A QoS-Aware 3D Point Cloud Streaming from Real Space for Interaction in Metaverse

Hiroki Ishimaru[1], Yugo Nakamura[2], Manato Fujimoto[3,4], Hirohiko Suwa[1,4], Keiichi Yasumoto[1,4]
[1] Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
E-mail: {ishimaru.hiroki.ie3, h-suwa, yasumoto}@is.naist.jp
[2] Kyushu University, Fukuoka, Fukuoka 819-0395, Japan
E-mail: y-nakamura@ait.kyushu-u.ac.jp
[3] Osaka Metropolitan University, Osaka, Osaka 558-8585, Japan
E-mail: manato@omu.ac.jp
[4] RIKEN, Chuo-ku Tokyo 103-0027, Japan

***Abstract*—In this paper, we present a point cloud streaming method of real-space objects such as humans and animals for real-time 3D reconstruction in VR space. The system uses a depth camera to scan a human or animal, divides the point cloud into parts of the body, and then controls the quality of the point cloud (i.e., resolution and frame rate) for each part in real-time according to the object's motion and context. This enables point cloud streaming with limited resources (computational and network resources) and maximizes the user's quality of experience (QoE). We have implemented and evaluated a series of systems incorporating the proposed method to enhance the user experience in realistic environments and scenarios while maintaining interactivity in VR-based online communication. The results show that the proposed system is feasible under resource-constrained environments without significantly affecting the user's QoE.**

***Index Terms*—Real-time communication, 3D point cloud, QoE, Streaming, VR**

I. Introduction

In recent years, attention to the metaverse and digital twin has increased, and how to incorporate real space situations into virtual space has been actively discussed. However, because real space is vast and complex, projecting it into the virtual world is a difficult task. Especially the tourism field is one of the most challenging use cases to virtualize because of the importance of experiences and sensations in real space. In the virtualization of tourist attractions, interactivity is essential to recreate the interaction with local people and animals, so it is ideal for sharing space in real-time between virtual and real space. To achieve real-time spatial sharing, real space must be quickly scanned and projected in 3D into virtual space. 3D Point cloud data streaming is one promising approach to achieving this goal.

However, real-time streaming of 3D point cloud data has resource limitations. The size of point cloud data is large, and it is nearly impossible to stream the raw data on existing mobile network infrastructure. For example, the point cloud data generated by the Azure Kinect depth camera is approximately 430 MB/s (720P/15FPS), which requires bandwidth at the Gbps level. Furthermore, there are all kinds of locations and environments in real space, and not all of them necessarily have high-quality infrastructure. In particular, in the field of tourism, which is our premise, the use of cellular networks and mobile devices which have limited resources is essential. Therefore, there is a need for an efficient point cloud streaming method with low bandwidth consumption while maintaining the quality of the user experience [1].

There are many studies and approaches to point cloud streaming, including point cloud data encoding/decoding, tiling, and transmission optimization using view angle prediction. For example, V-PCC which standardizes point cloud compression like a video compression standard [2], a prototype for real-time point cloud capture and streaming on mobile devices [3], and a client-driven dynamic point cloud streaming system based on DASH (Dynamic Adaptive Streaming over HTTP) [4] have been proposed. However, most of these studies focus on point cloud compression and preventing unnecessary data transmission on a client basis. A few studies evaluate the impact of these controls on user QoE [5]–[7].

In this paper, we propose an approach to extend point cloud quality control to context-adaptive driving of objects and to realize point cloud streaming on limited resources while maintaining high QoE. In this paper, we describe the implementation and evaluation results of the proposed system and confirm that the system is capable of streaming point clouds under cellular networks. The impact of the system's quality control on the user's QoE is also investigated and evaluated.

This paper introduces a system that enables remote users to move freely within a space and interact with local people and animals without needing fixed infrastructure, despite remote participation, in situations such as tourism where experience and sensation in real space are important. The system is built around the use of mobile devices so that it can be used at any tourist site or location and enables users to easily use and participate in the system.

The main contributions of this paper are as follows:

- Dynamic control of point cloud quality based on objects in real space is realized.
- A prototype system with real-time quality control under resource limitation is implemented.

