Viewport Adaptation-Based Immersive Video Streaming: Perceptual Modeling and Applications

Shaowei Xie, Qiu Shen, Yiling Xu, Qiaojian Qian,
Shaowei Wang, Zhan Ma, and Wenjun Zhang

arXiv:1802.06057, Submitted to ISCAS'18

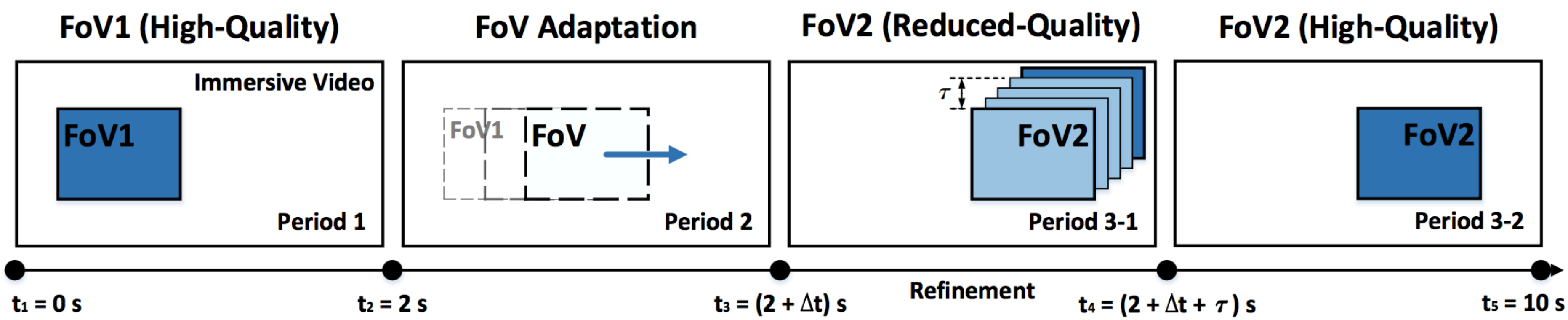
Introduction

- Netflix suggests a 5 Mbps connection speed for the broadcasting quality of a typical FHD (1080p) video at 30 fps
- How about an immersive video at 32K×16K, 120 fps, and 25 depth levels? 10Gbps <- unreliable
- Sol: apply the adaptive viewport streaming instead of delivering the bulky immersive video entirely

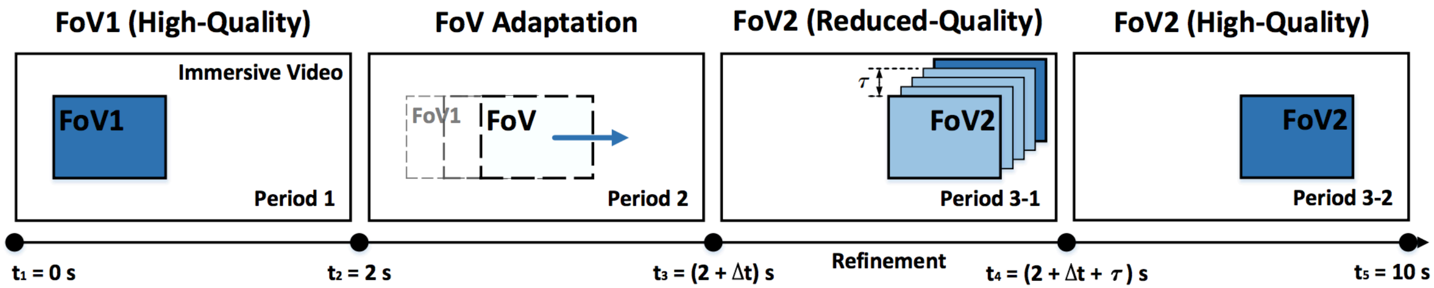
Adaptive Viewport Streaming

- Strategy:
  - Content within current FoV at the highest quality
  - Content outside the current FoV at the reduced quality

  ⇒ avoid the blackout caused by switching the FoV suddenly



Goal



- Model the perceptual quality using the Mean Opinion Score (MOS)

