GLOBAL OPTIMIZATION SOLUTION FOR DYNAMIC ADAPTIVE 360-DEGREE STREAMING

Xuekai Wei[1], *Mingliang Zhou*[2,3,*], *Weijia Jia*[3,4,*] *IEEE Fellow*

[1]School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China
[2]School of Computer Science, Chongqing University, Chongqing 400044, China
[3]BNU-UIC Institute of Artificial Intelligence and Future Networks
funded by Beijing Normal University (Zhuhai)
[4]Guangdong Key Lab of AI and Multi-modal Data Processing, UIC

ABSTRACT

This paper proposes a global optimization solution that combines different types of perceptual information to further improve transmission efficiency. First, a new rate-distortion (R-D) model is proposed to reflect the characteristics of 360-degree video. Second, a globally optimized adaptive bitrate control algorithm is proposed using both the R-D models to adjust the bitrate for each tile in a segment. Finally, a new model parameter updating strategy that is robust to quality variations is proposed to further reduce prediction errors. Comparison results indicate that the proposed method can outperform state-of-the-art methods in terms of various quality of experience (QoE) objectives[1].

***Index Terms*—** 360-degree video streaming, quality of experience, content-adaptive, global optimization

1. INTRODUCTION

The rapid development of consumer virtual reality (VR) devices has improved the comfort and smoothness of the user experience and driven increasing user demand for immersive experiences based on VR technology [1, 2]. A 360-degree video contains all of the visual information in the entire surrounding space, requiring high definition (a video resolution of 4K or above), a high frame rate (40 frames per second, fps, or above), and a high bitrate (10 Mbps or above) to guarantee an immersive user experience and high quality of experience (QoE) [3, 4]. All of the above requirements cause the data volume of 360-degree streaming videos to be several times that of ordinary videos, and consequently, the transmission of such videos may encounter bandwidth bottlenecks [5, 6]. Improvements in 360-degree video streaming schemes to increase transmission efficiency are currently a crucial task [7, 8].

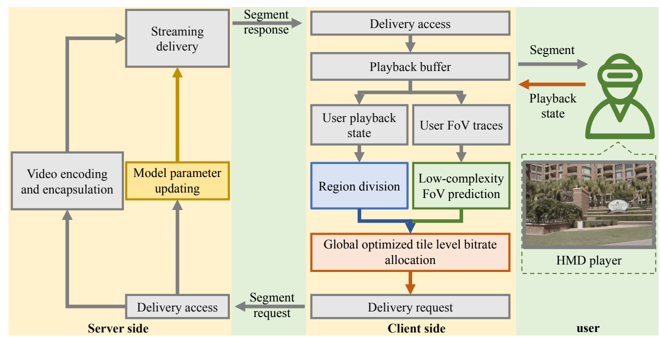


Fig. 1. Flowchart of the proposed 360-degree streaming solution.

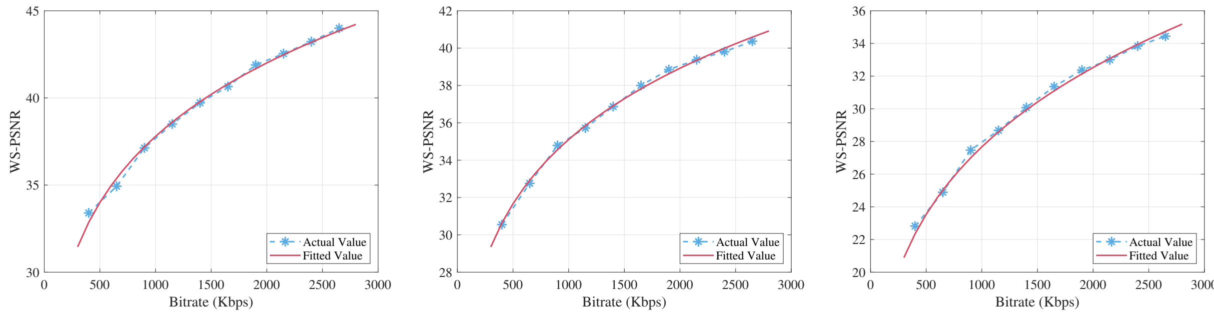
Tile-based or Non-tile streaming strategies are the most commonly used ones for 360-degree streaming [9, 10]. Scalable video coding is used in non-tile streaming strategies to achieve more smooth quality fluctuation [11]. In the tile-based streaming strategies, more human visual characteristics are analysed and utilised for tile-level bitrate allocation [12]. Due to the lower coding cost and higher flexibility, tile-based strategies are mainly studied in the proposed algorithm.

