text{预测概率}, \text{归一化的真实值}\}$

视点预测准确率 = 每个分块的预测率之和

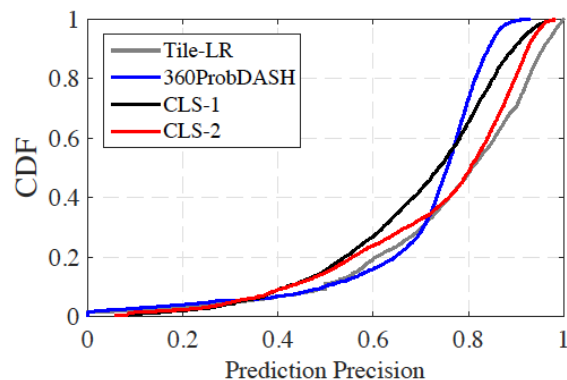
预测准确率提高26~40%以上



(a) 预测未来5秒钟的用户视点



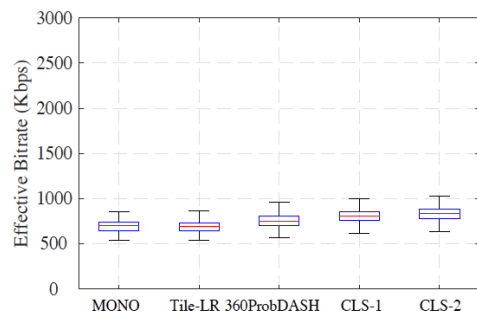
(b) 预测未来3秒钟的用户视点



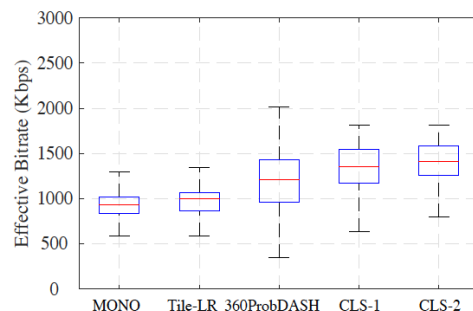
(c) 预测未来1秒钟的用户视点

实验结果

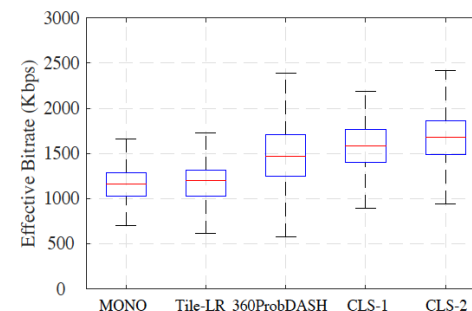
有效码率 ($eRate$) – 提升50%~100%



(a) 带宽= 2000Kbps

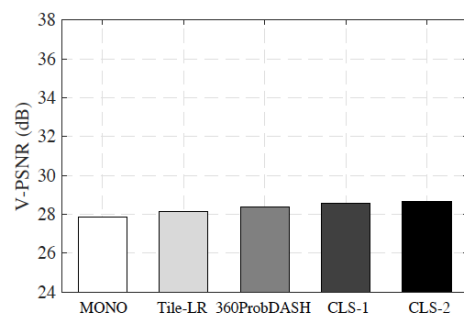


(b) 带宽= 3000Kbps

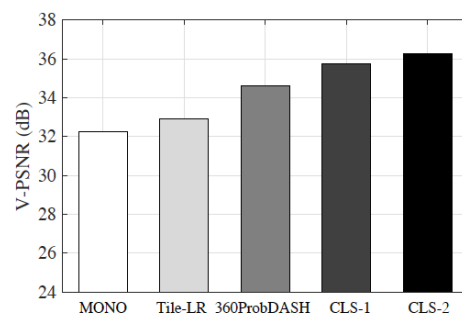


(c) 带宽= 4000Kbps

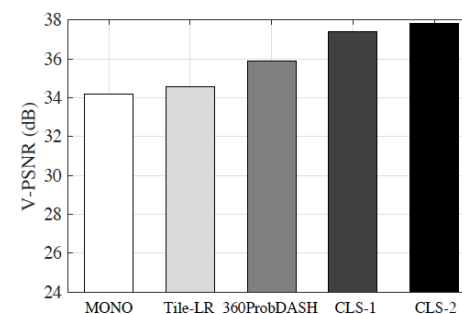
V-PSNR – 提升2~4dB



(a) 带宽= 2000Kbps



(b) 带宽= 3000Kbps



(c) 带宽= 4000Kbps

实验结果

■ 真实网络环境下性能比较

- 有效码率提升25%~50%
- 质量提升2~3dB

方法	视频1			视频2			视频3		
	sRate	eRate	V-PSNR	sRate	eRate	V-PSNR	sRate	eRate	V-PSNR
MONO	2418.1	808.2	27.16	3458.1	1015.5	25.54	3312.5	1072.5	28.74