⇒quantify the perceptual impact of the quality variations between consecutive FoVs

  - quantization stepsize q ($q = 2^{\frac{QP-4}{6}}$)
  - spatial resolution s
  - refinement duration τ

- Devise the model to guide the bandwidth constrained immersive video streaming
  - maximizing the subjective quality under the rate constraint

Considered Videos

- 5 from JVET test sequences
- 4 from Youtube


(a) KiteFlite*†


(b) AerialCity*

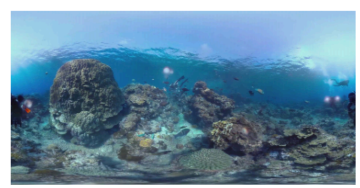

(c) Gaslamp*

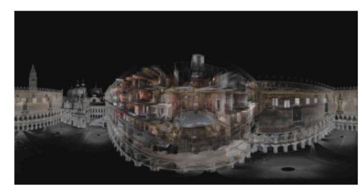

(d) Harbor*


(e) Trolley*

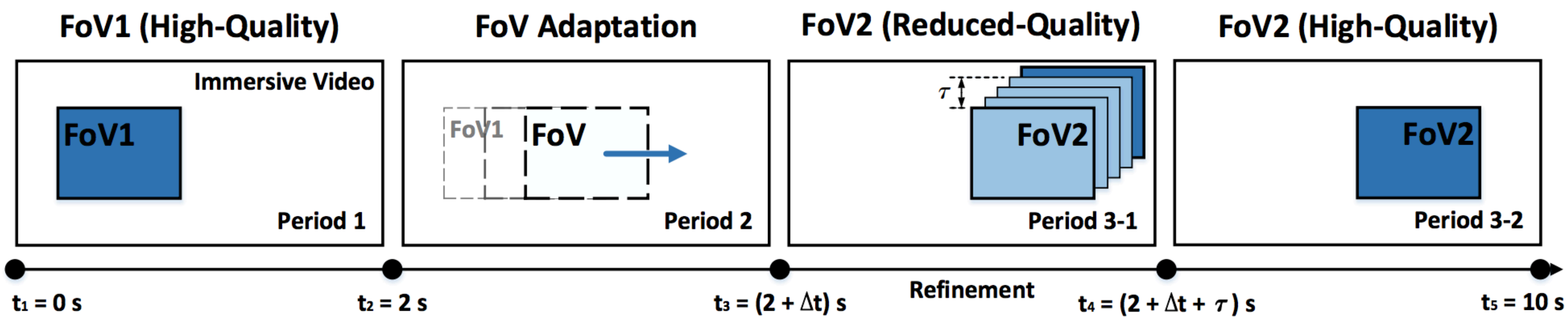

(f) Elephants

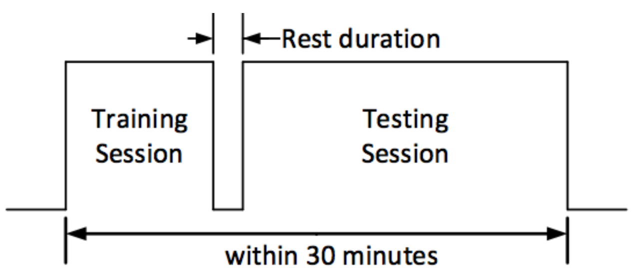

(g) Rhinos


(h) Diving

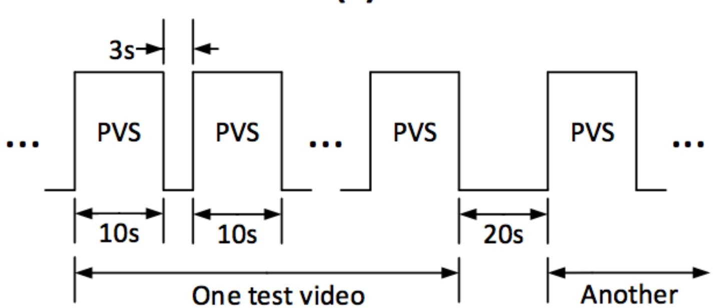

(i) Venice

Test Procedure

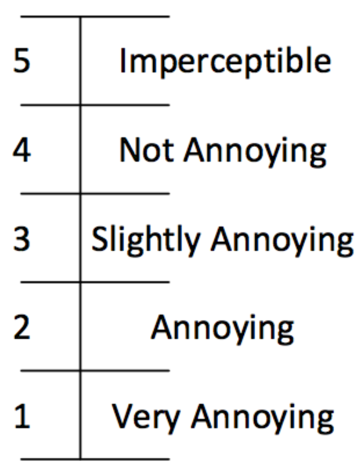




(a)