Although existing algorithms achieve some QoE gains, there remain challenges to be overcome. First, the variable quality nature of 360-degree frames represents an overlooked challenge in the design of quality adaptation algorithms. How to model the R-D relationship of different regions in one frame is important to avoid incorrect quality expectations, which will further help to make correct bitrate selection decisions [13, 14]. Second, most of the state-of-the-art bitrate selection methods do not take video content heterogeneity into consideration, which are top-down processes that lack an optimal global bit allocation between tiles; thus it is hard for them to be applied to streaming scenarios with mass

Xuekai Wei and Mingliang Zhou contributed equally to this work.
∗ Weijia Jia and Mingliang Zhou are the corresponding authors.

This work is supported by Guangdong Key Lab of AI and Multi-modal Data Processing, United International College (UIC), Zhuhai, Project No. 2020KSYS007, Chinese National Research Fund (NSFC) Project No. 61872239; The Institute of Artificial Intelligence and Future Networks funded by Beijing Normal University (Zhuhai) Guangdong, China, the National Natural Science Foundation of China under Grant 62176027, the General Program of National Natural Science Foundation of Chongqing under Grant cstc2020jcyj msxmX0790, the Zhuhai Postdoctoral Research Fund under Grant 110321001, and the International Postdoctoral Exchange Fellowship Program, the Zhuhai Postdoctoral Research Fund under Grant 110321001, and the International Postdoctoral Exchange Fellowship Program.

[1]Testing materials are available on the following website: https://github.com/OC007/Gcode01.



(a) Viewpoint Region (b) marginal Region (c) Not-viewed Region

Fig. 2. R-D models of the different regions.

video contents [9, 10, 11, 12]. Finally, commonly used video quality prediction designs do not reflect high video content fluctuation. Without suitable parameter updating strategies, the bitrate control algorithms may be deployed easily, but they currently may cause unexpected quality degradations under the actual network environment.

To address these problems, in this paper, a novel global optimization solution for dynamic adaptive 360-degree streaming is proposed. To maximize the transmission efficiency, a new R-D model, a globally optimized bitrate control algorithm, and a model parameter updating strategy are adopted to enhance the user QoE under various bandwidths with different QoE objectives. The contributions of the proposed optimization solution are listed as follows:

- A new R-D model considering 360-degree video characteristics is proposed to improve the transmission performance.
- A globally optimized adaptive bitrate control algorithm is proposed using both the R-D models and viewpoint accuracy to adjust the bitrate for each tile in a segment.
- A new R-D model parameter updating strategy that is robust to quality variations is proposed to further reduce prediction errors.

The remainder of this paper is organized as follows: Section 2 describes the details of the proposed method. The performance evaluation results are presented in Section 3. Finally, conclusions are given in Section 4.

