thinh

Thích ứng khung nhìn theo QoE cho việc truyền phát video 360 độ với đơn hướng và đa hướng

TÓM TẮT

Nhờ vào khả năng cung cấp trải nghiệm nhập vai cho người dùng, video 360 độ đã trở thành một trong những công cụ hỗ trợ chính của thực tế ảo (VR). Truyền phát trực tiếp video 360 độ mang đến một chất lượng trải nghiệm (QoE) hấp dẫn bằng cách cho phép người dùng thay đổi trường nhìn (FoV) một cách tức thì. Các nghiên cứu gần đây đã đề xuất một mô hình truyền phát video 360 độ thích ứng khung nhìn sử dụng khả năng thích nghi tốc độ bit trong miền không gian và thời gian để sử dụng hiệu quả các nguồn băng thông hạn chế. Trong bài báo này, chúng tôi đề xuất một sự ước tính về truyền phát video 360 độ trực tiếp với đơn hướng và đa hướng trên mạng di động nhằm tối ưu hóa việc phân bổ băng thông cho các nhóm người sử dụng để nâng cao trải nghiệm của họ trong khi tận dụng sự đa dạng vốn có trong các điều kiện mạng và trường nhìn (FoV) của người dùng. Kết quả thí nghiệm của chúng tôi cho thấy giải pháp đề xuất nâng cao ít nhất từ 10,05% đến 23,04% về tốc độ bit FoV khi so sánh với một số giải pháp đang tồn tại.

Từ khóa: *Truyền tải thích ứng, Chất lượng trải nghiệm, Thực tế ảo, Video 360 độ.*

QoE-Driven viewport Adaptation for Live 360 Degree Video Streaming with Unicast and Multicast

Abstract

Thanks to its ability to provide immersive experience to users, 360-degree video has become one of the key enablers of Virtual Reality (VR). Live 360-degree video streaming offers an engaging Quality of Experience (QoE) by enabling users to change their field of view (FoV) instantly. Recent works have proposed a viewport adaptive 360-degree video streaming model that uses bitrate adaptation in spatial and temporal domains to utilize limited bandwidth resources efficiently. In this paper, we propose an estimate of Live 360 Degree Video Streaming in Unicast and Multicast over Mobile Networks that optimizes bandwidth allocation to user groups to enhance their experience while taking advantage of the inherent diversity in network conditions and users' field of view. Our experimental results show that the proposed solution enhances at least 10.05% up to 23.04% in term of FoV bitrate in comparison with some existing solutions.

Keywords: *Adaptive Streaming, Quality of Experience, Virtual Reality, 360-degree Video.*

1. INTRODUCTION

Virtual reality[1-4] is a simulated environment generated by computer technology that users can explore in 360 degrees. It enables users to view, navigate, and engage with three-dimensional (3D) worlds. By using a virtual reality headset or other display devices, VR offers an immersive and interactive experience that brings the virtual world to life. With technological advancements, Virtual Reality applications have been developed across multiple fields, including education, healthcare, manufacturing, and entertainment.

Viewed from a different angle, High-Efficiency Video Coding is a video compression standard designed to greatly enhance compression efficiency without compromising video quality[5]. Compared to H.265/AVC, HEVC can reduce the bitrate by as much as 50%, while accommodating a wide range of video resolutions, including 8K ultra-high definition.

In addition, the Evolved Multimedia Broadcast Multicast Services (eMBMS) supported by 4G-LTE mobile networks enable multicast transmission to multiple users[6]. This allows a group of users to stream and watch live 360-degree video over a mobile network using a Head-mounted Display. As shown in Figure 1, MBMS provides an effective solution for distributing high-quality video content over cellular networks by employing video standardization through H.265/HEVC to the encoder[7]. Through multicast or broadcast transmission, MBMS technology enhances network resource utilization and lowers the total bandwidth needed to deliver multimedia content to multiple users[8].