- The impact of point cloud quality control on users' QoE is evaluated.

II. Related work

This section describes existing research on XR collaboration and point cloud streaming as related studies.

A. XR Collaboration

Lee *et al*. proposed an approach that divides the real space into OOI (Object Of Interest) and AO (Ambient Object) and combines pre-reconstruction of OOI and real-time reconstruction of AO [8]. They proposed an approach that combines OOI pre-reconstruction and AO real-time reconstruction. For OOI, the posture (position and direction) information in space detected from the feature information of the object is shared with the remote side to reconstruct the polygon model in real-time. For AO, a point cloud masking the areas other than OOI is generated and sent to the remote side, which is used as a background during remote collaboration. WebSocket and WebRTC were used to transmit the attitude information for OOI and the point cloud information for AO, respectively. Each of the information transfer methods is selected according to the server load and data size. In addition, they show an example of using a ball valve as an OOI, which is useful for cases such as obtaining operating instructions from a remote expert. They also propose a system that enables collaboration between real physical space and virtual space with high reproducibility, but the use cases are limited due to the process of OOI recognition and pre-modeling.

B. Point cloud Streaming

High decoding complexity exists in point cloud compression, and Li *et al*. proposed an approach to solve this complexity by transmitting uncompressed tiles of different quality levels in addition to compressed tiles [6]. The system selects an appropriate quality level for each tile that evenly divides the point cloud and dynamically controls the trade-off between computational and communication resources. Although uncompressed tiles consume more bandwidth, they do not require decoding, thus reducing the amount of computation. The authors also claim that the QoE can be optimized by applying an approach that selects only tiles that exist within the user's FoV (Field of View) [9] , thereby reducing unnecessary data transmissions, as well as by selecting the quality of the tiles. They measured the QoE of this system and found that it achieved a higher QoE value than the conventional method. They also propose a method for allocating communication and computational resources for point cloud video streaming and show that it has superior performance compared to existing methods. However, this method does not consider the quality control of each detail of the target object according to the movement and context of the object. Our system extends that approach to dynamically control point cloud quality (that is, resolution and frame rate) according to the object's motion and context.

III. System Requirements

We propose a system that enables remote and local users to share space by scanning and transferring objects in real space in real-time and reconstructing them in virtual space. In this section, we describes the system's prerequisites and associated problems and also introduce our envisioned use cases.

A. Prerequisites

We assume that the system will be used in the field of tourism. This is because tourism is one of the most difficult use cases to virtualize because of the importance of experiences and sensations in real space. We believe that remotely interacting with people and animals in real space can be applied to any situation. It is challenging since there are not many studies targeting remote interactions with people and animals.

Several conditions are necessary to achieve the above goal. There are two major problems: a high-speed network may not be available, and using a fixed system is inappropriate.

First, let us discuss the problem related to high speed network availability. Tourist destinations are widely scattered from rural to urban areas and can be outdoors or indoors. Therefore, we believe the proposed system must be widely available in many locations. Hence, we assume the use of cellular networks. According to data from NTT docomo, the largest Japanese telecommunications carrier, effective cellular communication speeds in Japan from January to March 2022 are 137Mbps to 273Mbps for download and 17Mbps to 39Mbps for upload, which are not as fast as fiber-optic lines [10]. Therefore, we considered systems with these transmission speeds in mind.

Next, we explain the problem that a fixed system is not appropriate. We considered that a fixed installation of sensors and processing equipment at the site would not be appropriate from the standpoint of versatility. The reason for this is that we assume that humans and animals are the target objects when scanning objects in real space, but it is unlikely that they will remain in the same place for a long period of time. To extend the range of using the system, we concluded that the system should be composed of mobile devices in the field. From this condition, we consider that volunteers to handle mobile devices are needed at the site. We assume that there are volunteers at each tourist spot and that the system will be realized with their cooperation.

