

A 3D Framework for Improving Low-Latency Multi-Channel Live Streaming

Aizierjiang Aiersilan
 University of Macau
 Macau SAR, China
 ezharjan@outlook.com

Zhiqiang Wang
 Beijing University of Technology
 Beijing, China
 wzqcg@bjut.edu.cn

Abstract—The advent of 5G has driven the demand for high-quality, low-latency live streaming. However, challenges such as managing the increased data volume, ensuring synchronization across multiple streams, and maintaining consistent quality under varying network conditions persist, particularly in real-time video streaming. To address these issues, we propose a novel framework that leverages 3D virtual environments within game engines (e.g., Unity 3D) to optimize multi-channel live streaming. Our approach consolidates multi-camera video data into a single stream using multiple virtual 3D canvases, significantly increasing channel amounts while reducing latency and enhancing user flexibility. For demonstration of our approach, we utilize the Unity 3D engine to integrate multiple video inputs into a single-channel stream, supporting one-to-many broadcasting, one-to-one video calling, and real-time control of video channels. By mapping video data onto a world-space canvas and capturing it via an in-world camera, we minimize redundant data transmission, achieving efficient, low-latency streaming. Our results demonstrate that this method outperforms some existing multi-channel live streaming solutions in both latency reduction and user interaction responsiveness improvement.

Index Terms—live streaming, latency, multi-channel, data processing, virtual 3D space.

I. INTRODUCTION

Live video streaming is crucial in digital communication, encompassing social media, virtual events, and online education. The advent of 5G networks has heightened the demand for high-quality, low-latency streaming by reducing delays and enhancing data transmission. However, challenges remain in managing large data volumes, synchronizing multiple video streams, and maintaining quality under fluctuating network conditions—issues particularly critical in real-time streaming where even slight delays impair user experience and system performance.

To address these issues, Chen et al. [1] proposed a framework offering solutions for video streaming dialogue. However, implementing 3D virtual environments in live streaming introduces significant obstacles. Real-time processing and rendering of multiple video streams demand substantial computational power, especially in complex virtual settings. Maintaining synchronization across video sources is technically challenging, as consistent timing is vital to avoid desynchronization. Our approach ensures synchronization by rendering all video streams within a unified 3D scene in Unity 3D (U3D), ensuring consistent frame timing during capture, as

detailed in section IV-B. Furthermore, managing low latency with high data volumes from multi-channel setups remains difficult under variable network conditions.

Our approach utilizes 3D virtual environments within engines like U3D to optimize multi-channel live streaming. This method integrates video data from multiple cameras within a virtual 3D space, mapping each stream onto a virtual canvas—a 3D surface displaying video content—and capturing it with an in-world camera. By consolidating multi-camera data into a single stream, our approach reduces complexity and minimizes redundant data transmission. Leveraging U3D for real-time rendering and data management, our method decreases latency and enhances flexibility in video delivery.

Using 3D virtual environments for live streaming offers capabilities beyond traditional methods. Unlike conventional techniques, our approach enables a dynamic, interactive streaming experience where users can select different camera inputs, angles, or perspectives in real-time, creating a more immersive experience. This method supports the demand for interactive content in virtual reality (VR), augmented reality (AR), and mixed reality (MR), where real-time processing and low latency are critical. Consequently, our approach extends its impact to remote collaboration, virtual education, and telemedicine.

While existing methods may reduce latency, they often require significant bandwidth, which can be a bottleneck in less robust network environments. Our framework optimizes video data processing within virtual 3D environments to enhance scalability and reduce computational load. By efficiently mapping video data onto virtual canvases and using in-world cameras to consolidate streams, our method minimizes redundant processing and reduces data volume transmitted across the network. This approach not only reduces latency but also offers a more flexible and scalable solution for multi-channel live streaming. Experimental validation shows our method outperforms existing solutions in latency, user interaction, and system scalability.

We make 4 major contributions:

- Latency reduction of up to 68.7% compared to existing multi-channel live streaming methods.
- Improved scalability by optimizing multi-camera video data processing and transmission.

- Enhanced user interaction through real-time control within a 3D virtual environment, with response times averaging 600ms.
- Broader applicability across VR/AR/MR and remote collaboration, providing a robust solution for diverse industry needs.

II. RELATED WORK

The rapid expansion of live video streaming has driven extensive research to tackle challenges in multi-channel streaming, particularly regarding latency, synchronization, scalability, and video quality.