2. PROPOSED OPTIMIZATION ALGORITHM

The flowchart of the proposed optimization solution is shown in Fig. 1. The basic 360-degree streaming strategy includes the viewpoint prediction, the video quality estimation, and the tile-level bitrate selection part. In this paper, we mainly focus on proposing new video quality estimation, and the tile-level bitrate selection methods. The viewpoint prediction results are generated following the method proposed in [3]. As shown in Fig. 1, to maximize the transmission efficiency, a new R-D model, a globally optimized bitrate control algorithm, and a model parameter updating strategy are adopted to enhance the user QoE under various bandwidths with different QoE objectives. The predicted viewpoint and marginal regions are different from frame to frame. To avoid the issue that predicted not-viewed regions may fall within the viewpoint or marginal region caused by the prediction error, the tile in not-viewed regions will be transmitted using a much lower bitrate at R_o to guarantee user QoE. By selecting the suitable bitrate for every tile according to their own R-D model shown in Fig. 1, user QoE can be further improved.

2.1. R-D Modelling

The R-D characteristics of the 360-degree videos are firstly analyzed in this subsection as a basis for the streaming control. To derive the R-D models, the video coding experiments are performed by kvazaar [8] with HEVC under rate control mode on the Joint Video Experts Team (JVET) 360-degree video testing sequences [5]. The resulting R-D fitting curves in accordance with the proposed model for several example video sequences in different regions are shown in Fig. 2. From which it can be seen that good fitting results are obtained by the following relation:

$$D(br) = m \cdot br^{-n}. \tag{1}$$

where m and n are model parameters related to the features of the source. Let

$$\lambda = -\frac{\partial D}{\partial br}. \tag{2}$$

Given

$$br = \left(\frac{m}{D}\right)^{\frac{1}{n}}. \tag{3}$$

We have the distortion of the i-th tile is expressed as follows:

$$D_i = \left(\frac{\lambda_i m_i^{\frac{1}{n_i}}}{n_i} \right)^{\frac{n_i}{n_i+1}}. \tag{4}$$

The R-D models for different regions are used to denote the R-D models for the viewpoint, marginal, and not-viewed regions, respectively. The model parameters are generally content-dependent.

2.2. Optimized Solution

The tile-level bitrate in the viewpoint region will be assigned at R_v. The tile-level bitrate in the marginal region will be assigned as a lower bitrate at R_l. Bitrate R_v and R_l will be solved using a globally optimized method. In the global bitrate allocation algorithm, the utility is firstly represented as the expected quality for the next segment, and the maximization problem can be formulated as a bargaining game [15, 11] that can be described as

$$\begin{aligned} &\{\lambda_1^*,\ \lambda_2^*, ...,\ \lambda_{N_{S_m}}^*\} = \text{argmin} \sum_{i=1}^{N_{S_m}} D(\lambda_i), \\ &s.t. \quad \sum_{i=1}^{N_{S_m}} br(\lambda_i) \le C_S. \end{aligned} \tag{5}$$

where C_S is the predicted available bitrate allocating for the tiles. Then it can be constructed into a Lagrangian cost function:

$$L = \sum_{i=1}^{N_{S_m}} \left(\frac{\lambda_i m_i^{\frac{1}{n_j}}}{n_i} \right)^{\frac{n_i}{n_i+1}} + \mu \left(\sum_{i=1}^{N_{S_m}} \ln \left(\frac{\lambda_i}{n_i m_i} \right)^{-\frac{1}{n_i+1}} - N_{S_m} \ln \left(\frac{C_S}{N_{S_m}} \right) \right), \tag{6}$$

where μ is the Lagrange multiplier. The optimal solution of Eq. (6) can be obtained by solving the Karush-Kuhn-Tucker (KKT) conditions[2] [16]. By solving Eq. (6), we can obtain the bitrate set $\{br_i^*\}$ as follows:

$$br_i^* = \left(\frac{\lambda_i^*}{n_i m_i} \right)^{-\frac{1}{n_i+1}}. \tag{7}$$

Thus, the bitrate R_v and R_l can be generated.

2.3. R-D Model Parameters Updating

As it is difficult to obtain the model parameters by relying on the video content before the encoding process, the optimal parameters are estimated using an updating strategy. Assuming that the current to-be-encoded frame is i, we aim to estimate the parameter based on the coding statistics of the i-1-th frame. In particular, we assume that the distortion of the co-located tiles is D_r and aim to minimize the difference between the real and estimated distortion D_p, which can be expressed as the squared error between $\ln D_r$ and $\ln D_p$:

[2]"Karush-Kuhn-Tucker conditions", [Online]. Available: http://en.wikipedia.org/wiki/Karush-Kuhn-Tucker_conditions.