Tile-LR	2442.5	868.2	27.44	3513.9	1035.6	25.69	3306.6	1099.4	28.85
360ProbDASH	2333.1	958.1	28.12	3311.7	1177.9	25.86	3215.3	1186.1	29.39
CLS-1	2431.8	1044.3	28.40	3465.4	1396.7	26.54	3315.8	1360.3	30.61
CLS-2	2435.5	1174.3	29.10	3473.1	1468.3	27.09	3319.6	1443.9	31.12

小结

- VR视频视点预测三大方法
 - 视频内容、运动趋势、跨用户行为
- 跨用户行为学习
 - 提高视点预测精确度26~40%
 - 节省码率25~80%以上

[1] Lan Xie, Xinggong Zhang, Zongming Guo. “CLS: A Cross-user Learning based System for Improving QoE in 360-degree Video Adaptive Streaming”, ACM MM, 2018.

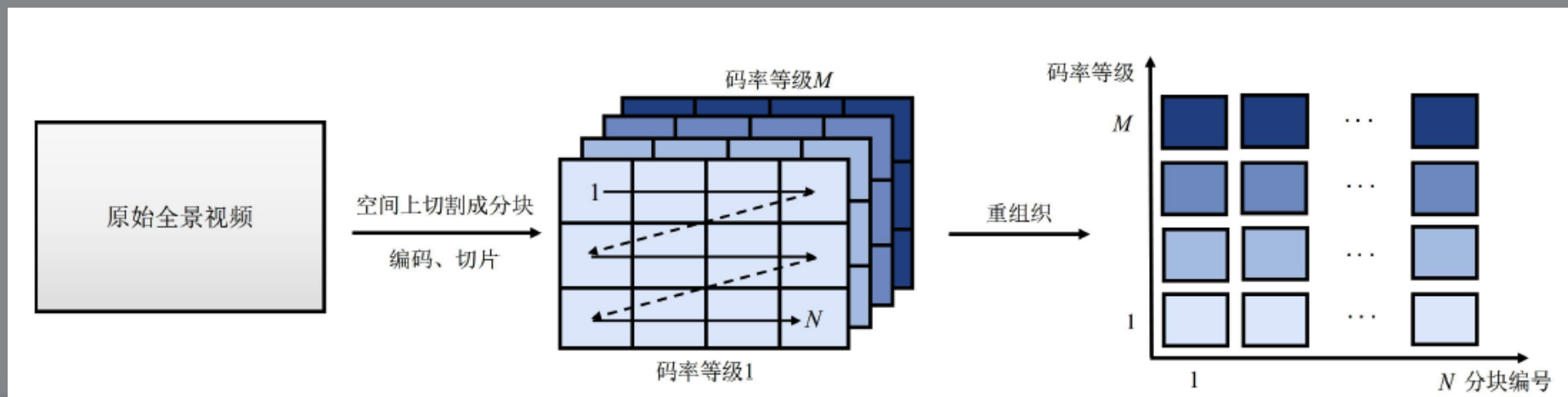
[2] Yixuan Ban, Lan Xie, Zhimin Xu, Xinggong Zhang and Zongming Guo, “CUB360: Exploiting Cross-Users Behaviors for Viewport Prediction in 360 Video Adaptive Streaming”, ICME’2018.