B. Assumed Use Cases

We aim to make traveling and tourism more accessible by reproducing the local tourist experience in a virtual space and building a system from which travelers and locals can benefit. The system is expected to be used as follows:

1) Volunteers with mobile terminals equipped with LiDAR are at the target tourist sites.
2) The remote user moves freely within the virtual space (provided in advance) corresponding to the sightseeing spot and requests the projection of dynamic objects (people, animals, etc.) in the real space included in his/her field of view onto the virtual space.

3) Upon receiving a request, the volunteer scans the object with the mobile device's LiDAR and sends it to the system.
4) The object is projected into the virtual space.

By placing a volunteer who scans objects between the remote user and the local object, interaction with the local area through the volunteer becomes possible. This enables users to have an experience similar to that in real space when using facilities or purchasing products. We believe these are potential use cases in tourism and can also be applied to education, training, and inspection tours.

C. Technical challenge

When the system is used in a mobile environment, various problems arise. For example, network bandwidth and computing resources may be limited. To maximize QoE with limited resources, factors that affect quality must be adjusted.

In this study, a depth camera is used to scan the target object, but transmitting all the resulting point clouds in real-time would exceed the available bandwidth. For example, the Azure Kinect color point cloud stream uses approximately 430 MB/s bandwidth at 720P/15FPS, which cannot be transmitted in real-time over cellular communication as described in Sect. III-A. Compression of point cloud data is unacceptable in a mobile environment, as it can cause FPS degradation and delays, and non-delivery of other data. In addition, the larger the number of points in the point cloud data, the more it will affect the processing time, which may also lead to a decrease in real-time performance. Therefore, even when sending selected point cloud data according to the OOI (Object Of Interest), it is ideal to be able to reduce the data to the extent that it does not significantly affect the user's QoE.

IV. Proposed System

In this section, we describe the design of the system that satisfies the requirements described in section III-A and its components, such as hardware and functions.

A. Approach

To realize online tourism, it is necessary to convert real space with various tourist objects into real-time data and reconstruct it in virtual space. Still, it is difficult to acquire and use all objects in real space in real-time due to computational and network resources in mobile environments. Lee *et al.* proposed an approach in which a portion of the real space is cut out, and only the necessary objects are reflected in real-time, while static data is used for the other objects [8]. We adopted this approach in our system, dividing the real space into the objects that the remote participants wanted to project and the surrounding objects that would serve as their backgrounds. Fig. 1 shows this approach. The necessary objects and the surrounding objects are called OOI (Object Of Interest) and AO (Ambient Object), respectively.

In addition, we considered that using cameras and communication infrastructure pre-installed at the site would be inappropriate in terms of ubiquity and versatility. Since there

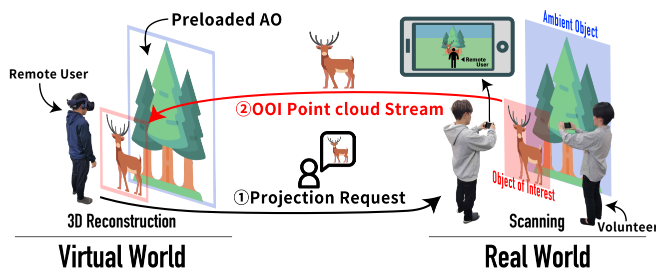

Fig. 1. Our approach in streaming real space objects to virtual space

are countless tourist spots, the system must ultimately be flexible enough to accommodate them. Therefore, we assumed that volunteers would be present in the vicinity of the target real space. That OOI point cloud scanning and transmission would be conducted with their cooperation, using their mobile terminals, such as smartphones and tablets. Under these conditions, even when sending only OOI point clouds, there can be delays that degrade the QoE in terms of communication bandwidth and computational resources. Hence, we propose a method to dynamically control the quality of the OOI point cloud to keep the resources used under the limit. Unlike the previous tiling approaches [6] divide the point cloud object evenly, the proposed method recognizes the object's attribute and divides the point cloud at the body part scale to perform finer-grained quality control of the object to maximize the user experience while reducing bandwidth consumption.