A. Latency Reduction in Live Streaming

Latency is critical in live streaming, especially for real-time interactions in gaming, telemedicine, and VR/AR/MR. Strategies to minimize latency include integrating Model Predictive Control with deep reinforcement learning, achieving 2 to 5 seconds of latency [2]. However, this is high for real-time applications. The *Learn2Adapt-LowLatency* achieves 1.04 seconds latency without parameter tuning [3], but is insufficient for multi-channel scenarios. Adaptive Bitrate Streaming (ABR) dynamically adjusts video quality to match bandwidth, reducing buffering delays but introducing latency due to quality adjustments [4]–[7]. *MultiLive* reduces end-to-end delay to 100 ms [8], but bandwidth increases with channel count. Recent approaches like *LLL-CAdViSE* [9], *FastEmit* [10], *HxL3* [11], and *LiveNet* [12] focus on ultra-low-latency but require significant resources and are less suitable for multi-channel scenarios. Our method uses U3D to consolidate multiple streams, reducing latency while maintaining synchronization.

B. Synchronization Across Multiple Streams

Synchronization across multiple video streams is a major challenge, as minor discrepancies cause desynchronization artifacts. Traditional methods use timestamps for alignment, but varying network delays pose issues. *MMT-based Multi-channel Video Transmission System* [13] and cloud-edge collaboration [14] enhance synchronization but increase computational complexity and latency. Our framework centralizes synchronization within the 3D virtual environment, ensuring consistent timing with minimal overhead, as detailed in section III-B.

C. Scalability in Multi-Channel Streaming

Scalability is crucial as the number of video streams grows. Traditional architectures struggle to scale efficiently, as each stream adds to computational and bandwidth demands. Scalable Video Coding [15] allows incremental decoding for better bandwidth use but introduces latency and complexity. Edge computing ([16]–[18], etc.) offloads processing closer to data sources, reducing server load and improving responsiveness, but requires substantial infrastructure. Our approach optimizes video data processing within a virtual 3D environment, efficiently managing multiple streams without extensive infrastructure.

D. User Interaction in Real-Time Streaming

User interaction is increasingly important in immersive environments like VR and AR, where real-time control and feedback are crucial. Traditional architectures struggle to provide responsiveness, leading to disruptive delays. Some approaches [19]–[21] reduce these delays by streaming from edge servers, but require significant infrastructure and may not scale across regions. Our framework enhances interaction by integrating real-time control within the 3D virtual environment, allowing immediate updates based on user inputs without extensive infrastructure.

E. Advantages of Our Approach

Despite advancements in multi-channel live streaming, existing methods often compromise on latency, synchronization, scalability, and video quality. Our framework harnesses 3D game engines to consolidate video streams, reducing latency, improving synchronization, enhancing scalability, and maintaining an acceptable quality under varying conditions while enabling real-time user interaction.

III. 3D FRAMEWORK FOR LIVE VIDEO STREAMING

Our method adopts the strategy of multi-channel/multi-input data fusion in 3D space, leveraging the U3D engine to map video streams onto virtual canvases, which are then captured by an in-world camera to create a consolidated data stream. To aid clarity, we define U3D-specific terms: a *virtual canvas* is a 3D surface in the virtual environment displaying video content; a *Raw Image display board* is an object showing a single video stream; an *interactive control group* is a set of interactive elements controlling stream display or user interactions. Our framework addresses latency reduction, synchronization, and scalability in real-time video streaming.

A. Design Principles

The design of our framework is grounded in three core principles: modularity and scalability, low latency, and spatial awareness.

Modularity and Scalability. Our framework is structured with a modular architecture, where each video channel operates independently. Such modularity ensures the system's scalability, allowing it to handle an increasing number of video channels without significant performance degradation. The modular approach also enhances flexibility, enabling live streaming systems to adapt to various operational environments and requirements.

Low Latency. Latency is minimized in multi-channel scenarios using our method, which ensures that the communication latency for multi-channel video streaming is equivalent to that of single-channel streaming. This reduction in latency is achieved by streaming multiple input video streams through a single channel, effectively lowering the communication latency typically associated with multi-channel streaming.

Spatial Awareness. By employing a 3D virtual environment, our framework enhances the processing and display of multiple video streams. Video data is mapped onto a virtual canvas

within the 3D space, and different virtual cameras' capabilities inside U3D are utilized for real-time rendering and manipulation. This spatial approach not only improves user experience but also facilitates more intuitive interactions with multi-channel content.