This paper introduces a novel viewport prediction method designed for multi-user 360-degree video streaming over mobile networks. The client-based approach, referred to as QoE-ViLA (Viewport-based Live-streaming Adaptation in unicast and multicast), allows for the assessment of spectral efficiency to enhance video quality on the client side in the context of QoE-Driven Viewport Adaptation for Live 360° Video Streaming in both Unicast and Multicast modes. Within QoE-ViLA, video tiles are encoded into multiple representations using High-Efficiency Video Coding (HEVC). The main contributions of the proposed method are outlined as follows:

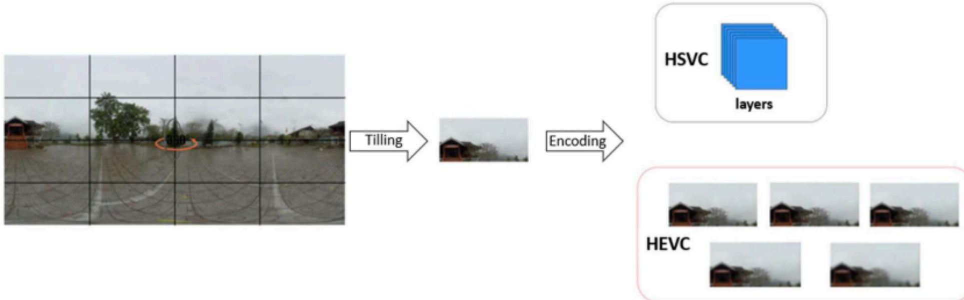


Figure 1. Video encoding standardized using H.265/HEVC

- Throughout a video streaming session, tiles outside the user's viewport are delivered only in a base version that ensures basic quality, while tiles within the viewport are transmitted in several enhanced versions. As a result, QoE-ViLA achieves more efficient use of available network resources compared to traditional methods.
- Impact of spectrum efficiency on different versions of multicast transport modes for minimizing network bandwidth usage.

The rest of the paper is structured as follows: Section 2 discusses about the state of the art. The resource allocation problem is shown in Section 3. The proposed method - QoE-ViLA - is described in Section 4 followed by an evaluation in Section 5. Finally, the conclusion is discussed in Section 6.

2. RELATED WORKS

Currently, researchers are making significant efforts to provide smooth omnidirectional video streaming with low latency, ensuring a satisfying viewing experience for users even under constrained network bandwidth. In the referenced study[9], the authors introduce a 360-degree video streaming framework that adjusts bitrate according to the user's viewport. This approach incorporates the spatiotemporal rate-distortion properties of 360-degree videos along with user navigation patterns. By analyzing user behavior to predict viewport transitions, the system adapts video quality accordingly, resulting in an enhanced quality of experience[10]. Another study[11] proposes an optimized video streaming approach using a chunk download scheduling framework, which enables the adjustment of each parallel stream's quality based on current bandwidth availability and the probability of stream switching.

Several recent studies[12-14] have addressed the challenges associated with multicasting/unicasting 360-degree video to multiple mobile users, including large video data volumes, varying network conditions, limited bandwidth resources, and levels of user interaction. Paper[14] presents a multi-quality multicast streaming approach for 360-degree videos targeting multiple users via time division multiple access (TDMA). To enhance transmission efficiency, the authors propose dividing encoded tiles into several representations at varying quality levels, enabling optimized decisions on quality selection, time allocation, and power distribution. Meanwhile, in paper[13], users are grouped into multiple multicast clusters, each assigned a specific quality representation to maximize perceived video quality. This approach ensures smooth playback quality of 360-degree content while conserving energy without compromising visual fidelity. In paper[12], the multi-session Omnicast system is employed to optimize the delivery of 360-degree video to multiple users under limited network resource conditions. Paper[15] proposes a framework for view-oriented, scale-optimized 360° video streaming, taking into account user viewport navigation behavior and the spatial scale distortion characteristics of 360° video content. Users are categorized into different classes based on their network bandwidth, with each layer corresponding to a multicast group receiving an identical set of tiles. In paper[16], a multicast fusion approach is introduced for adaptive multi-user 360-degree video streaming, aiming to maintain high quality of experience (QoE) in bandwidth-constrained scenarios. This solution allows for dynamic selection between multicast and unicast modes, as well as bitrate adaptation for each cell based on the user's available bandwidth and field of view (FoV).