B. System Design

Based on the proposed approach, targeting tourism application, we propose a system that enables remote and local users to share space by scanning and transferring objects in real space and reconstructing them in virtual space in real-time, without requiring fixed infrastructure. As shown in Fig. 1, a volunteer at a sightseeing spot performs scanning and data transmission of local objects (OOI) in the real space in real-time at the request of a remote user. The real space is reconstructed in the remote user's virtual space by using the transmitted point cloud data of the local objects (OOI) together with the background data (AO) prepared in advance, thereby sharing the space of the tourist site is realized.

To reduce bandwidth consumption of the point cloud data (OOI) handled by the system, we propose a dynamic quality control based on body part movements and incorporate it in the system. Based on the techniques used in video compression standards such as H.265 for 2D video and tiling approaches for 3D point clouds, we reduce the total amount of point cloud data by dynamically varying the resolution and frame rate of each body part according to context: high frame rate/low resolution when the object part is moving significantly, and low frame rate/high resolution when the part is nearly stationary. Although the absolute quality of the point cloud data (OOI) is lower than the raw data due to these processes, bandwidth consumption is optimized by controlling the quality so as not to affect the user's QoE as much as possible.

C. Implementation

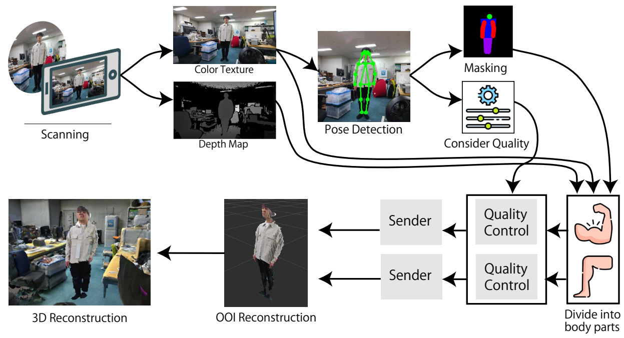

Fig. 2. System Data Flow

The following is a description of the prototype system we developed to evaluate the feasibility of the proposed system. Fig. 2 shows a data flow of the system. The system is built on ROS and uses Unity for 3D reconstruction.

First, we explain how to obtain the AO and OOI to reconstruct the real space in a virtual environment. First, we obtain the AO by using tools such as a 3D scanning application that utilizes LiDAR installed in the iPhone, or a camera that can capture 360-degree images. In recent years, 3D scanning of various locations has been progressing, so scanning of AO objects may not be necessary in the future.

Next, we describe the actual method of acquiring OOI point clouds for transmission and real-time reconstruction. As mentioned above, we assume that mobile devices will be used to acquire the OOI point clouds. The depth and texture of the OOI are continuously acquired using a depth camera and treated as a color point cloud. Our system uses Azure Kinect as well as iPhone for evaluation.

After scanning an OOI object, quality control is dynamically performed according to its context. First, as shown in Fig. 2, pose estimation is performed on the 2D color video input from the camera to segment body parts and track their movement. For the pose estimation, a trt-pose estimation model is used for human pose estimation [11], [12], and a model trained by DeepLabCut is used for animals (deer) pose estimation [13], [14], as shown in Fig. 3.

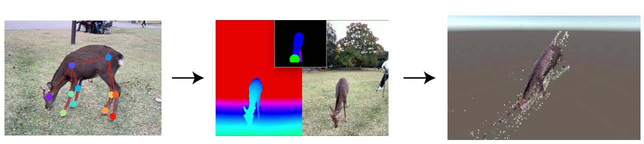

Fig. 3. In case of animals: masks are generated from animal pose estimation results to get a point cloud

Context-aware point cloud quality determination mechanism: The quality of the point cloud is determined by the value of the displacement of the corresponding body part. The system constantly monitors the difference between the position of the body part in the current frame and the previous frame. When this difference exceeds a threshold value, the system judges that the body part is moving and immediately reduces the resolution to give priority to the frame rate. Conversely, when the amount of movement of a body part remains below the threshold value for several frames, the system determines that the object is near stationary and gradually increases the resolution (frame rate decreases).

