Edge-Assisted Multi-User 360-Degree Video Streaming

Deng Yixuan (23% System design), Lin Zexin (23% Writing of CFP),
Liu Junhua (31% Organize the whole project, idea of LTO/CFP), Wang Xi (23% Writing of LTO)

I. Background

RECENT years have witnessed the booming development of virtual reality (VR), which provides users an immersive and interactive experience to enjoy the virtual world. As one of the most important manifestations, 360° videos bring a new watching experience that has been widely adopted in many multimedia applications such as gaming, education, tourism, and sports. It is also a key technology that supports the development of the new paradigm *Metaverse*. Cisco Mobile Visual Networking Index (VNI) Forecast [1] indicates that 360° videos mobile data traffic will grow nearly 12-fold from 2017 to 2022. According to a recent market research report published on Globe Newswire [2], the global market size was USD 4.42 billion in 2020, exhibiting a significant growth of 42.2% compared to the average year-on-year growth during 2017-2019. They also predict that the market is projected to grow to USD 84.09 billion in 2028 at a compound annual growth rate (CAGR) of 44.8% in the 2021-2028 period.

In 360° videos, users usually need to wear a professional VR headset like HTC VIVE or simple VR-compatible mobile devices such as Google Daydream. How to stream 360° videos have recently captured great attention in academia [3]–[5]. Given the spherical features of 360° videos, a user is allowed to freely move his/her head and eye gaze to watch the most attractive portion of the video. And such a portion is called Field of View (FoV). Given the huge bandwidth overhead of transmitting the whole spherical angles of a 360° video, tile-based HTTP live streaming is widely used to balance the video transmission and user QoE [6]–[8]. With this standard, the video segment can be divided into many tiles with different qualities so that we can assign high-quality content in the FoV while low quality (or even blank content) for the rest of the video. However, the user's FoV can be quite dynamic and affected by many factors such as personalized watching preferences and video content, making it difficult to achieve accurate FoV prediction.

Besides, providing high-bandwidth and low-latency streaming to support multi-user viewing is also critical to 360° video watching. The development of 5G and mobile edge computing (MEC) provides a promising opportunity. The geo-distributed edge servers, e.g., base stations, can be used to cache the frequently requested tiles, which can significantly decrease network latency and bandwidth consumption, thus both improving the users' QoEs and saving the service cost [9]–[12]. Designing such a caching system, however, is non-trivial. Given an edge node serves multiple users within the proximity, how to satisfy the users' overall QoE with fairness into account and maintain a long-term cost-effective strategy to avoid back and forth replacement is still a challenging problem.

Existing solutions to these challenges are inadequate because they do not adequately address the problem from a holistic perspective. Most works [13]–[16] in this field focus on a single client and use a non-cooperative approach to video requesting, which leads to unnecessary waste of bandwidth and computation resources. Other efforts [9], [10], [17], [18] aim to improve the cache hit rate for multiple users by using information from FoV data, but these approaches are often short-sighted because they only consider the current state without considering long-term planning. Additionally, these works often prioritize user QoE at the expense of users whose FoVs are significantly different from the majority of users. This can lead to suboptimal performance for these maverick users. Furthermore, these works do not consider the potential benefits of collaborative FoV prediction for multiple users, which could improve performance for all users.

To this end, we propose `Our model` to entrap targets(video tiles) in a bubble(edge cache), a novel intelligent edge caching framework that aims to optimize user QoE and system cost by combining user FoV prediction and 360° video tile caching. `Our model` consists of two modules: a collaborative FoV prediction (CFP) module and a long-term tile caching optimization (LTO) module. We first conduct a comprehensive spatial and temporal analysis of the user FoV trajectory and find it is highly correlated with the video content, the target user's historical trajectory, and the FoV of other users who have watched the video. Thus, we design an intelligent CFP module that integrates these three features toward collaborative FoV prediction. This collaborative approach allows for more accurate FoV prediction and improved QoE for users. Once the FoV prediction has been made, the LTO module determines a caching policy that maximizes overall user QoE and minimizes system cost. It uses the Lyapunov framework, dual composition, and subgradient descent to solve the long-term optimization problem and determine the optimal caching policy.

We have conducted extensive evaluations and the real-trace driven experiments show that compared to state-of-the-art solutions `Our model` can achieve 36% improvement in QoE under similar traffic consumptions and outperforms in FoV prediction.

The contributions and novelty can be summarized as follows.