(b)

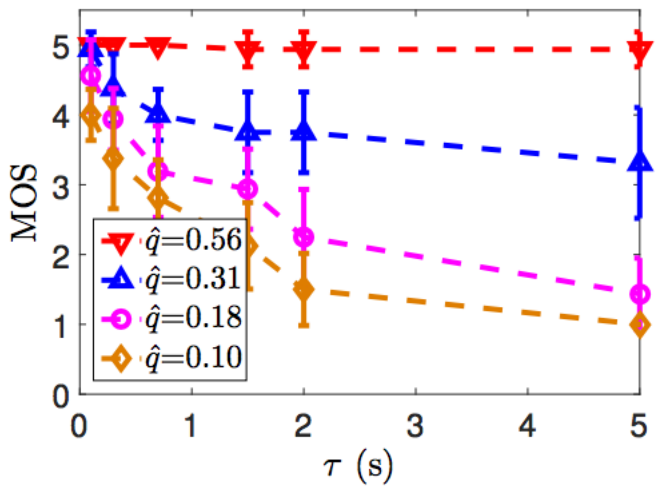


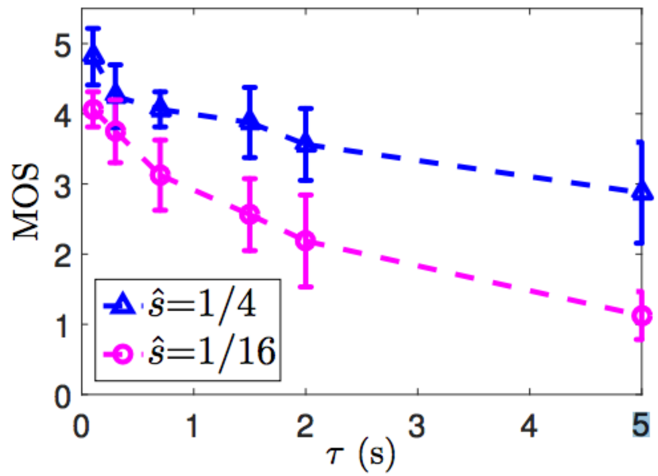
(c)

- Mainly crop and edit FoV sequences from the original immersive video to emulate the FoV adaptation
    - 5 QPs: 22, 27, 32, 37, 42
    - 3 resolutions: naive, 1/4, 1/6
    - 6 refinement durations τ: 0.1, 0.3, 0.7, 1.5, 2, 5 secs