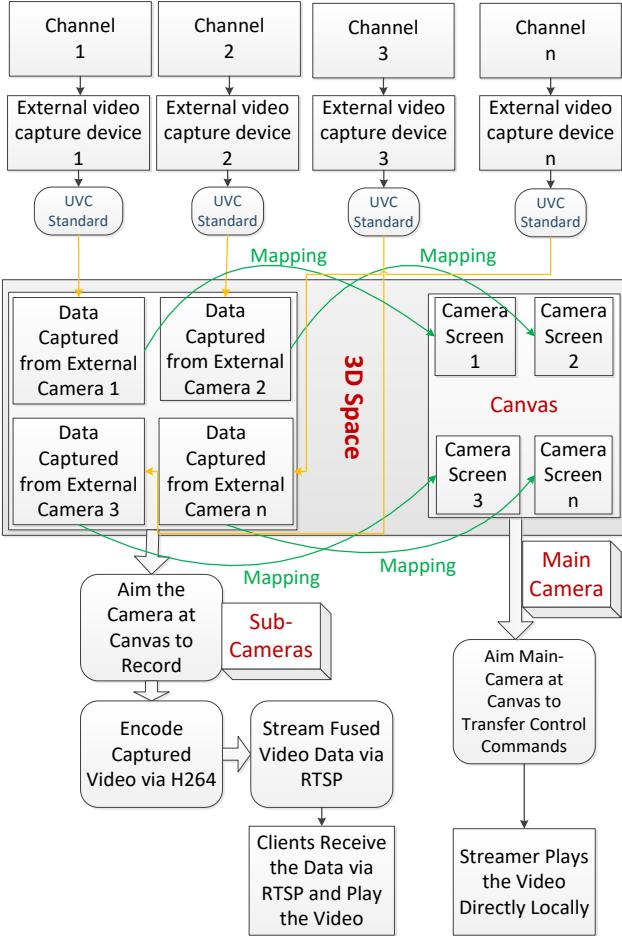


Fig. 1. Our framework leverages multiple input devices to create channels for clients. Video content is seamlessly distributed across canvases in a virtual 3D environment, captured by sub-cameras and then consolidated by a main camera. Users' interactions with the cameras are mapped to an interactive control group, ensuring a responsive experience. The final stream is transmitted to clients via RTSP. To acquire video data from external cameras, we utilize the UVC standard (USB Video Class), which enables plug-and-play functionality across various operating systems.

B. System Architecture

Our framework (Fig. 1) emphasizes the preprocessing of multi-channel data prior to transmission via streaming protocols (e.g., RTSP). Unlike traditional methods, which rely on multiple channels to transmit multi-channel data, our method consolidates the data into a single channel. This is achieved by leveraging the capabilities of virtual 3D components within 3D engines (e.g., U3D), enabling efficient data fusion and transmission while maintaining low latency and high performance. The client directly receives the streamed data via the chosen protocol and plays the video. In our experiment, the RTSP protocol was selected for demonstration purposes. The framework is designed with a modular architecture, allowing developers to easily integrate other protocols such as RTCP,

RTMP, WebRTC, HLS, etc., based on specific application requirements. It ensures that the chosen protocol on the streaming side decouples from our data preprocessing method in the framework, allowing the client-side to receive the streaming data based on the selected protocol. During the preprocessing of the multi-channel input data, each video stream is treated as a separate raw image, which is then mapped onto a virtual canvas within the 3D environment. The canvas is rendered in world space and captured by an in-world main camera to produce a single, consolidated video stream. This approach reduces the complexity associated with handling multiple video streams separately and minimizes redundant data transmission.

C. Core Components

Our approach involves constructing a theoretical model to establish the structure of data transmission, followed by video data preprocessing.

Building the Model. We formulate a model for low-latency live video data transmission, accounting for the relationship between external video capture devices and total streamed channels. Let F denote the number of live content channels and n the number of external video capture devices. Under USB-3.0 specifications, a single interface supports up to 256 devices [22], so $n \in [0, 255]$.