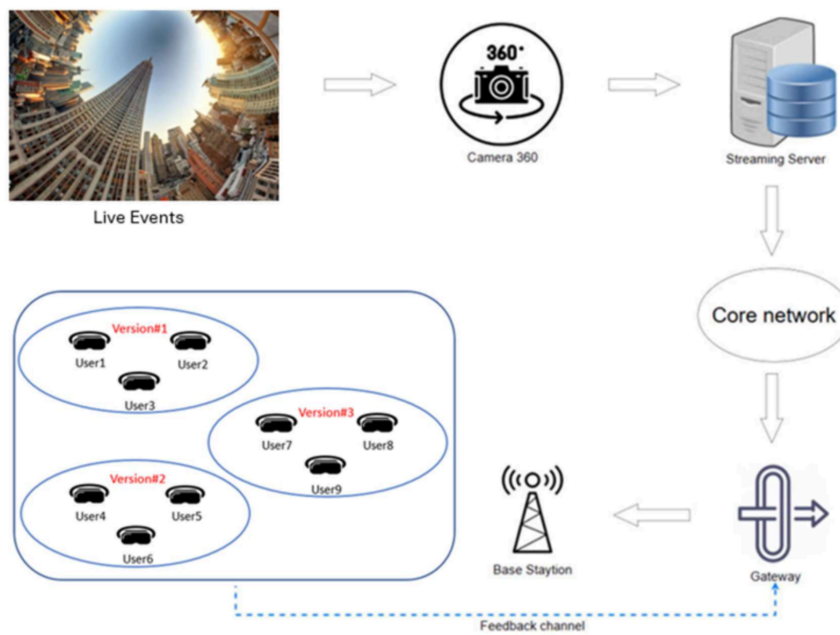


Figure 2. System Architecture Overview

Alternatively, Viewport Adaptive Streaming (VAS) enhances user experience under fluctuating network throughput and dynamic head movements. Besides, VAS approaches also rely on accurate head movement data and highly reliable viewport motion prediction[17]. By leveraging users' past viewport data, cross-user behavior patterns[18], and video content characteristics[19,20], it is possible to forecast users' future viewport positions. To address this challenge, another solution involves using Scalable Video Encoding, which encodes individual tiles into a base layer and multiple enhancement layers[21]. Initially, the base layers of all tiles are downloaded, while the enhancement layers of visible tiles are subsequently retrieved to enhance frame quality and ensure smooth video playback under varying frame and network conditions. This approach also allows users the flexibility to select which part of the video they wish to view at the highest Quality of Experience (QoE), limited to the visible portion within the video's viewport (View Field). In paper[22], the authors propose a method that maintains a high buffer level and ensures consistent quality in subsequent segments, even when the average bitrate drops, thereby improving overall QoE. Since users seldom view areas outside their current viewports, more than 80% of bandwidth is wasted during 360° video streaming[23]. Viewport Adaptive Streaming addresses this inefficiency by delivering the viewport region in high quality while encoding the remaining areas at lower quality, thereby optimizing bandwidth usage and reducing data waste[24].

3. RESOURCE ALLOCATION PROBLEM

As context, Figure 2 illustrates our proposed system for live 360-degree video streaming over mobile networks, which enables flexible distribution of multimedia content in multiple versions. This system leverages HEVC (High-Efficiency Video Coding) to address the resource allocation challenge.

Let B_{nh} denote the bitrate, and let Δ_{nh} represent the quality of layer h for tile n. The parameters H, M, and N correspond to the number of layers, users, and tiles, respectively. A resource block (RB) is denoted as R. Furthermore, each group's traffic is allocated R resource blocks (RBs), with the group's minimum spectral efficiency represented by σ.

Suppose that R network resource blocks are allocated to a specific user group. The spectral efficiency of user m (where $1 \leq m \leq M$) characterizes the quality of the wireless link between the base station and the user's terminal device. A live 360-degree video stream is divided into N tiles, each of which is encoded into $(H + 1)$ layers of different quality levels, with each layer representing a distinct version. For each tile n ($1 \leq$

$n \leq N$), layer h ($0 \leq h \leq H$) is associated with a bitrate β_{nh} and quality Δ_{nh}, corresponding to version h. Furthermore, each version is partitioned into segments, with each segment having a playback duration of τ seconds. The following formula specifies the number of network resource blocks, denoted as R_{mnh}, required to transmit version h of each tile to user m:

$$R_{mnh} = \frac{\beta_{nh} \times \tau}{\sigma^m} \quad (1)$$

where σ^m denotes the spectral efficiency of the wireless communication link between the base station and the user's terminal device.

