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 $\tau$

• 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
Test Procedure

- 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 $\tau$: 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

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

\[
Q(\tau, \hat{q}, \hat{s}) = Q_{\text{max}} \cdot \hat{Q}_{\text{NQQ}}(\tau, \hat{q}) \cdot \hat{Q}_{\text{NQS}}(\tau, \hat{s}),
\]

where

\[
\hat{Q}_{\text{NQQ}}(\tau, \hat{q}) = a(\hat{q}) \cdot e^{-b(\hat{q}) \cdot \tau} + (1 - a(\hat{q})),
\]

\[
\hat{Q}_{\text{NQS}}(\tau, \hat{s}) = a(\hat{s}) \cdot e^{-b(\hat{s}) \cdot \tau} + (1 - a(\hat{s})).
\]
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
Quality-Bandwidth Optimized Streaming
Problem Formulation

\[
\begin{align*}
\max_{\tau, \hat{q}, \hat{s}} & \quad Q, \\
\text{s.t.} & \quad R_i^{F_{oV}} + R_{mi}^{R_{L}} \leq B, \\
& \quad 0 < \hat{q}, \hat{s} \leq 1. 
\end{align*}
\] (8)

Thereinto,

\[
\tau = \frac{R_i^{F_{oV}} + R_{mi}^{R_{L}}}{B} \cdot T, \\
R_i^{F_{oV}} = \sum_{j=1}^{n} R_{i,j}^{R_{L}}, \\
R_{mi}^{R_{L}} = R(\hat{q}, \hat{s}).
\] (11-13)
Optimal Solution Under Continuous $q$

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

$$R_{\text{RL}}(\hat{q}) = R_{\text{max}} \cdot \hat{q}^\alpha,$$

(14)

$$\max_{\hat{q}} \hat{Q} = a(\hat{q}) \cdot e^{-b(\hat{q}) \cdot \left(\frac{R_{\text{FoV}} + R_{\text{max}} \cdot \hat{q}^\alpha}{B}\right)} + 1 - a(\hat{q}),$$

(15)

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

(16)

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

(17)

![Graphs showing the relationship between $q_{\text{opt}}$, $Q_{\text{opt}}$, and $B$ for Balboa* and PoleVault*](image)
Optimal Solution Under Discrete $s$ and Continuous $q$

\[ R_{mi}^{RL}(\hat{q}, \hat{s}) = R_{\text{max}} \cdot \hat{q}^\alpha \cdot \hat{s}^\beta. \] (18)
Performance Evaluation for Practical Adaptation

- Discrete quantization stepsize $q$ and discrete spatial resolution $s$: $3 \times 51 = 153$ possibilities
- Compared to heuristic: $s = 1/16$ when $B < 1$ Mbps, $s = 1/4$ when $1 \leq B < 4$ Mbps, $s = 1$ when $B \geq 4$ Mbps
Conclusion

• investigated the perceptual impact of the quality variations when performing the refinement within a period of time $\tau$

• Future work: FoV adaptation prediction and apply the proposed model in practical immersive streaming system