Algorithm 1: Proposed optimization solution

Input: Playback state s_i, FoV traces, and network conditions.
Output: Tile-level bitrate in the next requested segment.
begin
  initialize: Bitrate set; bitrate for the first segment; *etc.*
  for $i \in N-1$ **do**
    if Segment i is fully downloaded **then**
      Generate the FoV prediction map and the total available bitrate [3].
      Select FoV prediction results and make region division.
      Allocate bitrate R_o for tiles in the not-viewed region.
      Allocate the bitrate R_l for tiles in the marginal region using Eq. (7).
      Compute the total bitrate C_S of the viewpoint region.
      Perform bitrate allocation R_v for every tile in the viewpoint region using Eq. (7).
      Next segment request.
      Perform parameter updating for every region using Eq. (9) & (10).
return bitrate set $\{br_i^*\}$.

$$e^2 = (\ln D_\text{r} - \ln D_\text{p})^2 \tag{8}$$

The above equation can be solved by the adaptive Least Mean Square (LMS) [17] method, and, we can obtain m_new and n_new:

$$\begin{aligned} m_\text{new} &\approx m_\text{old} \cdot (1 + \delta_m (\ln D_\text{r} - \ln D_\text{p})) \\ &= m_\text{old} + \delta_m (\ln D_\text{r} - \ln D_\text{p}) \cdot m_\text{old} \end{aligned} \tag{9}$$

$$\begin{aligned} n_\text{new} &= n_\text{old} - \delta \cdot (-2 (\ln D_\text{r} - \ln D_\text{p}) \cdot \ln br) \\ &= n_\text{old} + 2\delta \cdot (\ln D_\text{r} - \ln D_\text{p}) \cdot \ln br \\ &\triangleq n_\text{old} + \delta_n \cdot (\ln D_\text{r} - \ln D_\text{p}) \cdot \ln br \end{aligned} \tag{10}$$

2.4. Summary

This subsection summarizes the proposed R-D optimization solution for 360-degree streaming, shown in Algorithm 1. The tile in not-viewed regions will be transmitted using a much lower bitrate at R_o to guarantee user QoE. The tile-level bitrate in the marginal region will be assigned as a lower bitrate at R_l. The tile-level bitrate in the viewpoint region will be assigned at R_v. Compared with other algorithms [18,

Table 1. QoE Comparison between different methods in 360-degree Video Contents. The best, second and third results are in red, green and blue colors, respectively.

Method	L-Fix	L-LR	L-Pen	Abl1	Proposed
FoV Quality (dB)	34.06	39.23	41.06	42.84	43.67
FoV Bitrate (Mb)	2.98	3.35	3.41	3.93	4.30
Quality Temporal Difference (dB)	1.14	1.78	1.72	1.54	1.45
Playback Freezing Length (ms)	18.02	19.56	20.41	19.70	19.03
Prediction Precision	100.00	92.71	92.71	95.53	95.53
Prediction Error	0.00	0.17	0.17	0.08	0.08

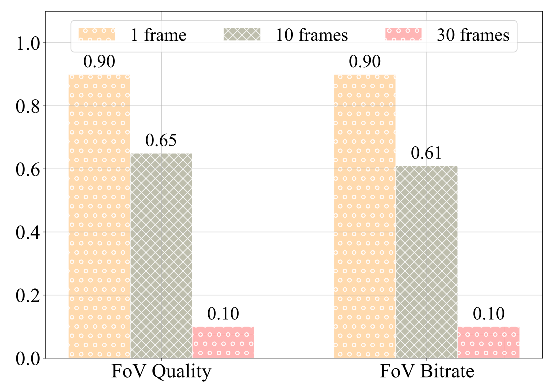


Fig. 3. Normalized comparison of FoV quality and bitrate in network traces with different segment durations.

