Towards Holographic Video Communications in 5G Networks: A Promising AI-driven Solution

Yakun Huang, Yuanwei Zhu, Xiuquan Qiao, Xiang Su, *Member, IEEE,*
Schahram Dustdar, *Fellow, IEEE,* Ping Zhang, *Fellow, IEEE*

***Abstract*—Real-time holographic video communications enable immersive experiences for next-generation video services in the future metaverse era. However, high-fidelity holographic videos require large bandwidth and significant computation resources, which exceed the transmission and computing capacity of 5G networks. This article reviews state-of-the-art holographic video communications techniques and highlights the critical challenges of delivering such immersive services. We further implement a preliminary prototype of an AI-driven holographic video communication system and present some critical experimental results to evaluate its performance. Finally, we discuss some potential use cases and identify future research directions for provisioning real-time and high-quality holographic experiences.**

I. Introduction

Holographic media (e.g., 3D mesh and point cloud) evokes human beings to expect and aspire for an immersive service that profoundly integrates the virtual and real worlds [1]. In particular, the point cloud describes the objects only using a set of unstructured 3D points with coordinates and color without topological information, which is more flexible and straightforward than the 3D mesh. When compared to conventional immersive content delivery, the massive volume of point cloud video streamings (e.g., capturing one-second of raw point cloud video with one depth camera at 30 FPS (Frames Per Second) produces 2.06 Gb of data) poses the following challenges to existing network service infrastructures including: (1) the challenge of network transmission capacity, i.e., holographic point cloud video transmission requires a bandwidth capacity that is more than Gbps level, which is far beyond the current transmission capacity of 5G networks, and even future 6G networks with both breadth and depth of multi-layer coverage still cannot fulfill the requirements in some scenarios; and (2) the challenge of adaptive video streaming techniques. Traditional technologies such as ABR (Adaptive Bit Rate) [2] have limited effects on optimizing the transmission of vast amounts of holographic video.

We investigate the existing point cloud video transmission for holographic media in Table I, including point cloud compression and video streaming optimization. Traditional compression methods include Kdtree-based and Octree-based solutions, such as popular PCL (Point Cloud Library) [3] and Draco [4]. Some deep learning-based compression methods provide lower accuracy loss and higher compression ratios [5], [6]. Unfortunately, these methods are only applicable for offline holographic video pre-processing and compression due to the required high memory and computational overhead. Besides, most video streaming optimization methods expand solutions used in VR (Virtual Reality) and 360-degree video streaming for encoding, tiling, and view angle prediction. For encoding, although ISO/IEC MPEG (Moving Picture Experts Group) provides two international standards of V-PCC (Video-based Point Cloud Compression) and G-PCC (Geometry based Point Cloud Compression) for point cloud video encoding, both methods require higher computational resources and costs than conventional encoding methods. Compared to 3-DoF 360-degree/VR videos, point cloud video adds 3-DoF location information, requiring the adaptive adjustment of the quality of the point cloud video stream with the dynamic change in physical distance between the user and the scene. This also leads to more complex and challenging user motion trajectory and perspective prediction.

TABLE I: Techniques for point cloud video transmission.

	Methods	Advantages	Disadvantages
Point Cloud Compression	Kdtree-based/ Octree-based	Easy-to-use	High computing time
	PointNet++/ PU-GAN	High compression ratio	High computing resources
Video Streaming Optimization	V-PCC/ G-PCC	Extension of MPEG	High cost & low performance
	Tiling/ View angle prediction	Extension of VR/360-degree	Poor adaptability

Some research investigates the combination of point cloud compression and transmission optimization [7], [8]. For instance, PCC-DASH [9] explores the dynamic adaptive point cloud streaming with several bitrate adaptation heuristics. However, these solutions cannot be run in real-time on contemporary hardware due to the massive cost of video compression. This article introduces the landscape and characteristics of holographic point cloud video communication and analyses the technical challenges associated with supporting holographic services in 5G networks. We start with presenting the architecture and workflow of an immersive media service for typical

Y. Huang, Y. Zhu, X. Qiao and P. Zhang are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China. E-mail:{ykhuang, zhuyw, qiaoxq, pzhang}@bupt.edu.cn.

X. Su is with the Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway and University of Oulu, 90570, Oulu, Finland. Email:xiang.su@ntnu.no.

S. Dustdar is with the Distributed Systems Group, Technische Universität Wien, 1040 Vienna, Austria. E-mail:dustdar@dsg.tuwien.ac.at.