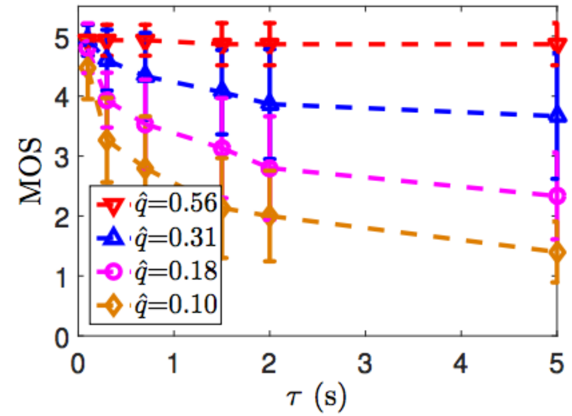
Versus Refinement time

$$z_{mij} = \frac{x_{mij} - \mu(X_i)}{\sigma(X_i)}$$





Analytical Models



- Least squared error

$$\hat{Q} = \frac{Q}{Q_{\max}} = a \cdot e^{-b \cdot \tau} + c$$

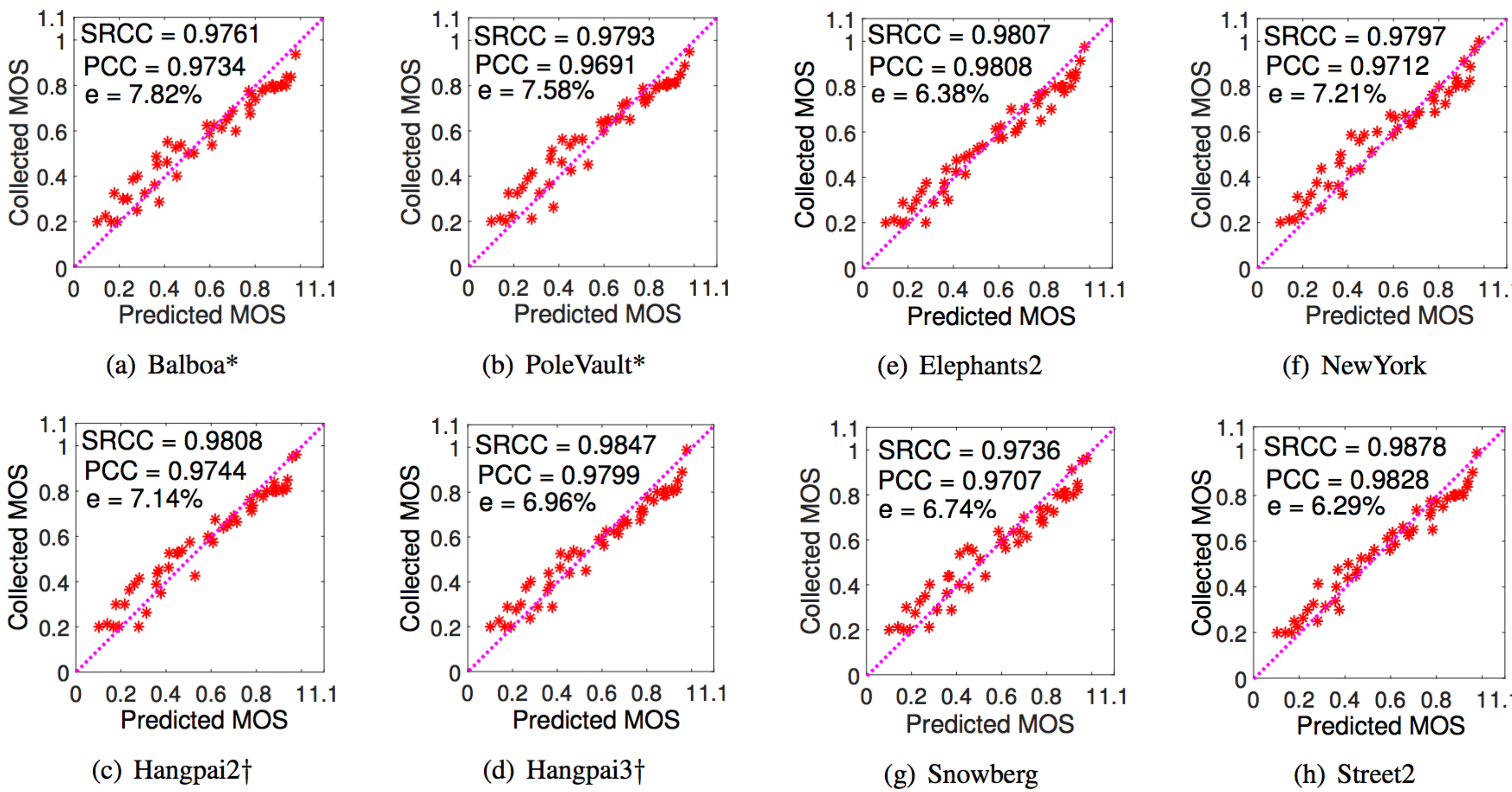
$$Q(\tau, \hat{q}, \hat{s}) = Q_{\max} \cdot \hat{Q}_{\mathrm{NQQ}}(\tau, \hat{q}) \cdot \hat{Q}_{\mathrm{NQS}}(\tau, \hat{s}),$$

where

$$\hat{Q}_{\mathrm{NQQ}}(\tau, \hat{q}) = a(\hat{q}) \cdot e^{-b(\hat{q}) \cdot \tau} + (1 - a(\hat{q})),$$
$$\hat{Q}_{\mathrm{NQS}}(\tau, \hat{s}) = a(\hat{s}) \cdot e^{-b(\hat{s}) \cdot \tau} + (1 - a(\hat{s})).$$