8, 17, 19, 20], the emphases of the proposed approach are entirely different. First, to deal with the fact that the variable quality nature of 360-degree frames is overlooked, a new R-D modelling strategy for the streaming optimization problem is proposed. Second, most of the streaming methods do not take global video content heterogeneity into consideration, which results in sub-optimized streaming solutions. The proposed algorithm directly utilizes the observed R-D models and FoV prediction results to obtain the globally optimized bitrate decision-making algorithm for 360-degree streaming control. Finally, existing methods can not handle the video contents with strong spatio-temporal fluctuations. A model parameter updating strategy is proposed to further reduce the streaming control errors to avoid unexpected quality degradations under the actual network environment. The proposed algorithm outperforms other strategies in terms of various QoE components.

3. PERFORMANCE EVALUATION

The video dataset and network traces from [8] are selected to generate the streaming dataset. The following 360-degree streaming algorithms are compared and analyzed: The *L-FIX* method requests bitrates without considering the viewing behaviour [18]. The *L-LR* methods use linear regression FoV prediction and the probabilistic optimization model [8] to request bitrate. The *L-Pen* method is a modified streaming framework of the “Pensieve” method [6]. The *Alb1* method is the ablation method that does not use the proposed model parameter updating algorithms.

3.1. Effectiveness of R-D Optimization Solution

The effectiveness of the R-D optimization solution algorithm is evaluated in this subsection. Experimental results are shown in Fig. 3, which indicates that the proposed solution outperforms other methods achieving higher rendered FoV quality with lower network bandwidth waste. The FoV quality means the average effective FoV WS-PSNR, which is the instantaneous video quality metric. The FoV bitrate represents the average bitrate of tiles to render the FoV areas. An FoV quality decreasing trend can be seen from the results, and the main reason lies in the increase of the prediction error and the decrease of the prediction precision. By jointly considering the above analysis, the one frame segment length streaming will be further used to evaluate the QoE performance.

3.2. QoE Gain Analysis

Detailed QoE gains are shown in Table 1, including the FoV quality, the FoV bitrate, the quality temporal difference between two continuous adjacent segments, the playback freezing length and frequency, the prediction precision, and error. Comparison results show that the proposed global optimized solution, together with the proposed R-D method and the low-complexity FoV prediction method, can outperform state-of-the-art algorithms, achieving higher QoE achievements. Meanwhile, the proposed algorithm directly utilizes the observed frame R-D model and FoV prediction results to obtain the optimal solution, which will further guarantee QoE.

4. CONCLUSION

A global solution in the R-D domain was proposed for dynamic adaptive streaming optimization in 360-degree streaming with the aims of removing imperceptible redundancies and achieving efficient interactive transmission under various network conditions while significantly improving the user QoE. Bitrate allocation was performed considering different R-D models based on the video contents and visual importance to enhance the transmission performance. Evaluation results confirmed that jointly optimizing the R-D model and global bitrate allocation was beneficial in terms of various QoE objectives.

5. REFERENCES

[1] Pierre Lebreton and Kazuhisa Yamagishi, "Network and content-dependent bitrate ladder estimation for adaptive bitrate video streaming," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 4205–4209.

[2] X. Wei, C. Yang, and S. Han, "Prediction, communication, and computing duration optimization for vr video streaming," *IEEE Transactions on Communications*, vol. 69, no. 3, pp. 1947–1959, 2021.

[3] Kashyap Kammachi Sreedhar, Alireza Aminlou, Miska M. Hannuksela, and Moncef Gabbouj, "Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications," in *2016 IEEE International Symposium on Multimedia (ISM)*, 2016, pp. 583–586.

[4] M. Zink, R. Sitaraman, and K. Nahrstedt, "Scalable 360° video stream delivery: Challenges, solutions, and opportunities," *Proceedings of the IEEE*, vol. 107, no. 4, pp. 639–650, 2019.