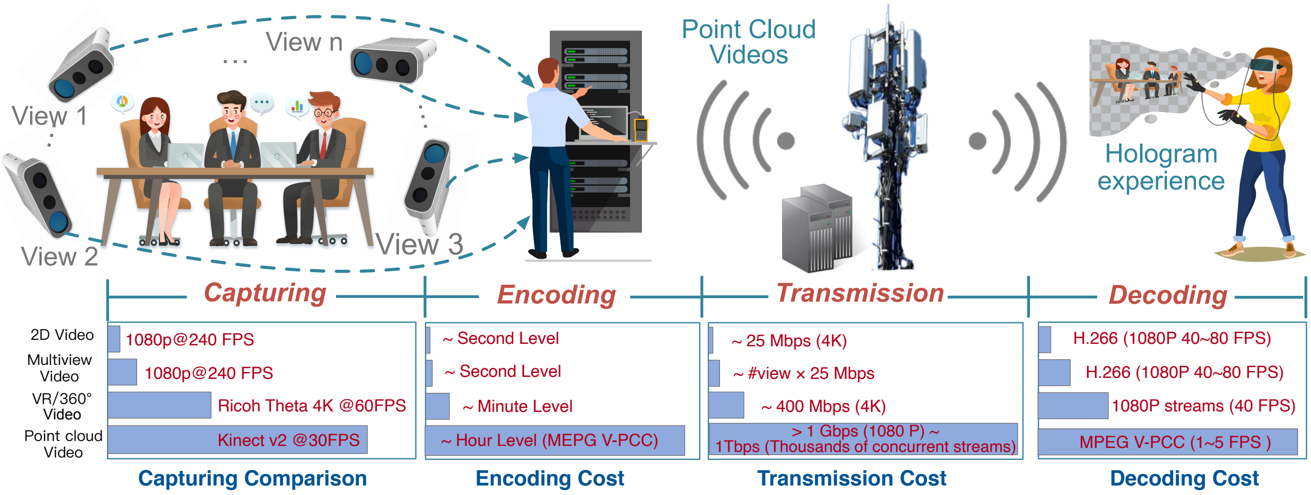

Fig. 1. The workflow of typical holographic video communications and comparisons with other video services.

holographic video communications. Afterwards, we propose an AI-driven holographic transmission solution as a prototype for preliminary exploration and verify its performance through experimental simulations. Finally, we discuss the proposed AI-driven communication technique and identify future directions for high-quality holographic video communication services.

II. Holographic Communication: characteristics, challenges

A. Characteristics

Fig. 1 presents a typical holographic video communication system consisting of data capturing, raw data fusing and encoding, data transmission, data decoding, and video rendering [10]. We compare the holographic point cloud video and alternative types of video services in terms of computation and transmission overheads at different stages to better understand the characteristics of holographic video communication. We observe that holographic video services clearly require higher network bandwidth and computing resources than traditional videos, especially requiring at least 1 Gbps and reaching 1 Tbps with thousands of concurrent streams.

Holographic video production requires capturing as many natural scenes as possible from different angles and merging these raw video streamings in real-time with a high performance edge server. Thus, the capturing requires abundant computation resources, including GPUs, to accelerate the real-time generations of holographic video streams. For fusing and encoding, we observe that the fusion of various views and extended MPEG encoding methods is highly time- and resource-consuming to provide efficient processing. For real-time transmission, in addition to provide basic network communication services with large bandwidth and low transmission delay, adaptive point cloud streaming technology is crucial to effectively guarantee the QoE (Quality of Experience). The primary method is the expansion of VR and 360-degree video streaming. Holographic terminals execute video decoding, rending, interaction commands, etc. In particular, decoding is a computationally intensive and time-consuming step, an essential factor affecting rendering, interaction experience, and energy consumption.

Undoubtedly, holographic video far exceeds the requirements of traditional video streaming services in terms of network bandwidth, transmission latency, and computational complexity. Although the interactive terminal for holographic video is theoretically compatible for usage on conventional PC, smartphones, and HMDs (Head-Mounted Displays), users have to compromise on the point cloud resolution to mitigate the resource consumption required for network bandwidth and decoding computation.

B. Challenges

Holographic communication poses significant demands on network transmission infrastructure, such as requiring bandwidth, ultra-low delay, ultra-complex computation, and mobility and ubiquity of devices.