QoE驱动视点自适应流传输

分块传输优化框架

■ 基于视频分块传输的视点自适应最优化框架

- 视频分块、分段
- 切分为多个码率
- Client根据MPD选择合适码率、分块



QoE 目标函数

- 以QoE为目标，建立最优化问题，求解视频分块码率
 - 建立用户视窗概率模型，提高视点预测准确率
 - 最小化视频失真的期望，提高视频质量
 - 最小化视频空间质量抖动的期望，减少质量抖动

$$\min_{\mathbf{X}} \Phi(\mathbf{X}) + \eta \cdot \Psi(\mathbf{X})$$

$$\text{s.t.} \quad \sum_{i=1}^N \sum_{j=1}^M x_{i,j} \cdot r_{i,j} \leq R,$$

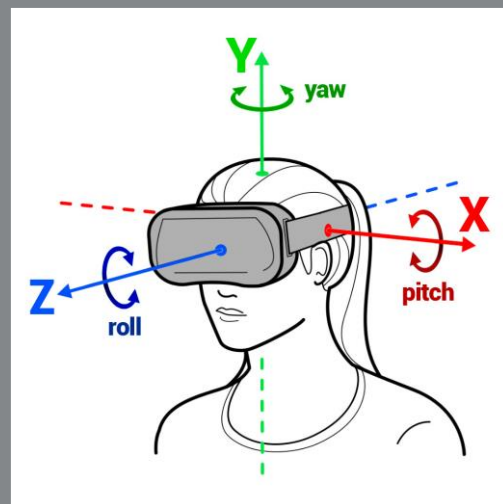
$$\sum_{j=1}^M x_{i,j} \leq 1, \quad x_{i,j} \in \{0, 1\}, \quad \forall i.$$

概率预测

- 目标
 - 提升头部运动预测准确性
- 预测方法
 - 线性回归
 - Yaw α , Pitch β , Roll γ
 - 预测 Δt 之后的欧拉角

$$\begin{cases} \hat{\alpha}(t_0 + \Delta t) = v_{\alpha} \cdot \Delta t + \alpha(t_0), \\ \hat{\beta}(t_0 + \Delta t) = v_{\beta} \cdot \Delta t + \beta(t_0), \\ \hat{\gamma}(t_0 + \Delta t) = v_{\gamma} \cdot \Delta t + \gamma(t_0). \end{cases}$$