The resolution of the point cloud is adjusted by resizing the original input (color image, depth map: 720P) using nearest neighbor completion (three steps: 25%, 50%, and 75%), and the frame rate is switched by adjusting the transmission timing. This process is performed for each body part. Table I shows the bandwidth consumption of a sample human video sequence when our system controls the quality. In this sample, the bandwidth consumption (i.e., the maximum value) is 20.2 Mbps when all body parts are set to 75% quality, which is within the possible transmission range with the existing infrastructure.

The resulting point cloud data for each body part is then sent to Unity as a ROS topic and reconstructed in 3D together with the AO prepared in advance. To ensure water tightness in the representation of the point cloud, the size of the points is dynamically changed according to the quality of the point cloud. The reconstructed tourist attraction space is presented to the user through a VR-HMD.

V. Experiment

This section describes experiments conducted to evaluate the impact of quality control on user QoE.

A. Outline of Experiment

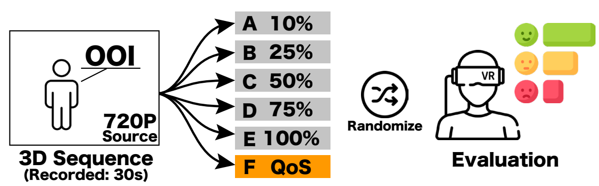

Fig. 4. Quality Evaluation Experiment

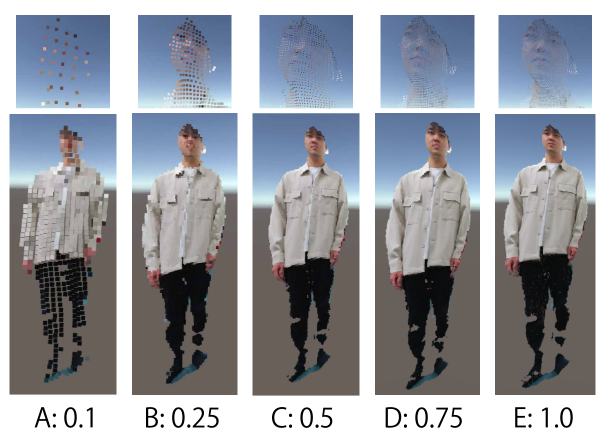

Fig. 5. Change in OOI due to Difference in Compression Ratio

TABLE I
SPECIFICATIONS OF POINT CLOUD QUALITY OUTPUT BY PROPOSED SYSTEM

	Head	Body	Arms	Legs	All
25%	0.28Mbps (13FPS)	2Mbps (13FPS)	2.3Mbps (14FPS)	1.5Mbps (14FPS)	6.1Mbps
50%	0.9Mbps (12FPS)	5.5Mbps (9FPS)	6.1Mbps (9FPS)	4.8Mbps (12FPS)	17.3Mbps
75%	1.8Mbps (9FPS)	6.5Mbps (5FPS)	6.4Mbps (5FPS)	5.6Mbps (6.5FPS)	20.2Mbps

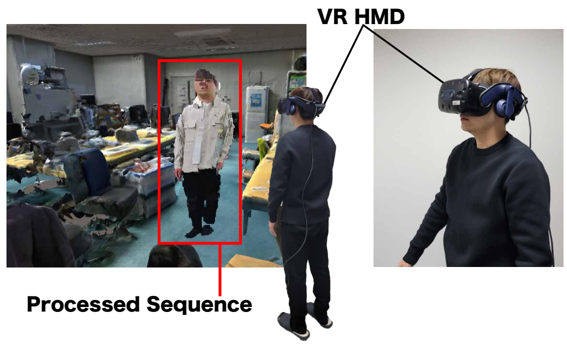

Fig. 6. Subjects wear VR HMD and view point clouds at 6DoF

Fig. 4 shows an overview of the QoE evaluation experiment. In this experiment, we asked users to view several different quality levels of output from a point cloud sequence that we had prepared in advance for about 30 seconds and to rate the quality of the output. The point cloud sequence consists of three states: human standing (stationary), waving, and stepping. Each output stream is based on the input source (720P) with compression ratios of $rate = 0.1, 0.25, 0.5, 0.75, 1$, respectively (Condition A-E). The compression ratio here is the ratio of the resize to the video input to the system. Nearest neighbor interpolation was used to resize both color and depth video (For rate=0.1, the quality is 10% of 720P, i.e., 128x72).Fig. 5 is an example of point cloud display for each compression ratio.