1) Ultra-high bandwidth. Volumetric video streaming leads to skyrocketing network bandwidth demands at the Gbps level. Take the example of using a depth camera such as the Microsoft Azure Kinect to capture holographic contents and for interacting with AR/VR terminals. The camera outputs an RGB 1080PP image (1920×1080) and a depth image (512×424). The color data and depth data per pixel are represented by 4 bytes and 2 bytes, respectively [11]. Thus, streaming one-second raw point cloud video with one depth camera at 30 FPS requires at least 2.06 Gbps of transmission bandwidth. Besides, compared with 360-degree and VR video, holographic video provides 6-DoF (degree of freedom) experiences, including 3-DoF of translational movement (X, Y, Z) and 3-DoF of rotational movement (yaw, pitch, roll). This requires numerous depth cameras to capture data and incurs more data volume than other types of videos. It is still unavailable to high-fidelity holographic videos captured with dozens of depth cameras that exceed the 5G network's bandwidth using the enhanced Mobile Broadband (eMBB) at a 10 Gb/s rate.

2) Ultra-low delay. Holographic video communications present more stringent delay requirements than other delay-tolerant services. One of the most crucial factors that impairs the experience is the MTP (Motion-To-Photon) latency. MTP should be below the human perceptible limit and allow users to interact with holograms directly and seamlessly. High latency may result in virtual objects lagging behind the intended position, causing dizziness, especially using HMDs. When comparing 6-DoF point cloud video with 360-degree or VR videos, 6-DoF movement and orientations make the holographic video more sensitive to latency than the 3-DoF services. In particular, the ideal delay recommended for the streaming of a holographic point cloud video is less than 5 ms, which is more stringent than the latency requirement for 360-degree or VR videos (i.e., <20 ms) [12]. Therefore, more efficient transport protocols or physical deployment solutions need to be proposed to alleviate dizziness caused by high transmission delay.

3) Ultra-complex computation. Capturing a complete holographic video stream, then transmitting, and finally rendering in high quality on mobile terminals requires multiple steps in the pipeline. First, the raw point cloud streams from different multi-view cameras should be synchronized, spliced, and aligned into a complete video stream. Compared to the 360-degree and VR videos that deal with 2D pixels, the point cloud video uses 3D voxels or unstructured points, incurring extra computational overhead. Second, the huge amount of raw holographic video streams introduces high computing costs to compress and encode the streams at the sender along with decoding and rendering at the receiver. For example, encoding a one-second video from the *longdress* dataset with lossy compression requires 11 to 42 minutes with MPEG V-PCC on a generic desktop computer [13].

4) Mobility and ubiquity of devices Device mobility and ubiquity also introduce challenges for holographic video communication. Current holographic communication scenarios force users to stand or sit in a fixed position. Around them, there is an array of multi-view cameras capturing the central object (e.g., Google Relightables). Although the price of such devices for acquisition has gradually become cheaper, guaranteeing a complete hologram usually requires at least nine cameras which far outnumbers that of 360-degree and VR video. To the best of our knowledge, there are only a few outdoor mobile scenarios using holographic communication, such as recording a user's movement when running, skiing, and performing other sports activities.

Besides, dedicated devices (e.g., HMDs and AR glasses) for rendering and interacting with holographic services are far less prevalent than smartphones. Meanwhile, the user requirement of portability is also hard to meet. Although the HMDs or AR glasses (e.g., Nreal, HoloLens) avoid occupying one user's hands and tracks the direction of the user's FoV (Field of View), the limited FoV available on HMDs is far smaller than the human vision, negatively influencing user experiences. Moreover, bulky HMDs occlude the user's peripheral vision, making them unaware of incoming dangers from their physical surroundings. Thus, we conjecture that overcoming this weakness will become the critical turning point of popularizing holograms in the real world.

III. Progressive real-time delivery: A novel AI-driven Solution

Intuitively, compression requires high computing cost and introduces high processing latency even with enhanced 2D projection-based and 3D tree-based methods. We propose a novel AI-driven solution that defines point cloud video streaming as an end-to-end neural network training problem entirely different from existing work. Our general idea is that each point in an arbitrary point cloud set contributes its feature to the whole content, but not all point features are key features. Then, we can reconstruct the original point cloud content with key features by an AI generating technique. Fig. 2 presents the overall AI-driven transmission system for point cloud video service, consisting of a generating module, extraction module, device-side feature reconstruction, and online adapter. Afterwards, we describe each module of AI-based holographic point cloud video transmission separately to better help understanding.

