IEEE Open Journal of the Computer Society

Received 29 May 2020; accepted 23 June 2020. Date of publication 1 July 2020;
date of current version 14 July 2020. The review of this article was arranged by Associate Editor Peng Li.

*Digital Object Identifier 10.1109/OJCS.2020.3006205*

# Fuzzy Logic-Based Adaptive Point Cloud Video Streaming

**ZHI LIU [1] (Senior Member, IEEE), JIE LI [2], XIANFU CHEN [3] (Member, IEEE), CELIMUGE WU [4] (Senior Member, IEEE), SUSUMU ISHIHARA [1] (Member, IEEE), YUSHENG JI [5] (Senior Member, IEEE), AND JIE LI[6] (Senior Member, IEEE)**

[1] Shizuoka University, Shizuoka 422-8529, Japan
[2] Hefei University of Technology, Hefei 230027, China
[3] VTT Technical Research Centre of Finland, 02044 Espoo, Finland
[4] The University of Electro-Communications, Tokyo 182-8585, Japan
[5] National Institute of Informatics, Tokyo 101-8430, Japan
[6] Shanghai Jiao Tong University, Shanghai 200240, China

CORRESPONDING AUTHOR: CELIMUGE WU (e-mail: celimuge@uec.ac.jp)

The work of Z. Liu was supported in part by JSPS KAKENHI under Grants 19H04092, 20H04174, in part by ROIS NII Open Collaborative Research 2020(20FA02). The work of Jie Li was supported in part by grants from the Fundamental Research Funds for the Central Universities under Grant JZ2019HGTB0089.

**ABSTRACT** Point cloud video provides 6 degrees of freedom (6DoF) viewing experiences to allow users to freely select the viewing angles of 3D scenes and is expected to be the next-generation video. This paper studies the point cloud video streaming and proposes a fuzzy logic-based point cloud video streaming scheme to solve the inherent technical issues. In particular, a point cloud video is first partitioned into smaller tiles, along with a low-quality base layer covering the entire video. Each tile is encoded into different quality levels, and both the compressed and uncompressed (i.e., decoded) versions of each tile are prepared for selection. Then, based on the user's viewing angle and predicted future network bandwidth condition, fuzzy logic empowered quality level selection, with properly defined novel *fuzzification, fuzzy rules*, and *defuzzification*, is conducted to maximize the received point cloud video quality under the communication resource, computational resource and quality requirements constraints. Extensive simulations based on real point cloud video sequences and network traces are conducted, and the results reveal the superiority of the proposed scheme over the baseline scheme. To the best of our knowledge, this is the first work studying point cloud video streaming using fuzzy logic.

**INDEX TERMS** Point cloud, hologram video, volumetric video, fuzzy logic, immersive video streaming.

## I. INTRODUCTION

With the development of video capturing and transmission technologies, immersive video such as virtual reality (VR) video [1], which allows free selection of the viewing direction, has become increasingly popular [2]. Point cloud video, as shown in Fig. 1, is composed of points represented in 3D space. Each point is associated with multiple attributes, such as coordinates and color. Point cloud video can provide the participants with an even more immersive viewing experience with 6 degrees of freedom (i.e., 6DoF, including forward/backward (surge), up/down (heave), left/right (sway)), and users can freely select any preferred viewing angle of the 3D scene, which is not applicable in VR video as it only provides 3DoF. Due to the advantages of point cloud video in terms of providing an extraordinary viewing experience and enabling viewing from different angles, it has drawn considerable attention from both academia and industry and is a typical user case of 5G beyond wireless communications [3]–[5]. The global market for Holography for Industrial Applications estimated at US$16.3 Billion in the year 2020 [6]. Thus, the study of point cloud video transmission, which is one fundamental research issue in these applications, is in critical demand.

However, point cloud video transmission over wireless networks [7] is nontrivial. The inherent technical challenges mainly lie in the lack of efficient and adaptive transmission schemes for handling the high bandwidth and low-latency

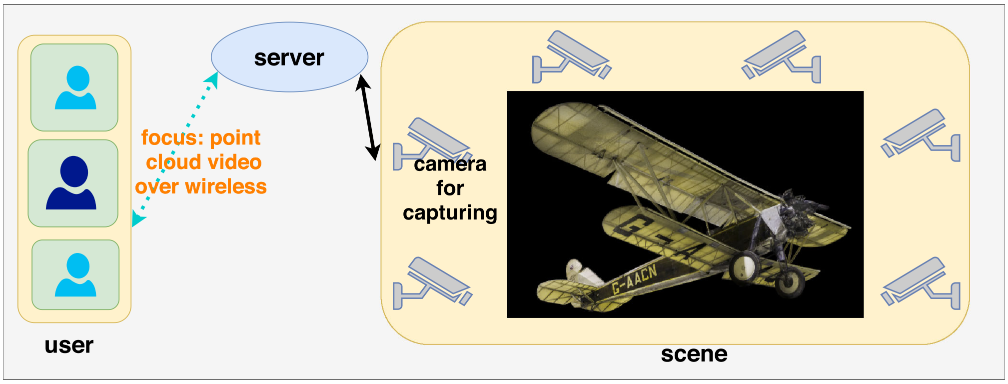



**FIGURE 1.** Illustration of a point cloud video system. In this system, the scene is simultaneously captured by cameras from different angles. The server generates the point cloud video from these captured videos. The server and users are connected via a wireless network, and we focus on the point cloud video streaming over this wireless network.