Consider a main camera S within a virtual 3D scene. A virtual canvas c holds Raw Image display boards (ρ), each corresponding to a video stream from a single device, representing a non-interactive channel. The total non-interactive channels on a canvas are:

$$c(\rho) = \sum_{n=0}^{255} \rho_n \quad (1)$$

Each interactive control group (β) on the canvas corresponds to an interactive board controlling stream display. Since each interactive board is paired with a Raw Image, the total channels on a single canvas are:

$$c = c(\rho) + c(\beta) = \sum_{n=0}^{255} \rho_n + \sum_{n=0}^{255} \beta_n = \sum_{n=0}^{255} (\rho_n + \beta_n) \quad (2)$$

The scene includes two canvases: c_1 for video streams and c_2 for user interaction observation. The main camera S captures both, so $S(c) = c_1 + c_2$. The total channels F are:

$$F(n) = S \left(c_1 \left(\sum_{n=0}^{255} \rho_n \right) + c_2 \left(\sum_{n=0}^{255} \beta_n \right) \right) \quad (3)$$

This model ensures multi-channel streaming achieves single-channel latency by consolidating streams before transmission. Equation (3) is derived by aggregating the channels from both canvases, captured by S , which fuses video and interactive data into a single stream.

Video Data Preprocessing. The preprocessing component captures, renders, and encodes video streams within the 3D space. It creates Raw Image objects for each video channel, merges them onto a virtual canvas, and renders the canvas

in world space, ensuring efficient multi-channel rendering and spatial transmission. Adaptive video encoding/decoding can be integrated, and functionalities like peer-to-peer communication, broadcasting, and text messaging are supported. Algorithm 1 details the preprocessing and fusion process.

Algorithm 1 Streaming Side Setup

```

Set up Recording Area:
1:  $\mathcal{S}_{3D} \leftarrow \text{InitNewScene}()$            ▷ Init 3D scene
2:  $\text{SetRootGameObjects}(\mathcal{S}_{3D})$           ▷ Set scene hierarchy
3:  $\mathcal{C}_{ws} \leftarrow \text{CreateCanvas}(\text{WorldSpace}, \text{ScreenSize})$  ▷ Create canvas
4:  $\mathcal{C}_{wsc} \leftarrow \text{AddWorldSpaceCamera}(\mathcal{C}_{ws})$     ▷ Add canvas camera
5:  $\mathcal{H}_{cameras} \leftarrow \text{AddMultiCamHolder}(\mathcal{C}_{wsc})$  ▷ Add multiple cameras

Set up Interaction Observer Module:
6:  $\mathcal{C}_{interaction} \leftarrow \text{SetInteractionCanvas}(\text{RenderMode} \leftarrow \text{ScreenSpaceOverlay}, \text{ScreenSize})$       ▷ Set interaction canvas
7:  $\text{Duplicate}(\mathcal{H}_{cameras}.\text{Pages}) \rightarrow \mathcal{C}_{interaction}$       ▷ Copy camera pages

Fuse All Cameras' Data into 1 Channel:
8:  $\text{Data}_{fused} \leftarrow \text{WebcamVideoPlayer.Fuse}()$       ▷ Init data fusion
   for  $i = 1$  to  $n$  do
      $\text{Data}_{fused} \leftarrow \text{Data}_{fused} + \mathcal{C}_{interaction}(i)$       ▷ Add interaction data
   end for
   for  $j = 1$  to  $n$  do
      $\text{Data}_{fused} \leftarrow \text{Data}_{fused} + \mathcal{H}_{cameras}(j)$       ▷ Add camera data
   end for
9:  $\text{RTSP.Send}(\text{Data}_{fused})$           ▷ Send fused stream

```

Our framework leverages U3D to optimize multi-channel video data processing and transmission. By mapping streams onto virtual canvases and consolidating them via in-world cameras, it minimizes redundant processing and network data volume, enhancing latency reduction and scalability for applications like virtual events and VR/AR/MR content delivery.

IV. EVALUATION

To validate our framework, we conducted comprehensive evaluations focusing on four critical performance metrics: latency, synchronization accuracy, scalability, and user interaction responsiveness. We compared our approach to existing works to contextualize performance. Evaluations were conducted in a stable campus network environment with 73.65 Mbps upload and 54.06 Mbps download bandwidth, aligning with Wi-Fi 6 capabilities.

A. Latency

Latency, the delay between video capture and playback ($\text{Latency} = t_{\text{render}} - t_{\text{capture}}$), is key for real-time applications [23]. We varied input cameras from 1 to 10 (Fig. 2),

showing consistent latency (~230 ms) regardless of channel count, confirming effective stream consolidation.