For generating point cloud video, we deploy multiple depth cameras at different angles to capture video streams containing RGB and depth maps, and then convert them to the point cloud format. Then, we align and fuse various point cloud frames to obtain a complete 6-DoF point cloud video frame according to the camera's pose and other parameters. The key technologies involved in generating point cloud video are computer vision and 3D reconstruction. This article focuses on how to provide efficient and adaptive transmission services for generated point cloud videos. For transmission, the proposed AI-based method extracts key features of the point cloud video to significantly reduce redundant data transmission, thus enabling real-time transmission. The feature extraction module takes each frame as input and outputs critical features at the intermediate layer for real-time transmission. Specifically, we use the hierarchical extraction structure of the PointNet++ [5] and employ an ensemble abstraction layer to capture the local structure from the original point cloud. The raw input frame is represented by fewer points and features when outputting a point-by-point feature matrix. Thus, we can use an AI-based deep extracting technique to transfer key features instead of the raw point cloud frames.

For the reconstruction module, since the end device receives the key features, directly decoding and recovering from these feature points for rendering is impossible. Besides, these feature points lose much detailed information compared with the original point cloud frames. We propose a lightweight GAN-based point cloud reconstruction network based on an AI-based generative technique. We design an upsampling-downsampling-upsampling layer [6] to generate more diverse point distributions and enhance the feature variations rather

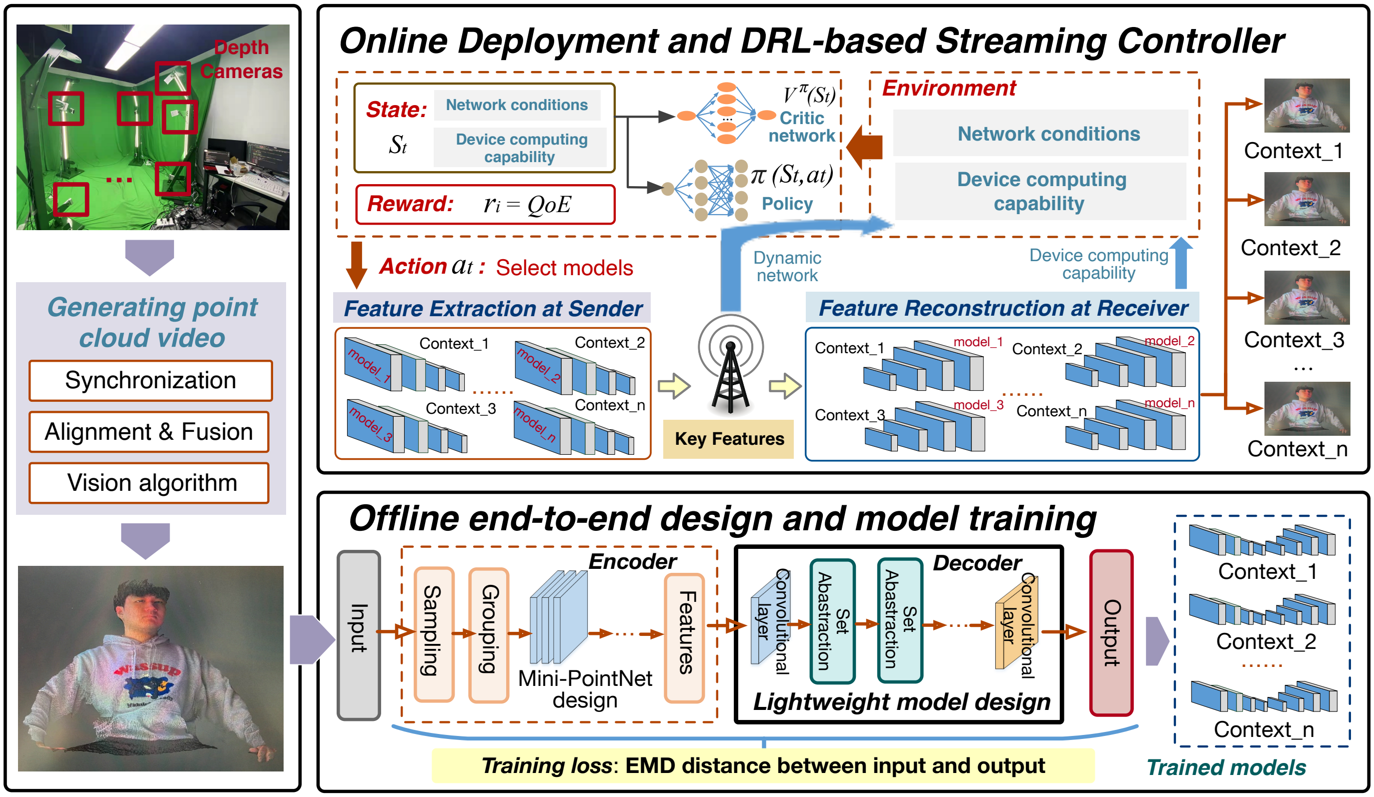

Fig. 2. Proposed AI-driven end-to-end network architecture.