$$\hat{J}(x^0 + \nabla x) = \nabla J \cdot \nabla x + J(x^0).$$

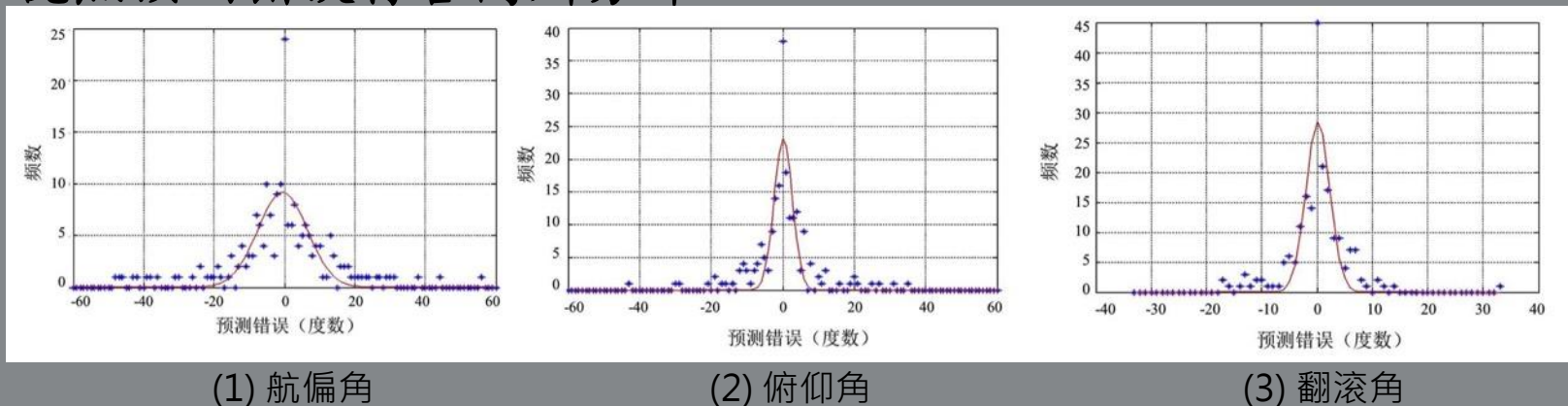


预测误差

■ 用户视窗概率模型

考虑视点预测错误概率，对视窗分布进行估计，从而计算视频分块被观看的概率

■ 视点预测错误符合高斯分布



■ 以航偏角为例，预测值 α 是真实值的概率为

$$P_{\text{yaw}}(\alpha) = \frac{1}{\sigma_{\alpha}\sqrt{2\pi}} \exp\left(-\frac{(\alpha - (\hat{\alpha} + \mu_{\alpha}))^2}{2\sigma_{\alpha}^2}\right)$$

Viewport变换

- 扩展到任意旋转角度，正确的概率

$$P_E(\alpha, \beta, \gamma) = P_{\text{yaw}}(\alpha)P_{\text{pitch}}(\beta)P_{\text{roll}}(\gamma)$$

- 球面上任意一点被观看到的概率

$$P_s(\varphi, \theta) = \frac{1}{|\mathbf{L}(\varphi, \theta)|} \sum_{(\alpha, \beta, \gamma) \in L(\varphi, \theta)} P_E(\alpha, \beta, \gamma)$$

包含点 (φ, θ) 的视窗集合

- 分块被观看到的概率

$$p_i = \frac{1}{|\mathbf{U}_i|} \sum_{(\varphi, \theta) \in \mathbf{U}_i} P_s(\varphi, \theta)$$

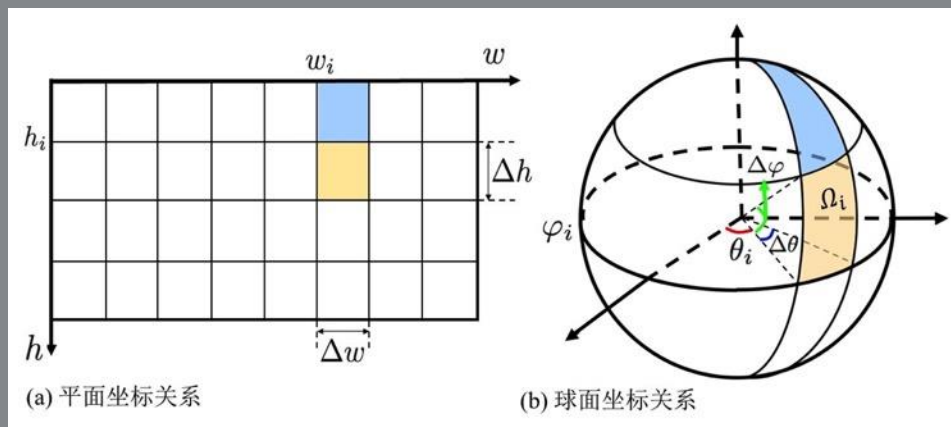
分块覆盖的球面点集

球面质量计算

■ 视频分块的质量失真计算

- 平面质量失真 \neq 球面质量失真
- 分块编码后MSE: d
- 根据球面积分计算分块球面面积

$$\begin{aligned}s_i &= \iint_{\Omega_i} \mathcal{R} d\varphi R \cos \varphi d\theta \\ &= \Delta\theta \mathcal{R}^2 [\sin(\varphi_i + \Delta\varphi) - \sin \varphi_i]\end{aligned}$$



- 分块的球面失真 = 分块编码后MSE ($d_{i,j}$) x 分块对应的球面面积 (s_i)

期望失真

■ 视频失真的期望

$$\Phi(\mathbf{X}) = \frac{\sum_{i=1}^N \sum_{j=1}^M D_{i,j} \cdot x_{i,j} \cdot p_i}{\sum_{i=1}^N \sum_{j=1}^M x_{i,j} \cdot s_i}$$