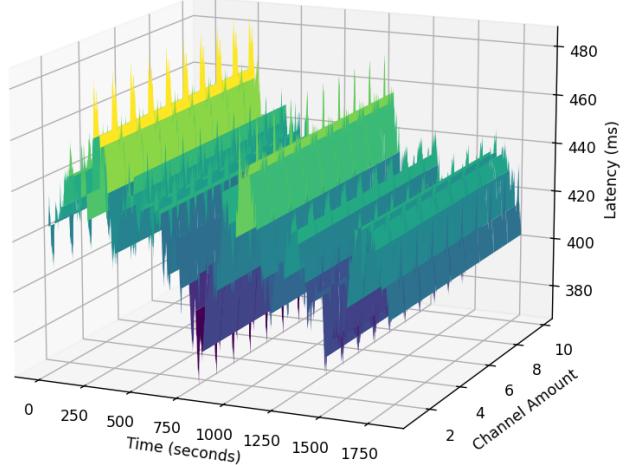


Fig. 2. Latency measured over a 30-minute period with different numbers of input camera devices functioning as separate channels for live video streaming.

We compared our approach with baselines: Sun et al. [2] (2314 ms), Learn2Adapt-LowLatency [3] (1040 ms), and Zhang et al. [24] (734 ms). Our method achieved a 68.7% latency reduction over Zhang et al. for single-channel streaming (table I). Unlike baselines, where latency increases with channels, our framework maintained consistent performance.

TABLE I
LATENCY COMPARISON WITH BASELINE SCHEMES (MS).

Channel Count	Single	Multi-Channel (5)	Multi-Channel (10)
Sun et al. [2]	2314	3462	5645
Learn2Adapt [3]	1040	2085	3203
Zhang et al. [24]	734	1788	3867
Ours	230	231	235

These findings highlight the effectiveness of our framework's design, which consolidates multiple video streams efficiently, minimizing delays and providing low-latency performance across a range of scenarios.

B. Synchronization Accuracy

Synchronization accuracy measures the time difference between corresponding frames from multiple video streams. Let t_{channel_1} and t_{channel_2} denote the timestamps of frames from two different channels rendered on the client. The synchronization offset is calculated as $\Delta t = |t_{\text{channel}_1} - t_{\text{channel}_2}|$. Since all channels are merged into a single stream, Δt is zero.

C. Scalability

Scalability is evaluated by measuring the system's performance as the number of video streams increases. Let n represent the number of streams, namely the number of input devices, and $L(n)$ be the latency as a function of n . The system's scalability is considered strong if $L(n)$ stays almost stable within a tolerable range with increasing n . The computational load on the system (e.g., CPU/GPU consumption) is

denoted by $U_{CPU}(n)$ and $U_{GPU}(n)$, and we expect these to increase linearly: $U_{CPU}(n), U_{GPU}(n) \propto n$.

Our evaluation (Fig. 3) on latency demonstrates that our method (230.29ms) surpasses [3] (1040ms) in single-channel live video streaming significantly, while also outperforming some traditional methods ([25]–[28]), as well as more recent methods ([8], [29], [30]). Notably, as the number of input devices increases, our approach maintains consistent bandwidth, showcasing its scalability and robustness. This is due to its design, which merges multiple channels into one before the streaming process.

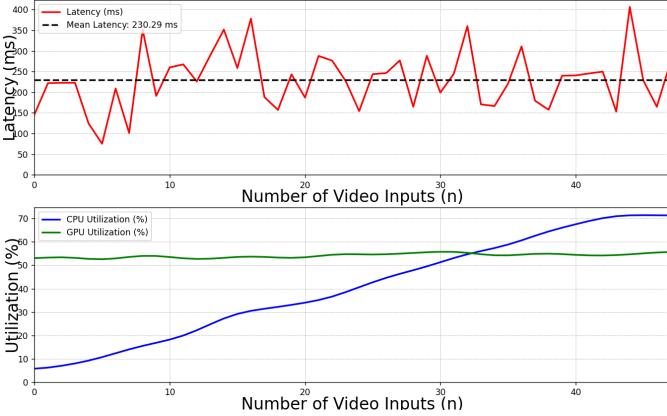


Fig. 3. Evaluation of system performance as the number of input devices increases to 50. The average latency is 230.29 ms. GPU consumption remains stable with minimal fluctuations, while CPU consumption rises progressively with more input devices. The GPU consumption stays almost still because the online-rendering is conducted if only the video is displayed. This indicates efficient GPU resource management and increasing CPU demands. Testing was conducted on machines with an NVIDIA GeForce GTX 1050 GPU and an Intel Core i7-7700HQ CPU.