The proposed system supports two transmission methods for delivering a version: unicast and multicast. In unicast mode, the version is transmitted to an individual user, whereas in multicast mode, it is distributed to a group of users. Furthermore, the spectral efficiency of multicast transmissions is constrained by the user with the lowest spectral efficiency in the group, which limits the overall multicast capacity.

4. PROPOSED SOLUTIONS (QoE-ViLA)

When viewing 360-degree videos under fluctuating mobile network bandwidth conditions that may cause buffering or playback interruptions, it becomes essential to adapt in real-time by dynamically modifying parameters like tile quality and transmission rate. By constantly tracking network performance and factoring in available bandwidth, latency, device capabilities, and the viewer's current viewport, the system can maintain smooth playback and high visual quality, ensuring a seamless and enjoyable viewing experience.

This subsection introduces the two transmission modes for each tile's layer using two binary variables, γ_{nh}, $\varphi_{mnh} \in \{0, 1\}$, $\forall h \in H, n \in N, m \in M$, as described below:

- γ_{nh} = 1: if layer h of tile n is sent in multicast mode to all users;
- φ_{mnh} = 1: if layer h of tile n is sent in unicast mode to user m;
- $\gamma_{nh} = 0$ or $\varphi_{mnh} = 0$: otherwise.

The objective function governing network resource allocation for streaming to multiple users is formulated as follows:

$$maxx \sum_{m \in M} \sum_{n \in N} \sum_{h \in H} \Omega_{mn} \Delta_{nh} (\gamma_{nh} + \varphi_{mnh}) \quad (2)$$

With constraints:

$$\sum_{n \in N} \sum_{h \in H} \frac{\beta_{nh} \times \tau}{min_{M \in \beta nh)=1} \sigma^m} \leq R \quad (3)$$

Where:

$$\beta_{nh} = \{ m \in \{1,\ldots,M\} \;/\; \gamma_{nh} + \varphi_{mnh} = 1 \}$$

$$\sum_{m} \sum_{n} \varphi_{mnh} \leq 1, \forall m \in M, n \in N, h \geq 1 \quad (4)$$

Table 1. The average bitrate (kbps) and quality (dB) of the four 360-degree videos utilized in the experiment for the Multicast-all and Multicast-sca approaches

Version	Diving		Paris		Venice		RollerCoaster	
	Bitrate	**Quality**	**Bitrate**	**Quality**	**Bitrate**	**Quality**	**Bitrate**	**Quality**
Version 1	158.73	34.50	61.45	38.20	50.35	32.74	55.34	39.46
Version 2	313.77	38.24	88.73	41.93	114.25	35.91	106.75	42.71
Version 3	504.53	40.84	114.01	44.62	216.23	38.50	179.10	44.74
Version 4	1036.98	43.88	191.32	47.54	483.59	41.56	389.13	27.13
Version 5	1418.45	46.22	293.74	50.41	824.69	44.52	632.93	49.23

Table 2. The average bitrate (kbps) and quality (dB) of the four 360-degree videos utilized in the experiment for the QoE-ViLA and JUMPS approaches

Version	Diving		Paris		Venice		RollerCoaster	
	Bitrate	**Quality**	**Bitrate**	**Quality**	**Bitrate**	**Quality**	**Bitrate**	**Quality**
Version 1	158.15	34.49	60.78	38.20	60.07	32.73	54.86	39.45
Version 2	393.63	38.22	130.38	42.19	183.02	36.07	131.71	42.86
Version 3	752.26	40.97	209.69	44.99	384.87	38.69	250.53	44.99
Version 4	1481.51	44.01	340.25	47.88	826.38	41.76	515.88	47.28
Version 5	2499.02	46.57	532.44	50.72	1520.15	44.76	933.84	49.42

$$\sum_{n} \gamma_{n0} + \varphi_{mn0} = 1, \forall m \in M, \forall n \in N \quad (5)$$

$$\sigma^{m} = \frac{R_m(1-\gamma)}{B} \quad (6)$$

Where: Ω_{mn} represents the weight of tile n for user m; B denotes the user's bandwidth, while B also refers to the channel bandwidth at the core; R_m indicates the number of network resource blocks required by user m; and σ^m denotes the spectral efficiency of the wireless communication between the base station and the user terminal.