- We propose a novel framework that combines user field of view (FoV) prediction and video tile caching with edge computing to improve quality of experience (QoE) and economic efficiency for 360° video streaming for multi-users.
- We propose a collaborative FoV prediction architecture that integrates video content, user trajectory, and other users' records for combined prediction, resulting in accurate and robust results.
- We formulate a video tile caching problem that considers both QoE maximization and cost minimization, and solve it effectively using Lyapunov optimization and subgradient descent. This allows for the optimization of both QoE and economic efficiency in 360° video streaming.

II. Literature Review

The main research related to our work can be divided into three areas. The first part is about 360° video streaming. Then we show previous efforts on FoV prediction. At last, we present works about edge computing and caching for video streaming.

A. 360° Video Streaming

360° video is a new kind of video representation that is recorded by omnidirectional cameras and can provide an immersive and interactive watching experience to viewers. The video content forms as a sphere rather than a plane, where the viewer can freely adjust his/her head angle to watch a portion of content in the sphere. But in the stage of coding, the sphere is actually projected to a 2D plane using such projection methods as equirectangular or cube-map projections because existing encoders work on 2D rectangles.

Given such spherical features, streaming 360° videos requires much higher network bandwidth than traditional videos. To accommodate the high resource consumption, tile-based video encoding/streaming together with FoV-adaptive tile selection is proposed. Each frame of a video is split into different tiles and only tiles inside the user's FoV are streamed at high quality [6]–[8], [19], [20]. Petrangeli et al. [21] show an HTTP/2-based adaptive streaming framework to achieve higher performance. Nasrabadi et al. [22] propose a Scalable Video Coding encoding method so that the number of video rebuffering events can be reduced. Zare et al. [23] use the motion-constrained tile set feature of High Efficiency Video Coding standard to tackle the problem of multiple decoders at the user side to decode each tile. Guan et al. [24] leverage the 360° video-specific factors and Dasari et al. [25] use super Super-Resolution to save the bandwidth.

Hou et al. [16] propose a predictive adaptive streaming approach for mobile 360° and VR experiences. And Fei et al. [26] focus on the evaluation of QoE for 360° video transmission, including online, offline and mixed scenarios that can meet the requirement of real applications.

B. FoV Prediction

Accurate FoV prediction for 360° video is the key to tile-based adaptive streaming where many pioneer efforts have been made toward this goal. Many works [27], [28] use regression-based methodologies to predict the future FoV according to the historical trajectory, but they are not quite capable of capturing the inherent correlations. Xie et al. [29] cluster users periodically based on the head movement trajectory and assign new users to the existing clusters to do the prediction. Sun et al. [11] propose a flocking-based methodology for a live 360° video streaming where a large number of users are available concurrently. DRL360 [14] and SR360 [30] introduce deep reinforcement learning frameworks to predict FoV.

The above works are just based on historical trajectory and other works take video content into consideration. Most of those works analyze the video contents through a saliency map that shows the properties of an image at the pixel level. Fan et al. [31] use LSTM based model that learns the sensor-related features and image saliency map to predict viewer fixation. Nguyen et al. [32] propose PanoSalNet to learn the saliency map from user FoV data using DCNN and uses the LSTM network to predict the FoV. Park et al. [33] use the same inputs and find a tile probability map by a CNN + LSTM network. Besides the saliency map, PARIMA [34] uses YOLOv3 [35] to detect the objects and obtain their bounding box coordinates, then predict the FoV based on the track of objects.

C. Edge Computing and Caching

The emergence of edge computing provides a new compute paradigm for multimedia streaming that the videos can be cached or instantly processed at the distributed edge servers for better services. In this way, the end-to-end latency, bandwidth consumption, and energy consumption can be reduced for high-quality video streaming. Hou et al. [36] shift extensive rendering to edge servers to address the challenge from bitrate and latency requirements. Mangiante et al. [37] take advantage of mobile edge computing (MEC) to process and render FoV in order to optimize bandwidth consumption and battery utilization. Chakareski et al. [38] study the delivery of 360°-navigable videos to 5G VR/AR wireless clients in future cooperative multi-cellular systems.

Cheng et al. [15] took a holistic approach to video coding, proactive caching, computation offloading, and data transmission. Maniotis et al. [13] focused on the live-streaming scenario. But their works are only for a single user.

Teng et al. [18] looked at a massive MIMO system with multiple users in a single-cell theater, with a focus on wireless communication. Maniotis et al. [17] used the Deep Q-Network (DQN) algorithm for cache policy. But their work did not consider the multi-user problem in terms of FoV prediction.