■ 视频空间质量抖动的期望

$$\Psi(\mathbf{X}) = \frac{\sum_{i=1}^N \sum_{j=1}^M x_{i,j} \cdot p_i \cdot (D_{i,j} - s_i \cdot \Phi(\mathbf{X}))^2}{\sum_{i=1}^N \sum_{j=1}^M x_{i,j} \cdot s_i}$$

实验结果

■ 实验设置

- 数据集：5个用户，1个全景视频
- 视频分块：6 x 12
- 评价指标：V-PSNR，空间质量抖动，黑窗百分比
- 对比算法
 - 全传输（ERP）
 - 分块传输，不进行视点预测（Tile）
 - 分块传输，视点预测使用线性回归（Tile-LR）

实验结果

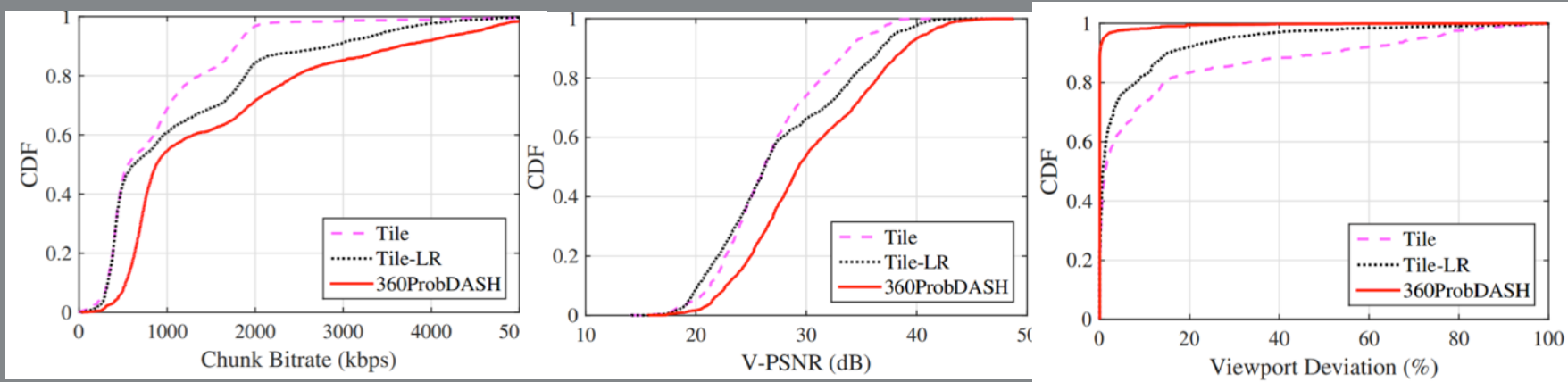
■ 实验结果（固定带宽2Mbps）

Metrics		ERP	Tile	Tile-LR	360ProbDASH
V-PSNR	Value (dB)	24.45	31.16	32.38	34.06
	Relative Imp.	+39.3%	+9.31%	+5.19%	-
空间质量 抖动	Value (CV)	1.05	1.89	1.81	0.97
	Relative Imp.	+8.25%	+48.67%	+46.41%	-
视窗 偏离	Value (%)	0	3.62	0.96	0.13
	Absolute Imp.	-0.13%	+3.49%	+0.83%	-

实验结果

■ 实验结果（真实网络环境）

- 服务器在香港，客户端在北京，通过互联网连接
- 每轮顺序执行三个分块传输算法，共10轮，每两轮使用同一个头部运动轨迹（共5个头部运动轨迹）
- 实验结果



Future Works

- **16K/VR在5G网络传输的质量评价**
 - 链路、网络、用户对视频质量影响
 - 端、网、云的测量和预测方法
- **Perceive Quality based Streaming**
 - VR视频的人眼视觉特性
 - 质量感知模型对传输的影响



北京
2019

遨游“视”界 做你所想
Explore World, Do What You Want



Thank you



张行功

北京大学NetVideo Group

<http://www.icst.pku.edu.cn/netvideo>

北京大学计算机科学技术研究所
Institute of Computer Science & Technology of Peking University

CSDN