Accurate prediction of the future viewport direction is essential for effective network resource allocation, as it guides the selection of γ_{nh} and φ_{mnh} values (as defined in Equation 2). The predicted direction influences γ_{nh}, which determines the weight assigned to each view or tile in a 360-degree video. By anticipating which areas the viewer is likely to focus on, higher γ_{nh} values can be allocated to those tiles, ensuring attention is directed toward the most relevant content. Likewise, φ_{mnh}—indicating the quality or level of detail of each tile—is also affected by the predicted direction, allowing for optimized visual fidelity where it matters most. By predicting the future viewing direction, higher quality or more detailed tiles can be assigned to regions likely to be viewed, thereby improving visual fidelity in those areas. In this study, two constraints are applied: the first ensures that the total network resources allocated in both multicast and unicast modes do not exceed the limit R, as defined in Equation 3. Additionally, the constraints in Equations 4 and 5 guarantee that each user receives every Field of View (FoV) tile only once. Moreover, each user is consistently provided with the base layer for all tiles.

5. PERFORMANCE EVALUATION

5.1. Experimental Settings

In this section, we prepare four 360-degree videos, namely *Diving, Paris, Venice* and *RollerCoaster*. Each video has a resolution of 3840×2048 and a duration of 60 seconds. Using the 360Lib software, we convert these videos into the CubeMap format with a resolution of 2890×1920. The videos are then divided into 24 tiles, each sized 480×480 pixels. Each tile is continuously encoded into five layers: one base layer and four enhancement layers. The Quantization Parameters for the respective layers are fixed as specified in Table 1 and Table 2.

Following the setup in paper[25], we set the number of users M to 15. The σ^m values are applied as defined in Equation 6. The experimental evaluation was conducted on a system running a 64-bit Windows 10 operating system with 8GB of RAM and a 3.8GHz Intel Core i7 processor. We compared the proposed QoE-ViLA method with reference methods, including the approach from paper[25] (referred to as Multicast-sca), Multicast-all[12], and JUMPS[26]. Additionally, we recalculated the performance of the proposed and reference methods using the average viewport PSNR (Peak Signal-to-Noise Ratio), based on the following formula:

$$UVPR = \frac{1}{M} \sum_{m=1}^{M} UVPR_m \quad (7)$$

$$UVPR_m = \sum_{m=1}^{M} \Omega_{mn}^{a} \times \Delta_{nh_m^a} \quad (8)$$

Where: $UVPR_m$ denotes the viewport PSNR for user m; Ω_{mn}^{a} represents the weighted sum of the quality of visible tiles; and h_m^a indicates the highest layer of tile n streamed to user m.

Table 3. Average bitrate (kbps) in the case of 15 users

Videos	JUMPS	Mul-all	Mul_sca	QoE-ViLA
Diving	597.00	609.39	637.68	737.50
Paris	711.78	715.97	732.35	863.02
Venice	587.74	608.11	630.63	718.01
Roller.	692.71	693.64	709.58	840.78

Table 4. Average bitrate (kbps) in the case of 30 users

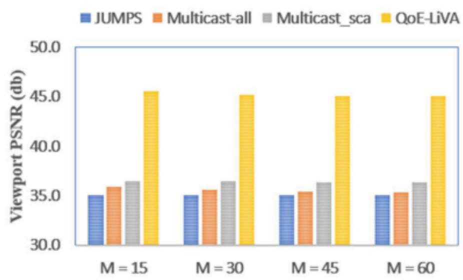
Videos	JUMPS	Mul-all	Mul_sca	QoE-ViLA
Diving	1185.20	1182.82	1263.17	1459.63
Paris	1417.94	1403.60	1456.70	1711.17
Venice	1375.00	1362.06	1410.78	1669.13
Roller.	1156.68	1181.48	1249.86	1420.57

Table 5. Average bitrate (kbps) in the case of 45 users

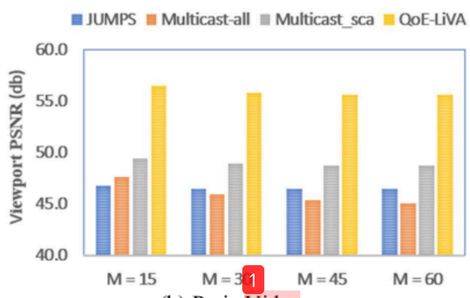
Videos	JUMPS	Mul-all	Mul_sca	QoE-ViLA
Diving	1777.80	1756.00	1890.65	2184.71
Paris	2126.90	2089.08	2180.62	2561.09
Venice	2062.48	2029.02	2112.20	2499.06
Roller.	1735.06	1752.80	1869.06	2125.93