Model Cross-Validation

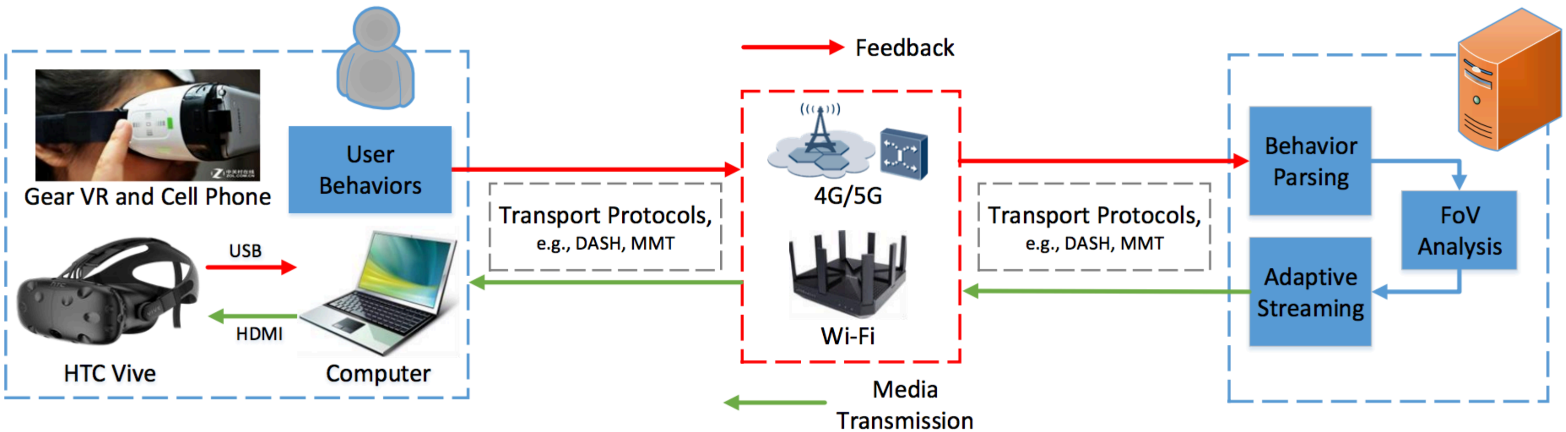
- 79 subjects, each watch one or two test videos
  - Pearson correlation coefficient (PCC) and Spearman's rank correlation coefficient (SRCC) close to 0.98



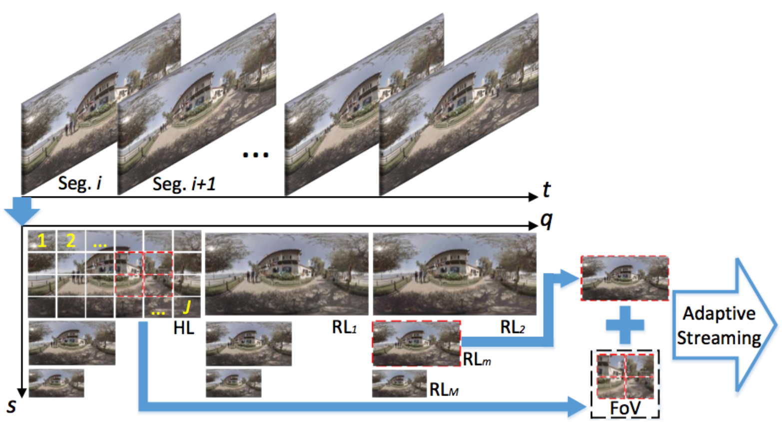
(a) Balboa* (b) PoleVault* (e) Elephants2 (f) NewYork