than a simple duplication strategy. Therefore, feature extraction, and feature reconstruction can be regarded as the encoding and decoding processes, respectively. More importantly, unlike traditional video encoding and decoding, which are relatively independent processes, AI-based processing requires offline joint pre-training to obtain the best-fitting feature extraction and reconstruction modules. We use EMD (Earth mover's distance) as the loss function, to produce generated points located on the target surface similar to the original input.

The online adapter adaptively chooses different extraction-reconstruction models for dynamic contexts rather than controlling bit rate in traditional streaming. We train various encoding-decoding model pairs with different compression ratios to match various networks. In other words, we fuse the encoding and decoding processes into the adaptive stream controlling of point cloud video. To this end, we define a hyper-parameter to adjust the size of transmission features, representing the compression ratio to match dynamic networks. Then, we also train multiple AI models with different hyper-parameters for various networks. Thus, satisfying the maximum QoE in dynamic networks requires a special transmission control algorithm different from conventional adaptive methods for such an AI-driven neural network transmission mechanism. To address this challenge, we propose an online self-learning neural network adapter based on DRL (Deep Reinforcement Learning), providing optimal feature extraction and recovery model for different networks. Specifically, we construct the self-learning adapter by taking the current network condition and the demands of the state, defining the reward, and selecting the matching codec neural network model as the action for DRL policy network training. We show the state, the action, and the reward defined in the DRL-based online adapter in Fig. 2. The reward plays an essential role in achieving fast convergence and obtaining the optimal global solution. Based on the experience, we define the QoE as the reward considering the transmission latency and reconstruction accuracy, which can be calculated as $QoE = [2 \cdot (1/d) \cdot (1/t)]/[(1/d) + (1/t)] = 2/(d + t)$. a and t are the reconstruction accuracy and transmission latency, respectively. We use the Chamfer Distance d between the original point cloud and the reconstructed point cloud to measure the reconstruction accuracy. Here, we normalize the values of these two objectives between 0 and 1.

IV. Experimental Analysis

The deep neural network described in Section III is a high-level framework. We design *feature extraction* and *feature expansion* modules to form an end-to-end network structure. This section provides an experimental analysis of our designed architecture that addresses the challenge of the ability to progressively deliver point coordinates (x, y, z) with a high compression ratio to reduce the overall data volume. Our specific architecture adopts the hierarchical extraction module based on the backbone of PointNet++ [5] as the *feature extraction*, and adopts the feature expansion component and point set generation component in the generator of PU-GAN

[6] as the *feature expansion*. More details and parameter settings can be found in [14].

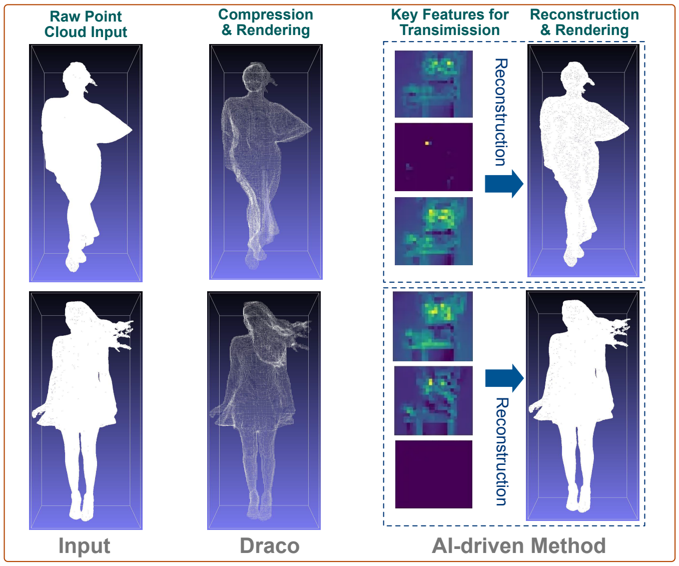

Fig. 3. Qualitative comparisons on the reconstruction results.