The challenge in existing works is how to effectively and fairly serve multiple users within proximity at an edge node while maintaining a cost-effective strategy to avoid frequent replacement with an accurate multi-user-based FoV prediction.

III. Complete Proof of LTO: Long-term Tile Caching Optimization

This section shows the detail of LTO. We first leverage Lyapunov optimization to transform the long-term optimization problem into a problem in one timeslot, then solve this problem by dual decomposition and subgradient descent.

A. Lyapunov Optimization

The optimization problem involves the long-term expected terms, so we seek the technique of Lyapunov optimization.

Firstly, we introduce two virtual queues

$$F_{t+1} = max\{F_t + B_t - \eta_1, 0\} \tag{1}$$

and

$$G_{t+1,n} = max\{G_{t,n} - U_{t,n} + \eta_2, 0\}, \forall n \tag{2}$$

Lemma 1. *If the virtual queue F_t and $G_{t,n}$ are rate stable, i.e.,*

$$\lim_{T\to\infty} \frac{F_T}{T} \leq 0, \lim_{T\to\infty} \frac{G_{T,n}}{T} \leq 0, \tag{3}$$

then we have

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} B_t \leq \eta_1, \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} U_{t,n} \geq \eta_2. \tag{4}$$

Proof. We first rewrite F_{t+1} as below

$$F_{t+1} = \begin{cases} F_t + B_t - \eta_1, & \text{if } F_t \leq -B_t + \eta_1 \\ 0, & \text{if } F_t < -B_t + \eta_1 \end{cases} \tag{5}$$

Then, we have

$$\begin{aligned} F_{t+1} - F_t &= \begin{cases} B_t - \eta_1, & \text{if } F_t \leq -B_t + \eta_1 \\ -F_t, & \text{if } F_t < -B_t + \eta_1 \end{cases} \\ &= max\{B_t - \eta_1, -F_t\} \\ &\geq B_t - \eta_1 \end{aligned} \tag{6}$$

So we have

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} F_{t+1} - F_t &\geq \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} B_t - \eta_1 \\ \lim_{T\to\infty} \frac{F_T}{T} &\geq \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} B_t - \eta_1 \end{aligned} \tag{7}$$

Because the virtual queue is rate stable, then we have

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} B_t \leq \eta_1 \tag{8}$$

Similarly, we can show that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} U_{t,n} \geq \eta_2 \tag{9}$$

if $G_{t,n}$ is rate stable. □

By assuming the virtual queues F_t and $G_{t,n}$ are rate stable, the origin problem can be rewritten as:

$$objective: \min_{\mathcal{S}_t} \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \left[\left(-\sum_n U_{t,n}\right) + \beta B_t\right] \tag{10}$$

$$s.t. 0 \leq \sum \mathcal{S}_t \leq C \tag{11}$$

$$B_t \leq B_{max,t} \tag{12}$$

$$\lim_{T\to\infty} \frac{F_t}{T} \leq 0 \tag{13}$$

$$\lim_{T\to\infty} \frac{G_{t,n}}{T} \leq 0, \forall n \tag{14}$$

Then we can utilize the penalty drift in Lyapunov optimization to solve Equation 10. We first introduce a concatenated vector of the virtual queues as

$$\Theta_t \triangleq [F_t, G_{t,1}, \cdots, G_{t,N}] \tag{15}$$

The corresponding Lyapunov function can be defined as

$$L(\Theta_t) \triangleq \frac{F_t^2}{2} + \frac{G_{t,1}^2}{2} + \cdots + \frac{G_{t,N}^2}{2} \tag{16}$$

Then, the Lyapunov penalty drift $\Delta(\Theta_t)$ can be obtained as follows

$$\Delta(\Theta_t) = E\left[L(\Theta_{t+1}) - L(\Theta_t)|\Theta_t\right] \tag{17}$$

According to the Lyapunov optimization theory, enforcing the virtual queue constraint is equivalent to minimize the drift penalty, and minimizing the objective function with the virtual queue constraints is equivalent to minimize the "drift-plus-penalty" defined as follows

$$\Delta(\Theta_t) + V \times E\left[\left(-\sum_n U_{t,n}\right) + \beta B_t \middle| \Theta_t\right] \tag{18}$$

where $V \geq 0$ is the penalty weight, which represents the importance of the objective function compared to the virtual queue constraints.