(c) Hangpai2† (d) Hangpai3† (g) Snowberg (h) Street2

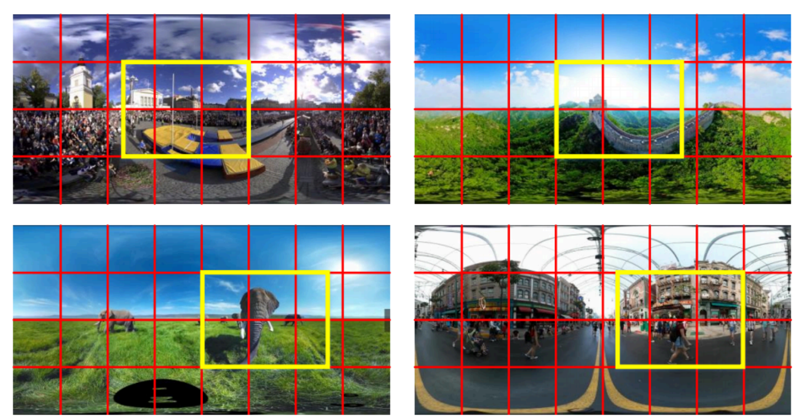
Quality-Bandwidth Optimized Streaming



(a)



(b)



(c)

Problem Formulation

$$\max_{\tau,\hat{q},\hat{s}} \quad Q, \tag{8}$$

$$s.t. \quad R_i^{\mathrm{FoV}} + R_{mi}^{\mathrm{RL}} \leq B, \tag{9}$$

$$0 < \hat{q}, \hat{s} \leq 1. \tag{10}$$

Thereinto,

$$\tau = \frac{R_i^{\mathrm{FoV}} + R_{mi}^{\mathrm{RL}}}{B} \cdot T, \tag{11}$$

$$R_i^{\mathrm{FoV}} = \sum\nolimits_{j=1}^{n} R_{ij}^{\mathrm{HL}}, \tag{12}$$

$$R_{mi}^{\mathrm{RL}} = R(\hat{q}, \hat{s}). \tag{13}$$

Optimal Solution Under Continuous q

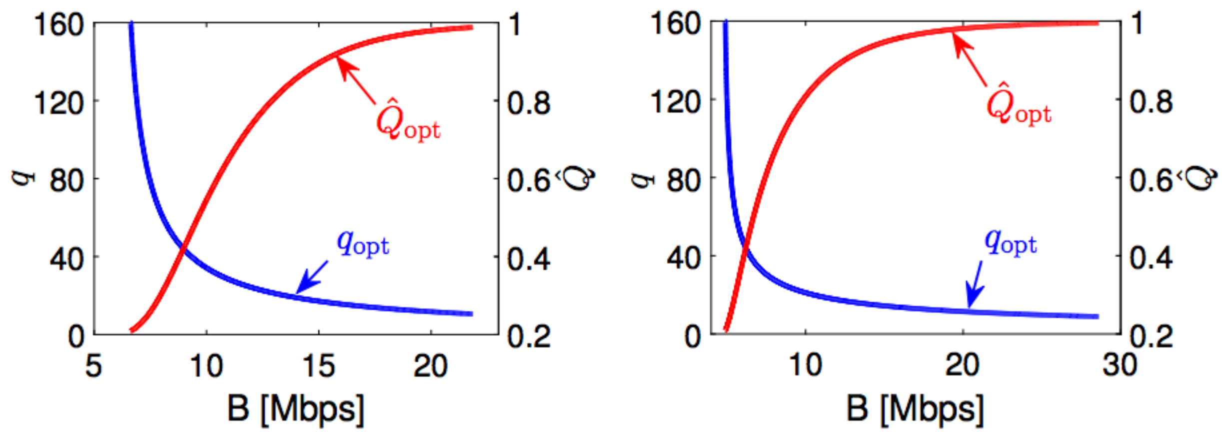
- numerically determine the optimal quantization stepsize q_{opt} and the corresponding normalized maximum perceptual quality Q_{opt} using (15).