To improve the generalization of the AI-driven model, we train our end-to-end neural network with utilizing 147 3D point cloud objects [6]. The dataset includes a rich variety of objects, from simple objects (e.g., Icosahedron) to high-detailed objects (e.g., Statue). In addition, we use four real-world point cloud video sequences for testing [15], each of which is a human body captured by 42 RGB cameras at 30 FPS over a 10 s period. Due to space constraints, we select the two typical longdress and redandblack datasets to show the reconstruction performance in Fig. 3. Due to the unity of the input dimension of the deep neural network, we decompose each point cloud frame into multiple patches of the same size in advance. For example, we group each patch with 256 points and normalize them in a unit sphere. Then, we compress the patch (256,3) (i.e., 256 points with 3D coordinates) into a (5,5) feature vector matrix. We compare our method with Draco [4], where the compression level parameter (cl) is set to 10, and the quantization parameter (qp) is set to 8 for a significant compression. As shown in Fig. 3, Draco performs a nonuniform and "blocky" phenomenon under 11.55x and 13.99x compression ratios. The quantization bits are not precise enough to represent the coordinate information. Moreover, our AI-driven solution can achieve a 30.72x compression ratio to the original video while ensuring a visually similar reconstruction result.

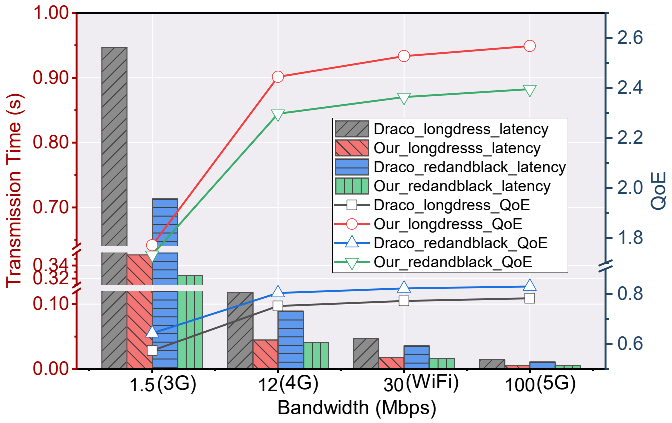

Fig. 4. Quantitative comparisons of the AI-driven method with Draco. The bars denote the transmission latency, and the lines denote the QoE performance.

Fig. 4 presents quantitative experimental results for evaluating the transmission latency and QoE. We conduct experiments under four downlink bandwidth conditions at 1.5 Mbps (3G), 12 Mbps (4G), 30 Mbps (WiFi), and 100 Mbps (5G), respectively. Due to the instability of the 5G networks, achieving a consistent bandwidth is sometimes challenging. We observe that transferring one point cloud frame using the proposed AI-driven framework significantly reduces latency compared with Draco. Meanwhile, our method achieves higher QoE than Draco on both typical datasets. The results illustrate the superiority and robustness of the AI-driven framework. Besides, to verify whether the online adapter can provide adaptive transmission under a dynamic network environment, we have trained more inference models, whose transmitted feature vector matrix sizes are represented as (06, 06) to (20, 20). To test the performance of the online adapter, we have trained the A3C network and used the trained actor-network to infer the transmission model for new point cloud video streaming. The accumulative discount rewards reach convergence at about 700 episodes in the training phase. We also show the model selection results in the testing phase in Fig. 5, illustrating the effectiveness of a DRL-based online adapter. It can be seen that the A3C network of the online adapter achieves convergence before the 1000^{th} episode demonstrating the advantage of the A3C framework. In addition, the red curve represents the bandwidth change over time and the blue bars represent the inference models over time outputted by the online adapter. The adaptive adjustment of the inference models has the same trend as the dynamic changes of network conditions, this then demonstrates the effectiveness of the online adapter.

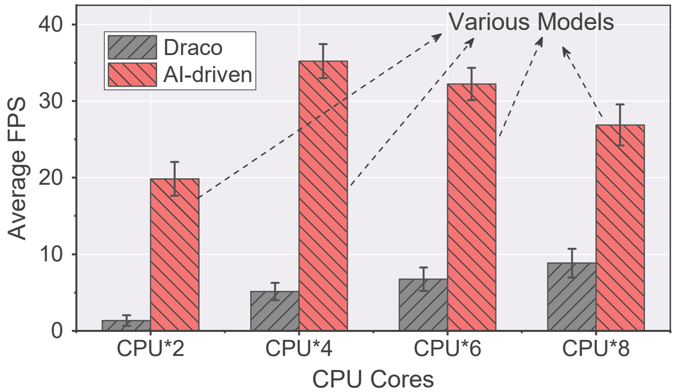

Fig. 5. Performance of DRL-based online Adapter.