requirements and accurate viewing angle prediction methods. Apart from these challenges, the encoding/decoding complexity introduced by state-of-the-art point cloud video coding schemes is very high due to the special data structure and lack of corresponding hardware support. These factors make point cloud video streaming very different from traditional video streaming [8], [9] and should be considered in transmission scheme design. This paper focuses on point cloud video streaming, which has recently become a research hotspot [10]–[15] and aims to solve the inherent technical issues to provide users with higher video quality.

The literature in this area mainly focuses on point cloud compression with the goal of optimizing coding efficiency, resulting in a low source video rate but with a high complexity [5], [16]. The point cloud video transmission has not been fully studied, and most of the existing point cloud video streaming schemes are based on VR/360 video streaming methods. In particular, Jounsup *et al.* proposed a volumetric media transmission scheme that spatially partitions the point cloud video into tiles. Different tiles have different quality levels, and the quality levels are selected depending on their relation to the user's view frustum and distance to the user [10]. [11] presented a number of rate adaptation heuristics that use information on the user's position and focus, the available bandwidth, and the client's buffer status to determine the most appropriate quality representation of each object. However, these works did not consider the decoding complexity [12], [13], i.e., the computational requirement of the existing point cloud video codec for decoding exceeds the computational capability of user devices. [12], [13] confirmed this research issue, and Li *et al.* first proposed including uncompressed tiles to reduce the computational burden of the user device and optimize the user's quality of experience (QoE) by selecting the proper quality levels under the communication and computation resource constraints [13]. [14] summarized the current research progress and confirmed these unsolved technical issues. Note that these schemes are model-based and are not adaptive to wireless networks with dynamic channel conditions.

Therefore, this article provides fuzzy logic empowered point cloud video streaming over wireless networks, where network conditions are dynamic, and real-time control is enabled. In particular, the point cloud video is first properly partitioned into tiles with different quality levels, along with a low-quality base layer covering the whole video. The base layer guarantees a smooth viewing angle switch, and tiling allows users to perceive the viewing area at a high-quality level. Both the compressed and uncompressed (i.e., decoded) versions of each tile are prepared for selection. The uncompressed tiles do not require decoding with higher bandwidth consumption, which decreases the user's computational burden and enables a tradeoff between available computational resource and transmission bandwidth. Then, based on the user's viewing angle, future network bandwidth, available computational resource and quality requirements, a fuzzy logic empowered quality level selection method is proposed with properly defined novel *fuzzification, fuzzy rules*, and *defuzzification*. This fuzzy logic empowered scheme considers the unique features of the point cloud video streaming system, which includes the status of the available computations for decoding and channel bandwidth changes, and helps maximize the received point cloud video quality under the communication resource, computational resource and user quality requirements constraints. Extensive simulations based on real point cloud video sequences and network traces are conducted, and the results reveal the superiority of the proposed scheme over the baseline scheme. To the best of our knowledge, this is the first paper studying point cloud video transmission over wireless networks using fuzzy logic. The contributions are summarized as follows:

- A point cloud video is properly portioned into tiles with different quality levels and computational requirements for decoding, along with a low-quality base layer covering the whole video. Each encoded tile is also decoded (i.e., the uncompressed version of each tile is ready for selection as well) at the server for selection.
- With properly defined novel *fuzzification, fuzzy rules*, and *defuzzification*, a fuzzy logic empowered quality

level selection scheme is proposed to help maximize the received point cloud video quality, which considers the future viewing angle, network condition, and available computational resource.

- Extensive simulations based on real point cloud video sequences and network traces are conducted, and the results reveal the superiority of the proposed scheme over the baseline scheme.

The remainder of this paper is organized as follows: Section II introduces the state-of-the-art works related to this article. Section III introduces the system details, including how to encode the point cloud video and conduct the bandwidth prediction. Section IV introduces the fuzzy logic empowered quality level selection. Section V shows the simulation results, and Section VI concludes this paper.

## II. RELATED WORKS

This section introduces the related works from the perspectives of point cloud encoding and processing, point cloud video transmission and fuzzy logic-based resource allocation.

### A. POINT CLOUD VIDEO ENCODING AND PROCESSING

The point cloud is composed of points represented in the 3D space, and each point is associated with multiple attributes, such as coordinates and color. Point cloud video can provide a 6DoF viewing experience and thus is expected to be widely used. The global holography market for industrial application alone will reach 22.5 billion USD by 2024 [6].

However, to the best of our knowledge, point cloud video streaming is still in its infancy. One challenge is the large data volume. Compared with traditional video streaming [17], the transmission bandwidth for a point cloud video application with a frame rate of 30 frames per second (i.e., the standard video frame rate) can be as high as 6 Gbps [18], leading to considerable pressure for transmission and storage. Efficient point cloud video encoding or compression has, therefore, become important for high-quality point cloud video applications.