Table 6. Average bitrate (kbps) in the case of 60 users

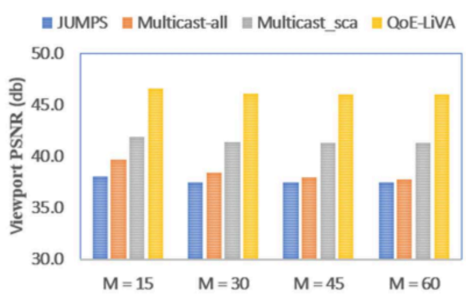
Videos	JUMPS	Mul-all	Mul_sca	QoE-ViLA
Diving	2370.39	2332.60	2518.09	2910.43
Paris	2835.86	2776.22	2905.89	3412.17
Venice	2749.95	2698.42	2814.74	3330.21
Roller.	2313.41	2326.90	2490.91	2831.55



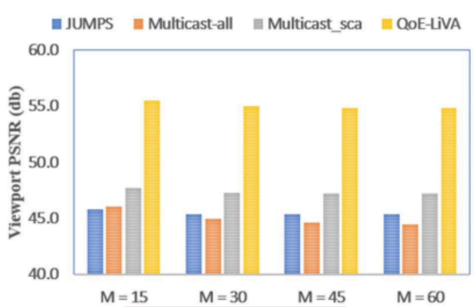
(a) Diving Video



(b) Paris Video

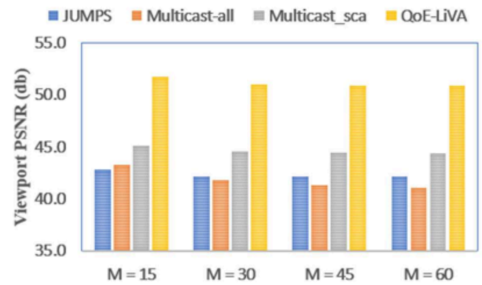


(c) Venice Video

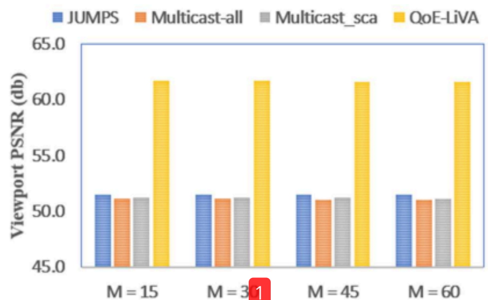


(d) RollerCoaster Video

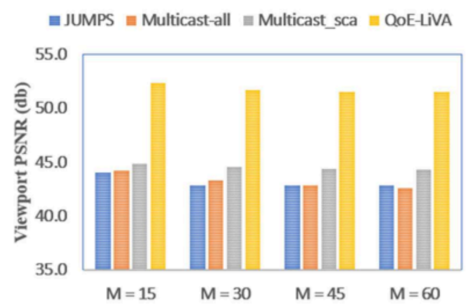
Figure 3. The performance across different numbers of users (with R = 200000 RBs) varies between the proposed method and the reference methods.



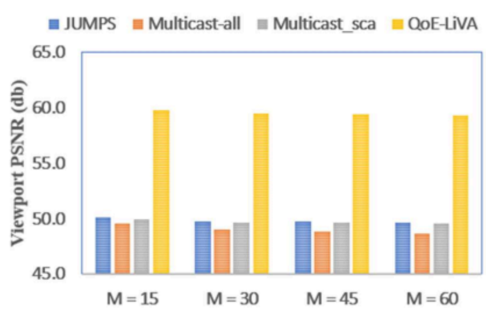
(a) Diving Video



(b) Paris Video



(c) Venice Video



(d) RollerCoaster Video

Figure 4. The performance across different numbers of users (with R = 700000 RBs) varies between the proposed method and the reference methods.