In addition to the above point cloud images, point cloud images in which the quality was dynamically switched according to the context of the object were also subject to evaluation, resulting in a total of six patterns of point cloud image evaluation (Condition F: Proposed system).

Fig. 6 shows a scene from the experiment. In this experiment, a VR headset (HTC VIVE) was used to view contents in a 3D space.

B. Evaluation Methodology

The evaluation method was based on the ACR (Absolute Category Rating) method [15], which is one of the representative video quality evaluation methods. After viewing a 30-second evaluation video, the experimental participants rated the quality of the video within the following 10 seconds on a 5-point quality scale (5.very good, 4.good, 3.normal, 2.poor, 1.very poor).

Since there is an order effect in which the quality of the images is affected by the quality of the previous images, the order in which the evaluated images are presented is randomized and varied for each participant in the experiment. The three evaluation contents are as follows: (i) The quality of the smoothness of the images, and (ii) Quality for the denseness of the point cloud. (ii) Quality in terms of point cloud density, (iii) Overall image quality.

The participants in the experiment were 15 undergraduate and graduate students in their 20s. Each participant watched six evaluation videos at random and rated them on a 5-point scale immediately after watching each video.

TABLE II
SIGNIFICANT DIFFERENCES BY EACH POINT CLOUD SEQUENCE

	Two pairs to compare	P-value	Significance
Smoothness	Compression A,F	0.008	✓
	Compression B,F	0.790	
	Compression C,F	0.002	✓
	Compression D,F	0.000	✓
	Compression E,F	0.000	✓
Resolution	Compression A,F	0.001	✓
	Compression B,F	0.007	✓
	Compression C,F	0.244	
	Compression D,F	0.020	✓
	Compression E,F	0.001	✓
Overall Quality	Compression A,F	0.046	✓
	Compression B,F	0.138	
	Compression C,F	0.129	
	Compression D,F	0.026	✓
	Compression E,F	0.012	✓

C. Result

The quality evaluation results for each evaluation criterion are shown in Fig. 7, Fig. 8, and Fig. 9. MOS was calculated and visualized based on the evaluation results obtained from the experiment participants. First, the Kruskal-Wallis test was used for the results of each evaluation criterion. The test results showed significant differences in smoothness, resolution, and overall quality in all three experiments. (p-value: Smoothness: 8.94×10^{-12}, Resolution: 1.14×10^{-9}, Overall Quality: 1.45×10^{-2}) Thereafter, the Mann-Whitney U-test is used to compare the quality of each criterion result. Table II show the test results. The results of the U-test on the smoothness quality results showed that condition F (proposed method) was not significantly different from condition B (compression ratio of 0.25). This result shows that the proposed method has the same level of smoothness as condition B. The results of the U-test for resolution quality showed that condition F (the proposed method) was not significantly different from condition C (compression ratio of 0.5). This indicates that the proposed method retains the same level of resolution as condition C. The results of the U-test on the overall quality

Fig. 7. Quality for Smoothness

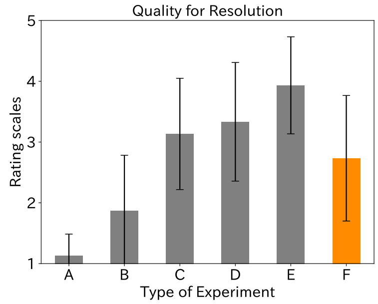

Fig. 8. Quality for Resolution

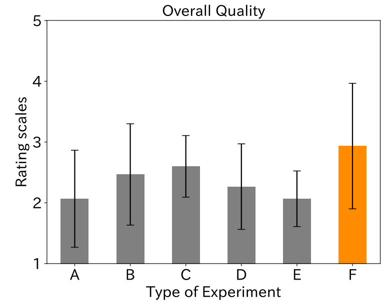

Fig. 9. Overall Quality

results showed that condition F (the proposed method) was not significantly different from condition B (compression ratio of 0.25) and condition C (compression ratio of 0.5). This indicates that the proposed method maintains the same level of overall quality as conditions B and C.