$$R_{mi}^{\mathrm{RL}}(\hat{q}) = R_{\max} \cdot \hat{q}^{\alpha}, \tag{14}$$

$$\max_{\hat{q}} \quad \hat{Q} = a(\hat{q}) \cdot e^{-b(\hat{q}) \cdot \frac{(R_i^{\mathrm{FoV}} + R_{\max} \cdot \hat{q}^{\alpha}) \cdot T}{B}} + 1 - a(\hat{q}), \tag{15}$$

$$s.t. \quad R_i^{\mathrm{FoV}} + R_{\max} \cdot \hat{q}^{\alpha} \leq B, \qquad \hat{q} = q_{\min}/q \tag{16}$$

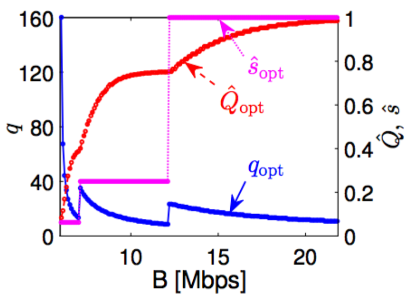
$$0.05 \leq \hat{q} \leq 1. \tag{17}$$



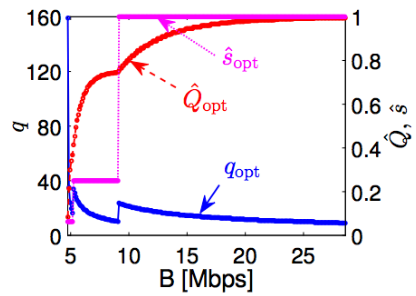
(a) Balboa* (b) PoleVault*

Optimal Solution Under Discrete s and Continuous q

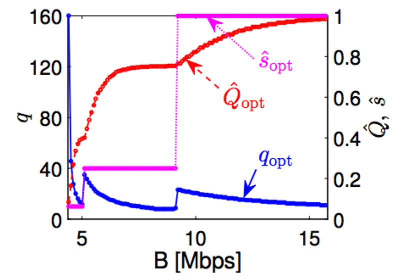
$$R_{mi}^{\mathrm{RL}}(\hat{q}, \hat{s}) = R_{\max} \cdot \hat{q}^{\alpha} \cdot \hat{s}^{\beta}. \qquad (18)$$