[5] Lisha Wang, Chenglin Li, Wenrui Dai, Junni Zou, and Hongkai Xiong, "Qoe-driven and tile-based adaptive streaming for point clouds," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1930–1934.

[6] Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh, "Neural adaptive video streaming with pensieve," in *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, New York, NY, USA, 2017, SIGCOMM '17, p. 197–210, Association for Computing Machinery.

[7] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "An optimal tile-based approach for viewport-adaptive 360-degree video streaming," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 29–42, 2019.

[8] M. Xu, C. Li, S. Zhang, and P. L. Callet, "State-of-the-art in 360° video/image processing: Perception, assessment and compression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 5–26, 2020.

[9] Alban Marie, Navid Mahmoudian Bidgoli, Thomas Maugey, and Aline Roumy, "Rate-distortion optimized motion estimation for on-the-sphere compression of 360 videos," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1570–1574.

[10] Robert Skupin, Kai Bitterschulte, Yago Sanchez, Cornelius Hellge, and Thomas Schierl, "Rate assignment in 360-degree video tiled streaming using random forest regression," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 3687–3691.

[11] L. Sun, F. Duanmu, Y. Liu, Y. Wang, Y. Ye, H. Shi, and D. Dai, "A two-tier system for on-demand streaming of 360 degree video over dynamic networks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 43–57, 2019.

[12] J. Lin, Y. Lee, C. Shih, S. Lin, H. Lin, S. Chang, P. Wang, L. Liu, and C. Ju, "Efficient projection and coding tools for 360° video," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 84–97, 2019.

[13] H. Yuan, X. Wei, F. Yang, J. Xiao, and S. Kwong, "Cooperative bargaining game-based multiuser bandwidth allocation for dynamic adaptive streaming over http," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 183–197, 2018.

[14] Hui Yuan, Shiyun Zhao, Junhui Hou, Xuekai Wei, and Sam Kwong, "Spatial and temporal consistency-aware dynamic adaptive streaming for 360-degree videos," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 177–193, 2020.

[15] Jong-Beom Jeong, Soonbin Lee, Il-Woong Ryu, Tuan Thanh Le, and Eun-Seok Ryu, "Towards viewport-dependent 6dof 360 video tiled streaming for virtual reality systems," in *Proceedings of the 28th ACM International Conference on Multimedia*, New York, NY, USA, 2020, MM '20, p. 3687–3695, Association for Computing Machinery.

[16] M. Homayouni, A. Aminlou, and M. M. Hannuksela, "On subpicture-based viewport-dependent 360-degree video streaming using vvc," in *2020 IEEE International Symposium on Multimedia (ISM)*, 2020, pp. 85–90.

[17] L. Cui, D. Su, S. Yang, Z. Wang, and Z. Ming, "Tclivi: Transmission control in live video streaming based on deep reinforcement learning," *IEEE Transactions on Multimedia*, vol. 23, pp. 651–663, 2021.

[18] Jizhe Zhou, Chi-Man Pun, and Yu Tong, "Privacy-sensitive objects pixelation for live video streaming," in *Proceedings of the 28th ACM International Conference on Multimedia*, New York, NY, USA, 2020, MM '20, p. 3025–3033, Association for Computing Machinery.

[19] Yixiang Mao, Liyang Sun, Yong Liu, and Yao Wang, "Low-latency fov-adaptive coding and streaming for interactive 360° video streaming," New York, NY, USA, 2020, MM '20, p. 3696–3704, Association for Computing Machinery.

[20] Liyang Sun, Yixiang Mao, Tongyu Zong, Yong Liu, and Yao Wang, "Flocking-based live streaming of 360-degree video," in *Proceedings of the 11th ACM Multimedia Systems Conference*, New York, NY, USA, 2020, MMSys '20, p. 26–37, Association for Computing Machinery.