D. User Interaction Responsiveness

User interaction responsiveness is measured by the time delay between user input and the corresponding action in the video stream. Let t_{input} be the time of user input, and t_{action} be the time when the action is reflected in the live stream. The responsiveness is then defined as: $ResponseTime = t_{action} - t_{input}$. Lower response times indicate better interaction responsiveness.

In the experiment, we performed over 500 interactions across 9 modules of our live streaming system implemented in U3D and recorded the corresponding interaction data (Fig. 4).

These results demonstrate our framework's responsiveness and robustness across diverse conditions. Our method achieves efficient response times, ensuring a smooth user experience. Its computational efficiency, indicated by stable GPU usage and linear CPU growth, validates its scalability and resource optimization for modern applications.

V. LIMITATION

Despite the notable advantages demonstrated by our multi-channel live streaming framework, certain limitations warrant consideration. A key trade-off observed lies in the reduction of video quality under multi-channel conditions. While the framework efficiently fuses data streams to optimize transmission

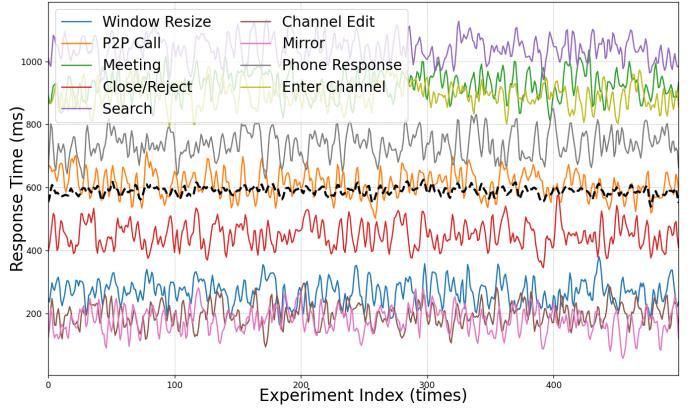


Fig. 4. Nine modules were selected for evaluation, with their response times predominantly ranging from 50 ms to 1300 ms. The mean response time, represented by a dashed line, is approximately 600 ms. Given the significant fluctuations observed in the recorded data, Gaussian smoothing [31] was applied to reduce noise and variability, utilizing a parameter (σ) of 1 for visualization purposes.

bandwidth, this design inherently results in a loss of quality as the number of input channels increases. Specifically, the consolidation process maintains fixed video sizes for transmission, leading to a proportional decline in visual fidelity with an increasing number of channels when end-users simultaneously request the display of multiple channels.

This limitation reflects the inherent compromise between maintaining low latency and achieving optimal video quality, particularly in scenarios involving a large number of channels associated with requests for multiple displays at the same time. While our framework achieves superior performance metrics in latency, synchronization, scalability, and user interaction responsiveness, the reduction in video quality may pose challenges in use cases where high-definition output is critical, such as medical imaging or high-resolution broadcasting.

Future work should explore adaptive encoding techniques or advanced compression algorithms tailored for multi-channel environments to mitigate this issue. Additionally, implementing quality-preservation strategies, such as dynamic resource allocation or channel prioritization based on content importance, may enhance the framework's overall effectiveness and broaden its applicability in quality-sensitive domains.

VI. CONCLUSION

Our 3D framework for enhancing low-latency multi-channel live video streaming, which integrates video within a virtual 3D space, demonstrates advantages over traditional live streaming methods across several key performance metrics. This framework achieves considerable reductions in latency, high synchronization accuracy, and robust scalability. These findings confirm the effectiveness of incorporating multi-channel video streams within a 3D virtual environment and underscore the potential of this approach for various multi-channel live video streaming applications, particularly in contexts that demand low-latency video delivery.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education of the People's Republic of China under the Industry-University Cooperative Education Program (BJUT Contract No. 40059000202520) through the New Engineering Construction Project Cooperation Agreement.