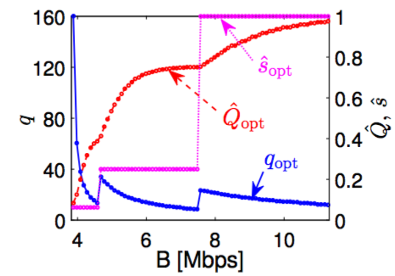
(a) Balboa*



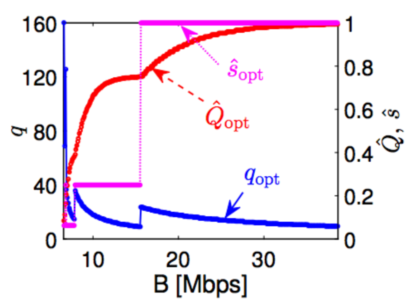
(b) PoleVault*

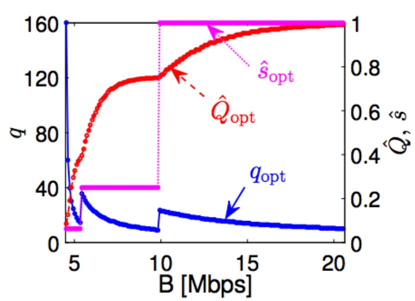


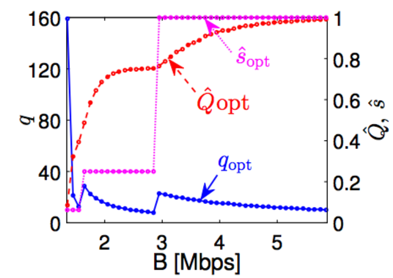
(e) Elephants2



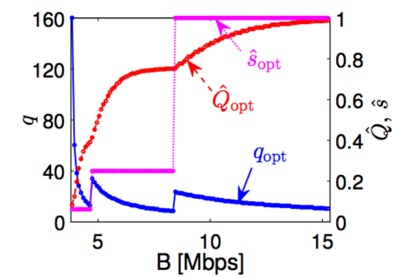
(f) NewYork







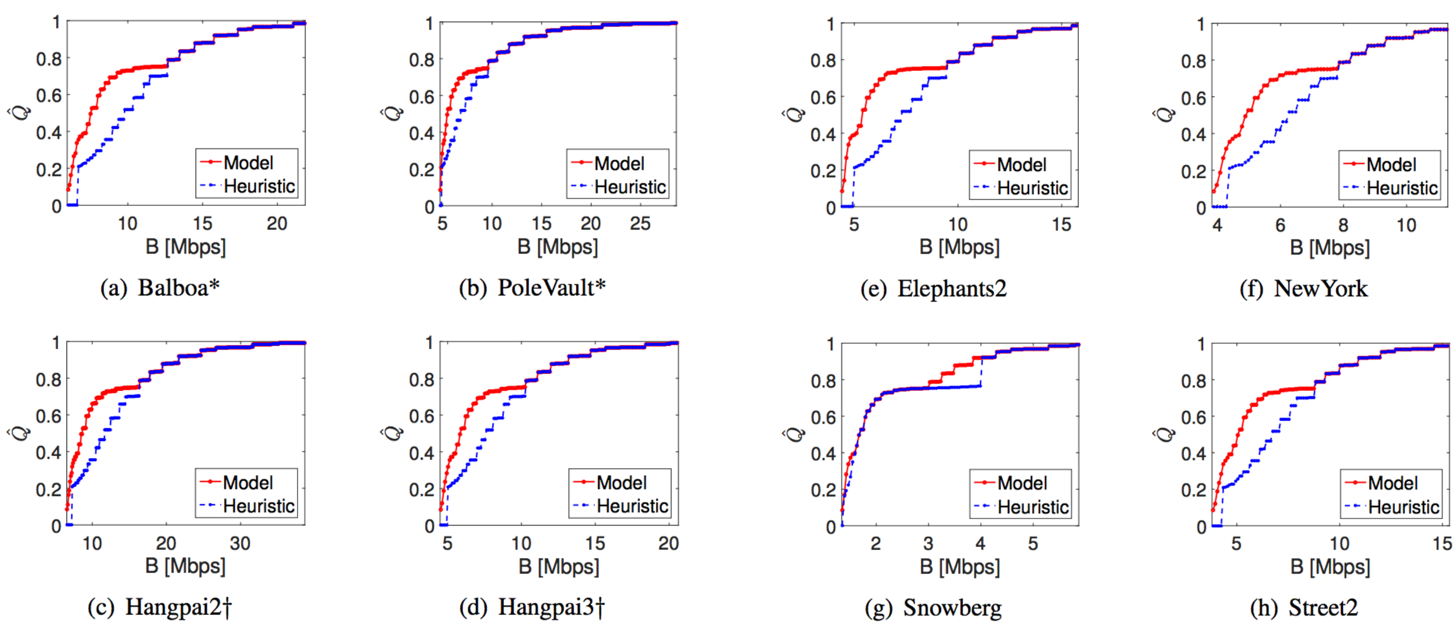
(g) Snowberg



(h) Street2

Performance Evaluation for Practical Adaptation

- Discrete quantization stepsize q and discrete spatial resolution s: 3×51 = 153 possibilities
- Compared to heuristic: s = 1/16 when B < 1 Mbps, s = 1/4 when 1 ≤ B < 4 Mbps, s = 1 when B ≥ 4 Mbps



(a) Balboa*
(b) PoleVault*
(e) Elephants2
(f) NewYork
(c) Hangpai2†
(d) Hangpai3†
(g) Snowberg
(h) Street2

Conclusion

- investigated the perceptual impact of the quality variations when performing the refinement within a period of time τ
- Future work: FoV adaptation prediction and apply the proposed model in practical immersive streaming system