Thus, the optimization problem can be rewritten as

$$objective:$$

$$\min_{\mathcal{S}_t}\left\{\Delta(\Theta_t) + V \times E\left[\left(-\sum_n U_{t,n}\right) + \beta B_t \middle| \Theta_t\right]\right\} \tag{19}$$

$$s.t. 0 \leq \sum \mathcal{S}_t \leq C \tag{20}$$

$$B_t \leq B_{max,t} \tag{21}$$

B. Dual Decomposition

A 360° video can be divided into E tiles both spatially and temporally. For convenience, we define two indicator variables

$$\gamma_t(e) = \begin{cases} 1, & \text{if } e \text{ in } \mathcal{S}_t \\ 0, & \text{if } e \text{ not in } \mathcal{S}_t \end{cases} \tag{22}$$

$$FoV_{t,n}(e) = \begin{cases} 1, & \text{if } e \text{ in } n\text{'s FoV} \\ 0, & \text{if } e \text{ not in } n\text{'s FoV} \end{cases} \tag{23}$$

Then we can rewrite $U_{t,n}$ and B_t as

$$U'_{t,n} = \frac{FoV_{t,n}(e)\left[\gamma_t(e) + \alpha'(1-\gamma_t(e))\right]}{\|FoV_{t,n}\|} \tag{24}$$

and

$$B'_t = \gamma_{t+1}(e)\left[1 - \gamma_t(e)\right] \tag{25}$$

where for convience, we combine the term denoted the traffic consumption caused by uncached but requested tiles in $U'_{t,n}$ by changing α'.

Algorithm 1 Cache Management Based on Dual Decomposition

Input: the number of time slots T, total cache size C, the number of tiles E, step size δ, threshold ϵ
Output: Cache Policy
  Initialize F'_1, G'_1
  for $t = 1$ to T **do**
    Initialize λ
    while $|\Delta\lambda| > \epsilon$ **do**
      for $e = 1$ to E **do**
        if e is in the predicted FoV **then**
          Solve Equation 34 to obtain $\gamma_t^*(e)$
        end if
      end for
      $\lambda = \left[\lambda - \delta(C - \sum_{e=1}^{E}\gamma_t^*(e))\right]^+$
      $\Delta\lambda = -\delta(C - \sum_{e=1}^{E}\gamma_t^*(e))$
    end while
    Update F'_{t+1}, G'_{t+1}
  end for

So the Equation 19 can be written as

$$objective:$$

$$\max_{\gamma_t(e)} \sum_e^E \left\{\Delta(\Theta'_t) + V \times E\left[\left(-\sum_n U'_{t,n}\right) + \beta B'_t \middle| \Theta_t\right]\right\} \tag{26}$$

$$s.t. 0 \leq \sum_e^E \gamma_t(e) \leq C \tag{27}$$

$$B_t \leq B_{max,t} \tag{28}$$

This optimization problem is clearly non-convex, to overcome this obstacle, we initiate the use of specific functions as described below

$$\begin{cases} c_0(\gamma_t(e)) = \Delta(\Theta'_t) + V \times E\left[\left(-\sum_n U'_{t,n}\right) + \beta B'_t \middle| \Theta_t\right] \\ c_1(\gamma_t(e)) = B_t - B_{max,t} \\ c_2(\gamma_t(e)) = \sum_e^E \gamma_t(e) - C \end{cases} \tag{29}$$

By utilizing Equation 29, the optimization problem presented in Equation 26 can be re-presented in a succinct manner as demonstrated below

$$objective: \max_{\gamma_t(e)} \sum_{e=1}^{E} c_0(\gamma_t(e)), \tag{30}$$

$$s.t. \; c_1(\gamma_t(e)) \leq 0, \tag{31}$$

$$c_2(\gamma_t(e)) \leq 0 \tag{32}$$

Next, we employ the technique of dual decomposition to address the complex coupling optimization problem. This approach transforms the original problem into a centralized master problem and several distrubuted subproblems. To begin with, we establish the definition of the Lagrange multiplier λ, which plays a crucial role in this process, as follows:

$$\min_{\lambda} \sum_{e=1}^{E} \psi_e(\lambda) - \lambda C, \ s.t. \lambda \geq 0 \tag{33}$$

where the subproblem for each tile is shown as follows

$$\psi_e(\lambda) = \sup_{\gamma_t(e)} \left\{ c_0(\gamma_t(e)) + \lambda \gamma_t(e) \middle| c_1(\gamma_t(e)) \leq 0 \right\} \tag{34}$$

For any given λ announced by the master problem, the edge can choose which tiles to cache by finding the near-optimal value $\gamma_t^*(e)$ by solving Equation 34. Then the Lagrange multiplier λ can be updated by the subgradient of the master problem Equation 33 shown as follows

$$\lambda_{K+1} = \left[\lambda_K - \delta(C - \sum_{e=1}^{E} \gamma_t^*(e)) \right]^+ \tag{35}$$

where δ is a positive stepsize, K is the iteration index, and $[.]^+$ denotes the projection onto the non-negative orthant.