REFERENCES

- [1] Joya Chen, Zhaoyang Lv, Shiwei Wu, Kevin Qinghong Lin, Chenan Song, Difei Gao, Jia-Wei Liu, Ziteng Gao, Dongxing Mao, and Mike Zheng Shou, "Videollm-online: Online video large language model for streaming video," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 18407–18418.
- [2] Liyang Sun, Tongyu Zong, Siquan Wang, Yong Liu, and Yao Wang, "Towards optimal low-latency live video streaming," *IEEE/ACM transactions on networking*, vol. 29, no. 5, pp. 2327–2338, 2021.
- [3] Theo Karagioules, Rafał Mekuria, Dirk Griffioen, and Arjen Wagenaar, "Online learning for low-latency adaptive streaming," in *Proceedings of the 11th ACM Multimedia Systems Conference*, 2020, pp. 315–320.
- [4] Farzad Tashtarian, Abdelhak Bentaleb, Hadi Amirpour, Sergey Gorinsky, Junchen Jiang, Hermann Hellwagner, and Christian Timmerer, "{ARTEMIS}: Adaptive bitrate ladder optimization for live video streaming," in *21st USENIX Symposium on Networked Systems Design and Implementation (NSDI 24)*, 2024, pp. 591–611.
- [5] Muhammad Hamza Bin Waheed, Faisal Jamil, Amir Qayyum, Harun Jamil, Omar Cheikhrouhou, Muhammad Ibrahim, Bharat Bhushan, and Habib Hmam, "A new efficient architecture for adaptive bit-rate video streaming," *Sustainability*, vol. 13, no. 8, pp. 4541, 2021.
- [6] Bekir Oguzman Turkkan, Ting Dai, Adithya Raman, Tevfik Kosar, Changyou Chen, Muhammed Fatih Bulut, Jaroslaw Zola, and Daby Sow, "Greenabr: Energy-aware adaptive bitrate streaming with deep reinforcement learning," in *Proceedings of the 13th ACM Multimedia Systems Conference*, 2022, pp. 150–163.
- [7] Zhengfang Duanmu, Wentao Liu, Zhuoran Li, Diqi Chen, Zhou Wang, Yizhou Wang, and Wen Gao, "Assessing the quality-of-experience of adaptive bitrate video streaming," *arXiv preprint arXiv:2008.08804*, 2020.
- [8] Ziyi Wang, Yong Cui, Xiaoyu Hu, Xin Wang, Wei Tsang Ooi, Zhen Cao, and Yi Li, "Multilive: Adaptive bitrate control for low-delay multi-party interactive live streaming," *IEEE/ACM Transactions on Networking*, vol. 30, no. 2, pp. 923–938, 2021.
- [9] Babak Taraghi, Hermann Hellwagner, and Christian Timmerer, "Lliv-adviser: live low-latency cloud-based adaptive video streaming evaluation framework," *IEEE Access*, vol. 11, pp. 25723–25734, 2023.
- [10] Jiahui Yu, Chung-Cheng Chiu, Bo Li, Shuo-yiin Chang, Tara N Sainath, Yanzhang He, Arun Narayanan, Wei Han, Anmol Gulati, Yonghui Wu, et al., "Fastemit: Low-latency streaming asr with sequence-level emission regularization," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 6004–6008.
- [11] Farzad Tashtarian, Abdelhak Bentaleb, Alireza Erfanian, Hermann Hellwagner, Christian Timmerer, and Roger Zimmermann, "Hxl3: Optimized delivery architecture for http low-latency live streaming," *IEEE Transactions on Multimedia*, vol. 25, pp. 2585–2600, 2022.
- [12] Jinyang Li, Zhenyu Li, Ri Lu, Kai Xiao, Songlin Li, Jufeng Chen, Jingyu Yang, Chunli Zong, Aiyun Chen, Qinghua Wu, et al., "Livenet: a low-latency video transport network for large-scale live streaming," in *Proceedings of the ACM SIGCOMM 2022 Conference*, 2022, pp. 812–825.
- [13] Yasuhiro Mochida, Takahiro Yamaguchi, and Ken Nakamura, "Mmt-based multi-channel video transmission system with synchronous processing architecture," in *2020 IEEE Symposium on Low-Power and High-Speed Chips (COOL CHIPS)*. IEEE, 2020, pp. 1–3.
- [14] Zhen Guo, Pengzhou Zhang, Junjie Xia, Zhe Zhou, and Juan Cao, "Multi-channel live video processing method based on cloud-edge collaboration," in *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2022, pp. 1404–1409.