V. Use Cases and Future Directions

A. Use Case: Holographic services for future metaverse

Combining 5G with holographic communications and display technologies brings images, body movements, and expressions of people from thousands of miles away in space and time to different social situations. The virtual holographic host or holographic image of the speaker can break through the limitations of space and time, and allow the speaker to communicate with users in real-time in an immersive manner. For example, affected by the COVID-19 epidemic, physical meetings have turned to participation in real-time remote and virtual meetings on smartphones or computers. However, there is a gap between the effectiveness of teleconferencing and live meetings, and many remote offices currently lack a sense of presence and immersion. In this context, the demand for "virtual holographic office" comes into being, and the holographic office is coming closer to reality. For example, when a person speaks, not only the delivery of voice, more body language, micro-expressions can be delivered in 3D elements. So the participants can coexist in the virtual holographic world in three-dimensional forms.

B. Future Directions

AI-driven point cloud streaming provides a foundation for efficient holographic transmission in 5G networks. However, higher resolution and more complex holographic content and interaction introduce challenges for future network services and open up some potential future research directions from data-capturing, adaptive streaming, and interaction involved in future truly holographic video communication.

1) Efficient generation of point cloud videos. Existing point cloud videos leverage multiple depth cameras to simultaneously obtain raw data from various angles and fuse them in real-time to form a 6-DoF video stream. This requires more depth cameras to acquire holographic video with higher resolution and more details. However, the traditional point cloud registration methods for multi-view cameras are only helpful for offline video generation, posing a massive challenge to real-time holographic video capture. Recently, AI-based research on multi-view fusion has provided a new way of thinking for real-time holographic video generation. Therefore, exploring efficient point cloud video generation based on AI is crucial for realizing future real-time holographic video communication and interaction.

2) Holographic point cloud encoding and decoding. Existing compression and encoding schemes and proposed AI-driven solutions require computation resources (e.g., expensive GPUs). A possible solution is to explore light point cloud feature extraction and fused neural networks for efficient compression and encoding. Naturally, exploring novel coding and decoding schemes is promising to improve efficiency, aiming at 6-DoF point cloud video characteristics. More importantly, possible solutions include exploring tiling and angle view prediction schemes, which reduce the transmission data volume and relieve the computing pressure of decoding and rendering at the terminals. For our proposed AI-driven point cloud transmission scheme, designing feature extraction models with a more robust generalization capability and smaller network model parameters and reconstruction models with efficient inference are interesting research directions.

3) Adaptive point cloud streaming. Although we can extend existing adaptive streaming techniques to 6-DoF point cloud videos, there are still barriers during adaptation, such as addressing the extra 3-DoF location information of point cloud videos. For the proposed AI-based point cloud transmission, this adaptive transmission is not only with the network environment and the transmission data volume, but also requires the fusion of multidimensional network resources, such as the computing resources and storage for the inference of AI models. Also, this implies that future adaptive streaming optimizations must be explored for AI-powered point cloud delivery mechanisms. Besides, point cloud video streaming is extremely sensitive to network delay. Therefore, it is crucial to provide transmission error-tolerant and smooth services in unstable network environments. Examples of possible solutions can be studied from encoding, transmission error correction, and reconstruction error recovery.

4) High concurrency processing capability. 5G and Beyond 5G networks support holographic video services closely related to application scenarios and interaction requirements. For instance, a user consuming a pre-registered holographic video service (e.g., visiting an art gallery and museum collection) requires high bandwidth to obtain a satisfying resolution holographic experience. While in multi-user holographic interaction, such as in virtual holographic conference scenarios, users have higher requirements for interaction latency than a better holographic resolution. In the primary stage of holographic video communication, different holographic scenarios and applications currently focus on ultra-bandwidth or ultra-low latency service provision, appropriately sacrificing part of the service performance provision to obtain a holographic video experience. For high concurrency processing, it is helpful to optimize the network transmission and lightweight point cloud codec or use fiber links to enhance the transmission capability of holographic video. Exploring the distributed multiplayer interaction framework is more important to reduce the ultra-high demand for various types of network resources from large numbers of concurrent users. Besides, how to improve the response latency and user experience under the distributed multiplayer interaction framework is the key element to solve the high concurrency processing.