Iterating all video items and tiles will use up a lot of computing power, making it difficult to deploy the system on an edge server. However, not all tiles need to be calculated. Only the tiles that appear in the predicted FoV are necessary, which is a small fraction compared to the total number of video items and tiles. Therefore, we only need to iterate over those effective tiles.

Algorithm 1 shows the procedure.

References

[1] Cisco, "Cisco visual networking index: Forecast and trends," 2018. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] F. B. Insights, "Virtual reality market size, share and covid-19 impact analysis, 2021-2028," 2021. [Online]. Available: https://www.fortunebusinessinsights.com/industry-reports/virtual-reality-market-101378

[3] C. Fan, W. Lo, Y. Pai, and C. Hsu, "A survey on 360° video streaming: Acquisition, transmission, and display," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 71:1–71:36, 2019.

[4] Y. Jin, J. Liu, F. Wang, and S. Cui, "Where are you looking?: A large-scale dataset of head and gaze behavior for 360-degree videos and a pilot study," in *Proceedings of the 30th ACM International Conference on Multimedia (MM), Lisboa, Portugal, October 10 - 14, 2022*.

[5] J. Liu, B. Zhu, F. Wang, Y. Jin, W. Zhang, Z. Xu, and S. Cui, "Cav3: Cache-assisted viewport adaptive volumetric video streaming," in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Shanghai, China, March 25 - 29, 2023*.

[6] M. Graf, C. Timmerer, and C. Müller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: design, implementation, and evaluation," in *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys), Taipei, Taiwan, June 20-23, 2017*.

[7] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer," in *Proceedings of the IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, December 11-13, 2016*.

[8] P. Rondao-Alface, J. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," *Bell Labs Tech. J.*, vol. 16, no. 4, pp. 135–147, 2012.

[9] A. Mahzari, A. T. Nasrabadi, A. Samiei, and R. Prakash, "Fov-aware edge caching for adaptive 360° video streaming," in *Proceedings of the ACM Multimedia Conference on Multimedia Conference (MM), Seoul, Korea, October 22-26, 2018*.

[10] A. Mahzari, A. Samiei, and R. Prakash, "Coopec: Cooperative prefetching and edge caching for adaptive 360° video streaming," in *Proceedings of the IEEE International Symposium on Multimedia (ISM), Naples, Italy, December 2-4, 2020*.

[11] L. Sun, Y. Mao, T. Zong, Y. Liu, and Y. Wang, "Flocking-based live streaming of 360-degree video," in *Proceedings of the 11th ACM Multimedia Systems Conference (MMSys), Istanbul, Turkey, June 8-11, 2020*.

[12] Z. Guo, J. Wang, S. Liu, J. Ren, Y. Xu, and Y. Wang, "Sponge-training: Achieving high efficiency and accuracy for wireless edge-assisted online distributed learning," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2022.

[13] P. Maniotis and N. Thomos, "Tile-based edge caching for 360° live video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4938–4950, 2021.

[14] Y. Zhang, P. Zhao, K. Bian, Y. Liu, L. Song, and X. Li, "DRL360: 360-degree video streaming with deep reinforcement learning," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Paris, France, April 29 - May 2, 2019*.

[15] Q. Cheng, H. Shan, W. Zhuang, L. Yu, Z. Zhang, and T. Q. S. Quek, "Design and analysis of MEC- and proactive caching-based 360° mobile VR video streaming," *IEEE Trans. Multim.*, vol. 24, pp. 1529–1544, 2022.

[16] X. Hou, S. Dey, J. Zhang, and M. Budagavi, "Predictive adaptive streaming to enable mobile 360-degree and VR experiences," *IEEE Trans. Multim.*, vol. 23, pp. 716–731, 2021.

[17] P. Maniotis and N. Thomos, "Viewport-aware deep reinforcement learning approach for 360° video caching," *IEEE Trans. Multim.*, vol. 24, pp. 386–399, 2022.

[18] L. Teng, G. Zhai, Y. Wu, X. Min, W. Zhang, Z. Ding, and C. Xiao, "Qoe driven VR 360° video massive MIMO transmission," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 1, pp. 18–33, 2022.