- [15] Z. Shahid, M. Chaumont, and W. Puech, "Scalable video coding," in *Effective Video Coding for Multimedia Applications*, Sudhakar Radhakrishnan, Ed., chapter 1. IntechOpen, Rijeka, 2011.
- [16] Muhammad Ali, Ashiq Anjum, Omer Rana, Ali Reza Zamani, Daniel Balouek-Thomert, and Manish Parashar, "Res: Real-time video stream analytics using edge enhanced clouds," *IEEE Transactions on Cloud Computing*, vol. 10, no. 2, pp. 792–804, 2020.
- [17] Xiaoteng Ma, Qing Li, Longhao Zou, Junkun Peng, Jianer Zhou, Jimeng Chai, Yong Jiang, and Gabriel-Miro Muntean, "Qava: Qoe-aware adaptive video bitrate aggregation for http live streaming based on smart edge computing," *IEEE Transactions on Broadcasting*, vol. 68, no. 3, pp. 661–676, 2022.
- [18] Nianzhen Gao, Jiaxi Zhou, Guoan Wan, Xinhai Hua, Ting Bi, and Tao Jiang, "Low-latency vr video processing-transmitting system based on edge computing," *IEEE Transactions on Broadcasting*, 2024.
- [19] Selin Nacaklı and A Murat Tekalp, "Controlling p2p-cdn live streaming services at sdn-enabled multi-access edge datacenters," *IEEE Transactions on Multimedia*, vol. 23, pp. 3805–3816, 2020.
- [20] Reza Farahani, Mohammad Shojafar, Christian Timmerer, Farzad Tashtarian, Mohammad Ghanbari, and Hermann Hellwagner, "Ararat: A collaborative edge-assisted framework for http adaptive video streaming," *IEEE Transactions on Network and Service Management*, vol. 20, no. 1, pp. 625–643, 2022.
- [21] Gang Shen, Jianhui Dai, Hassnaa Moustafa, and Lei Zhai, "5g and edge computing enabling experience delivery network (xdn) for immersive media," in *2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR)*. IEEE, 2021, pp. 1–7.
- [22] Lambert M. Surhone, Mariam T. Tennoe, and Susan F. Henssonow, *USB 3.0*, Betascript Publishing, Beau Bassin, MUS, 2011.
- [23] Mikko Uitto and Antti Heikkilä, "Evaluation of live video streaming performance for low latency use cases in 5g," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuC-NC/6G Summit)*. IEEE, 2021, pp. 431–436.
- [24] Bo Zhang, Thiago Teixeira, and Yuriy Reznik, "Performance of low-latency http-based streaming players," in *Proceedings of the 12th ACM Multimedia Systems Conference*, 2021, pp. 356–362.
- [25] Yongtao Shuai, Manuel Gorius, and Thorsten Herfet, "Low-latency dynamic adaptive video streaming," in *2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*. IEEE, 2014, pp. 1–6.
- [26] Viswanathan Swaminathan and Sheng Wei, "Low latency live video streaming using http chunked encoding," in *2011 IEEE 13th International Workshop on Multimedia Signal Processing*. IEEE, 2011, pp. 1–6.
- [27] Sheng Wei and Viswanathan Swaminathan, "Low latency live video streaming over http 2.0," in *Proceedings of Network and Operating System Support on Digital Audio and Video Workshop*, 2014, pp. 37–42.
- [28] Serhan Güç, Dimitri Podborski, Jangwoo Son, Gurdeep Singh Bhullar, Thomas Buchholz, Thomas Schierl, and Cornelius Hellge, "Cloud rendering-based volumetric video streaming system for mixed reality services," in *Proceedings of the 11th ACM multimedia systems conference*, 2020, pp. 357–360.
- [29] Kyungjin Lee, Juheon Yi, Youngki Lee, Sunghyun Choi, and Young Min Kim, "Groot: a real-time streaming system of high-fidelity volumetric videos," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.
- [30] Jaehong Kim, Youngmok Jung, Hyunho Yeo, Juncheol Ye, and Dongsu Han, "Neural-enhanced live streaming: Improving live video ingest via online learning," in *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication*, 2020, pp. 107–125.
- [31] Kazufumi Ito and Kaiqi Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE transactions on automatic control*, vol. 45, no. 5, pp. 910–927, 2000.