5) Advanced network services for performance boosting. Edge computing, SDN, NFV, 5G slicing, and other network services provided by 5G and future 6G networks can reduce the latency and enhance the computing capability for dense video streaming. However, point cloud video streaming requires networks and computation far greater than the processing capacity of existing service architectures. To this end, exploring more advanced network infrastructures (e.g., network service technologies with more robust communication capabilities and more stable service capabilities) to provide caching and computing resource allocation in line with dense point cloud video streaming is also an important future research direction.

VI. Conclusion

In this article, we reviewed the landscape of hologram video in the form of point clouds, clarified the differences between point cloud video with conventional videos, and revealed that existing technologies are still far from supporting real-time holographic video streaming. Then, we discussed the critical challenges of enabling holographic communication and providing immersive services in transmission technology, computing, mobility, and ubiquity. We further proposed a novel point cloud streaming method that is entirely different from existing delivery mechanisms from an AI perspective, extracting key semantic features for delivery and reconstructing for rendering at terminals. Nevertheless, the generalizability and inference overhead are two main limitations of applying such an AI-driven approach to a broader range of scenarios. Finally, we point out some future directions to help facilitate research in point cloud streaming and immersive services.

Acknowledgment

The authors would like to thank Jacky Cao for his diligent proof-reading of this article. This research was supported in part by the National Key R&D Program of China under Grant 2018YFE0205503, in part by the Funds for International Cooperation and Exchange of NSFC under Grant 61720106007, in part by the 111 Project under Grant B18008, and in part by Academy of Finland under projects 319670 and 326305.

References

[1] J. van der Hooft, M. T. Vega, T. Wauters, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, "From capturing to rendering: Volumetric media delivery with six degrees of freedom," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 49–55, 2020.

[2] C.-L. Fan, W.-C. Lo, Y.-T. Pai, and C.-H. Hsu, "A survey on 360 video streaming: Acquisition, transmission, and display," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–36, 2019.

[3] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation (ICRA)*. Shanghai, China: IEEE, May 9-13 2011.

[4] Google, "Draco 3d data compression. accessed:[2021-11-21]," https://google.github.io/draco/.

[5] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 5105–5114.

[6] R. Li, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng, "Pu-gan: a point cloud upsampling adversarial network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7203–7212.

[7] B. Han, Y. Liu, and F. Qian, "ViVo: Visibility-aware mobile volumetric video streaming," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–13.

[8] L. Wang, C. Li, W. Dai, J. Zou, and H. 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)*. IEEE, 2021, pp. 1930–1934.

[9] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, "Towards 6DoF HTTP adaptive streaming through point cloud compression," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2405–2413.

[10] Z. Liu, Q. Li, X. Chen, C. Wu, S. Ishihara, J. Li, and Y. Ji, "Point cloud video streaming: Challenges and solutions," *IEEE Network*, vol. 35, no. 5, pp. 202–209, 2021.

[11] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. De Turck, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, 2020.

[12] E. C. Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, "6g: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 42–50, 2019.

[13] K. Lee, J. Yi, Y. Lee, S. Choi, and Y. M. 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.

[14] Y. Huang, Y. Zhu, X. Qiao, Z. Tan, and B. Bai, "AITransfer: Progressive AI-powered transmission for real-time point cloud video streaming," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 3989–3997.

[15] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i voxelized full bodies-a voxelized point cloud dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006*, vol. 7, p. 8, 2017.

Yakun Huang is currently a Postdoctoral Researcher at the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. His current research interests include video streaming, mobile computing, and augmented reality.

Yuanwei Zhu is currently working towards a Ph.D. degree at the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. His current research interests include point clouds, video streaming, and deep reinforcement learning.

Xiuquan Qiao is currently a Full Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. His current research interests include the future Internet, services computing, computer vision, distributed deep learning, augmented reality, virtual reality, and 5G networks.

Xiang Su is currently an Associate Professor with the Department of Computer Science, Norwegian University of Science and Technology, Norway, and the University of Oulu, Finland. He has extensive expertise in the Internet of Things, edge computing, mobile augmented reality, knowledge representations, and context modeling and reasoning.

Schahram Dustdar (Fellow, IEEE) is a Full Professor of Computer Science and is heading the Distributed Systems Research Division at the TU Wien. He is an ACM Distinguished Scientist, ACM Distinguished Speaker, IEEE Fellow, and Member of Academia Europaea.

Ping Zhang (Fellow, IEEE) is currently a Full Professor and Director of the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. He is an Academician with the Chinese Academy of Engineering (CAE). He is also a member of the IMT-2020 (5G) Experts Panel and the Experts Panel for China's 6G development.