[19] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *Proceedings of the IEEE International Conference on Communications (ICC), Paris, France, May 21-25, 2017*.

[20] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, "Tiling in interactive panoramic video: Approaches and evaluation," *IEEE Trans. Multim.*, vol. 18, no. 9, pp. 1819–1831, 2016.

[21] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. D. Turck, "An http/2-based adaptive streaming framework for 360° virtual reality videos," in *Proceedings of the ACM on Multimedia*

Conference (MM), Mountain View, CA, USA, October 23-27, 2017.

[22] A. T. Nasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using scalable video coding," in *Proceedings of the ACM on Multimedia Conference (MM), Mountain View, CA, USA, October 23-27, 2017*.

[23] A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "Hevc-compliant tile-based streaming of panoramic video for virtual reality applications," in *Proceedings of the ACM Conference on Multimedia Conference (MM), Amsterdam, The Netherlands, October 15-19, 2016*.

[24] Y. Guan, C. Zheng, X. Zhang, Z. Guo, and J. Jiang, "Pano: optimizing 360° video streaming with a better understanding of quality perception," in *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM), Beijing, China, August 19-23, 2019*.

[25] M. Dasari, A. Bhattacharya, S. Vargas, P. Sahu, A. Balasubramanian, and S. R. Das, "Streaming 360-degree videos using super-resolution," in *Proceedings of the 39th IEEE Conference on Computer Communications (INFOCOM), Toronto, ON, Canada, July 6-9, 2020*.

[26] Z. Fei, F. Wang, J. Wang, and X. Xie, "Qoe evaluation methods for 360-degree VR video transmission," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 1, pp. 78–88, 2020.

[27] Y. Bao, H. Wu, T. Zhang, A. A. Ramli, and X. Liu, "Shooting a moving target: Motion-prediction-based transmission for 360-degree videos," in *Proceedings of the IEEE International Conference on Big Data (BigData), Washington DC, USA, December 5-8, 2016*.

[28] F. Qian, B. Han, Q. Xiao, and V. Gopalakrishnan, "Flare: Practical viewport-adaptive 360-degree video streaming for mobile devices," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom), New Delhi, India, October 29 - November 02, 2018*.

[29] L. Xie, X. Zhang, and Z. Guo, "CLS: A cross-user learning based system for improving qoe in 360-degree video adaptive streaming," in *Proceedings of the ACM Multimedia Conference on Multimedia Conference (MM), Seoul, Republic of Korea, October 22-26, 2018*.

[30] J. Chen, M. Hu, Z. Luo, Z. Wang, and D. Wu, "SR360: boosting 360-degree video streaming with super-resolution," in *Proceedings of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV), Istanbul, Turkey, June 10-11, 2020*.

[31] C. Fan, S. Yen, C. Huang, and C. Hsu, "Optimizing fixation prediction using recurrent neural networks for 360° video streaming in head-mounted virtual reality," *IEEE Trans. Multim.*, vol. 22, no. 3, pp. 744–759, 2020.

[32] A. Nguyen, Z. Yan, and K. Nahrstedt, "Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction," in *Proceedings of the ACM Multimedia Conference on Multimedia Conference (MM), Seoul, Republic of Korea, October 22-26, 2018*.

[33] S. K. Park, A. Bhattacharya, Z. Yang, S. R. Das, and D. Samaras, "Mosaic: Advancing user quality of experience in 360-degree video streaming with machine learning," *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 1, pp. 1000–1015, 2021.

[34] L. Chopra, S. Chakraborty, A. Mondal, and S. Chakraborty, "PARIMA: viewport adaptive 360-degree video streaming," in *Proceedings of the Web Conference 2021 (WWW), Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*.

[35] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018.

[36] X. Hou, Y. Lu, and S. Dey, "Wireless VR/AR with edge/cloud computing," in *Proceedings of the 26th International Conference on Computer Communication and Networks (ICCCN), Vancouver, BC, Canada, July 31 - Aug. 3, 2017*.

[37] S. Mangiante, G. Klas, A. Navon, G. Zhuang, R. Ju, and M. D. Silva, "VR is on the edge: How to deliver 360° videos in mobile networks," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, VR/AR Network@SIGCOMM, Los Angeles, CA, USA, August 25, 2017*.

[38] J. Chakareski, "VR/AR immersive communication: Caching, edge computing, and transmission trade-offs," in *Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, VR/AR Network@SIGCOMM, Los Angeles, CA, USA, August 25, 2017*.