5.2. Experimental Results and Discussion

In this section, we evaluate the performance of the proposed framework in comparison with reference methods. A bitrate analysis is conducted using *R* multicast transmissions for user groups of 15, 30, 45, and 60 users, with *R* varying from 20 kRBs to 700 kRBs. As shown in Tables 3 through 6, the QoE-ViLA method consistently outperforms all other referenced approaches across all four videos. Specifically, the proposed method improves the average bitrate by more than 14.10% and 17.89% compared to the Multicast-sca and JUMPS methods, respectively. In comparison with Multicast-all, the performance gain ranges from 16.85% with 15 users to 18.83% with 60 users.

Additionally, we examine the effectiveness of the proposed method in terms of viewport PSNR. In our experiments, four values of *M*—15, 30, 45, and 60—are specifically analyzed. Figures 3 and 4 illustrate the viewport PSNR results for both the proposed and reference methods at 200000 RBs and 700000 RBs, respectively. The results clearly show that our approach consistently outperforms all reference methods across all *M* values. The QoE-ViLA method achieves a viewport PSNR that is at least 5 dB and up to 10 dB higher than the other methods in all scenarios. Moreover, although the viewport PSNR of the proposed method slightly decreases as the number of users increases, the reduction remains minimal.

6. CONCLUSION

This paper proposes a novel viewport prediction approach for multi-user omnidirectional video streaming over cellular networks. The method leverages High-Efficiency Video Coding (HEVC) along with enhanced spectral efficiency to optimize bandwidth usage across users. Experimental results demonstrate that the proposed method can achieve significant performance improvements, ranging from 10.05% to 23.04% across four test videos. In future work, we aim to extend the capabilities of the proposed framework to support scenarios involving multiple users concurrently accessing various video assets, including live 360-degree video streams.

thinh

ORIGINALITY REPORT

45%

SIMILARITY INDEX

PRIMARY SOURCES

1 Nguyen Viet Hung, Pham Hong Thinh, Nguyen Huu Thanh, Tran Thanh Lam, Tran Thuy Hien, Vu Thanh Ninh, Truong Thu Huong. "LVSUM-Optimized Live 360 Degree Video Streaming in Unicast and Multicast Over Mobile Networks", 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN), 2023 — 1164 words — 34%
Crossref

2 Duc Nguyen, Nguyen Viet Hung, Nguyen Tien Phong, Truong Thu Huong, Truong Cong Thang. "Scalable Multicast for Live 360-degree Video Streaming over Mobile Networks", IEEE Access, 2022 — 224 words — 6%
Crossref

3 mingtang.site — 27 words — 1%
Internet

4 Nguyen Viet Hung, Tran Thanh Lam, Tran Thanh Binh, Alan Marshal, Truong Thu Huong. "Efficient Deep Learning-Based Viewport Estimation for 360-Degree Video Streaming", Advances in Science, Technology and Engineering Systems Journal, 2024 — 17 words — < 1%
Crossref

5 studentsrepo.um.edu.my — 14 words — < 1%
Internet

6 Nguyen Dinh Thuan, Van Thi Phuong Lam. "MAPPING OF RELATIONAL AND OBJECT DATABASE TO NOSQL", KỶ YẾU HỘI NGHỊ KHOA HỌC CÔNG — 10 words — < 1%

NGHỆ QUỐC GIA LẦN THỨ XIII NGHIÊN CỨU CƠ BẢN VÀ ỨNG DỤNG CÔNG NGHỆ THÔNG TIN - Proceedings of the 13th National Conference on Fundamental & Applied Information Technology Research, 2020
Crossref

7	dtc24h.com Internet	10 words — < 1%
8	nhathongminhplus.com Internet	10 words — < 1%
9	www.researchgate.net Internet	10 words — < 1%
10	123doc.org Internet	9 words — < 1%
11	docplayer.hu Internet	9 words — < 1%
12	quatangamnhac247.blogspot.com Internet	9 words — < 1%
13	Duc Nguyen, Nguyen Viet Hung, Nguyen Tien Phong, Truong Thu Huong, Truong Cong Thang. "Scalable Multicast for Live 360-Degree Video Streaming Over Mobile Networks", IEEE Access, 2022 Crossref	8 words — < 1%
14	anchor.fm Internet	8 words — < 1%
15	www.thenho.net Internet	8 words — < 1%
16	Vu Duy Hien. "An Efficient Solution for Privacy-preserving Naïve Bayes Classification in Fully Distributed Data Model", Journal of Science and Technology on Information security, 2022	7 words — < 1%