Condition B and C, which had the same overall quality, were discussed in detail. Condition B had the same level of smoothness, but the proposed method was evaluated higher in terms of resolution. Condition C has the same level of resolution, but the proposed method has a higher evaluation in terms of smoothness. From these results, we believe that the proposed method is better in the individual evaluations, although the overall evaluation was at the same level.

VI. Conclusion

In this paper, we proposed a point cloud streaming method for real-time 3D reconstruction of real-space objects such as people and animals in virtual space in order to realize the virtualization of tourism. We proposed a dynamic quality control approach that adjusts the resolution and frame rate of the point cloud according to the object's motion, and implemented and evaluated the proposed method. The results of evaluation experiments confirmed that the proposed method was superior in individual evaluations, although the overall quality of the proposed method was equivalent to that of conditions B and C that degrade FPS/resolution equally for all body parts. Although these results suggest the practicality of a new compression method for point cloud images, it is unlikely that the system will be effective for all types of users, and a more rigorous QoE evaluation that takes into account users' preferences must be conducted. In addition to the display of point clouds, the extent to which various interactions are possible must also be considered.

Acknowledgment

This work is supported in part by the Commissioned Research of National Institute of Information and Communications Technology (NICT) in Japan, Contract No. 222C03 and JSPS KAKENHI JP21H03431.

References

[1] 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.

[2] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, "Emerging mpeg standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2018.

[3] J. Hu, A. Shaikh, A. Bahremand, and R. LiKamWa, "Characterizing real-time dense point cloud capture and streaming on mobile devices," in *Proceedings of the 3rd ACM Workshop on Hot Topics in Video Analytics and Intelligent Edges*, ser. HotEdgeVideo '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 1–6. [Online]. Available: https://doi.org/10.1145/3477083.3480155

[4] M. Hosseini and C. Timmerer, "Dynamic adaptive point cloud streaming," in *Proceedings of the 23rd Packet Video Workshop*, 2018, pp. 25–30.

[5] L. Wang, C. Li, W. Dai, S. Li, J. Zou, and H. Xiong, "Qoe-driven adaptive streaming for point clouds," *IEEE Transactions on Multimedia*, 2022.

[6] J. Li, C. Zhang, Z. Liu, W. Sun, and Q. Li, "Joint communication and computational resource allocation for qoe-driven point cloud video streaming," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[7] J. van der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, "Objective and subjective qoe evaluation for adaptive point cloud streaming," in *2020 twelfth international conference on quality of multimedia experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[8] Y. Lee, B. Yoo, and S.-H. Lee, "Sharing ambient objects using real-time point cloud streaming in web-based xr remote collaboration," in *The 26th International Conference on 3D Web Technology*, 2021, pp. 1–9.

[9] Z. Liu, S. Ishihara, Y. Cui, Y. Ji, and Y. Tanaka, "Jet: Joint source and channel coding for error resilient virtual reality video wireless transmission," *Signal Processing*, vol. 147, pp. 154–162, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168418300185

[10] "Effective speed measurement results." (in Japanese) Accessed: 2022-12-13. [Online]. Available: https://www.docomo.ne.jp/area/effective_speed/

[11] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "Deeplabcut: markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, 2018. [Online]. Available: https://www.nature.com/articles/s41593-018-0209-y

[12] T. Nath*, A. Mathis*, A. C. Chen, A. Patel, M. Bethge, and M. W. Mathis, "Using deeplabcut for 3d markerless pose estimation across species and behaviors," *Nature Protocols*, 2019. [Online]. Available: https://doi.org/10.1038/s41596-019-0176-0

[13] B. Xiao, H. Wu, and Y. Wei, "Simple baselines for human pose estimation and tracking," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 466–481.

[14] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299.

[15] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," 1999.