There are two major classes of point cloud encoding methods according to different distributions of the point cloud data [5], [16]. For point cloud data with a relatively uniform distribution in the 3D space, we can leverage the well-known 2D video technologies by projecting the points into 2D frames. For the point cloud dataset with spare distribution, we can decompose the 3D space into a hierarchical structure of cubes and encode each point as an index of the cube to which it belongs. Note that point cloud encoding requires higher computational complexity than traditional video encoding [12], [13], which distinguishes the point cloud video streaming system from the traditional video streaming system.

### B. POINT CLOUD VIDEO STREAMING

Point cloud video streaming has drawn increasing attention recently [10]–[15]. Most of these works are based on VR/360 video streaming schemes [1], [19], i.e., a point cloud video is divided into smaller tiles, and only tiles inside the user's field of view (FoV) (denotes the part of the video the user is watching) are transmitted, and optimization can be conducted following this thread to optimize the defined objective function [10], [11]. However, these works do not consider the decoding complexity [12], [13], which makes them hard to deploy in real systems. Li *et al.* [13] considered the high computational complexity of point cloud video decoding, introduced uncompressed tiles to reduce the computational requirement for decoding, and studied the inherent transmission optimization problem. Note that these schemes are model-based and are not adaptive to wireless networks with dynamic channel conditions.

### C. FUZZY LOGIC EMPOWERED RESOURCE ALLOCATION

Fuzzy logic [20], [21] is an alternative for dynamic resource allocation with a very low complexity compared with reinforcement learning and many other methods [22] while providing satisfactory performances [23]–[25]. In particular, [23] studied the MPEG DSAH video streaming system and used fuzzy logic to control the bitrate with the objective of distributing video segments of the best quality, delivering undisrupted video playback, and avoiding frequent changes in video resolution. [24] extended [23] with more appropriate future bandwidth prediction and bitrate selection. Zhang *et al.* [25] studied fuzzy theory-assisted 3D video streaming over a mobile P2P network.

To the best of our knowledge, fuzzy logic has not yet been studied, particularly for immersive video (e.g., VR/360 video, point cloud video) streaming. The difficulties of fuzzy logic learning-based schemes mainly include how to formulate the resource allocation problem into a fuzzy logic problem, how to properly define the rules and select the parameters to obtain satisfactory performance. Therefore, this paper proposes a fuzzy logic empowered point cloud video streaming considering the features of point cloud video streaming.

## III. POINT CLOUD VIDEO STREAMING SYSTEM

This section introduces the point cloud video streaming system, including point cloud video tiling and network bandwidth prediction.

### A. SYSTEM OVERVIEW

As illustrated in Fig. 1, we consider a point cloud video system with dynamic network conditions and real-time control requirements. In this system, video cameras (RGBD cameras or RGB cameras [26]) capture the target scene from different angles. The captured videos are delivered to the server, and the server generates the point cloud video from these captured videos. The users can then subscribe to the point cloud video from the server via a last-mile wireless network.

We focus on point cloud video transmission over wireless networks. There are $N$ users in this system. User $i$ has a selected viewing angle at time instance $t$. The corresponding

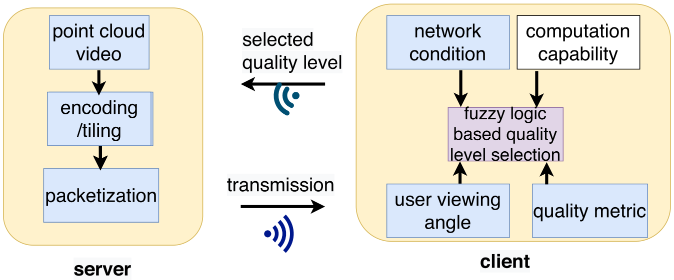



**FIGURE 2.** **Flowchart of the fuzzy logic-based point cloud video streaming system.**

viewing area is named FoV[1] and denoted as $F_{i,t}$. The user can switch to another FoV $F_{i,t+1}$ at the next time instance $t+1$. The received video quality should be no smaller than $Q_i$, which corresponds to the user quality requirement constraint. The bandwidth experienced by user $i$ is limited and dynamic. We assume user $i$'s bandwidth is $B_{i,t}$ at time instance $t$. $B_{i,t}$ can be known (with errors) by bandwidth prediction as explained in Section III-C. At time $t$, user $i$ has $C_{i,t}$ available computational capacity to decode the compressed tiles. Due to the high decoding complexity, the available computational capacity at the user device is insufficient for decoding the compressed frames at 30 frames per second, and we count $C_{i,t}$ in terms of how many compressed frames[2] the user can decode per second.

To comply with MPEG DASH [27], each user selects the quality levels using fuzzy logic, considering the buffer status, predicted channel bandwidth, viewing angle and available computation capacity. Then, the tiles with the selected quality levels are transmitted to the user via the wireless networks. We assume that the quality levels of the tiles inside each FoV at each time instance are the same to avoid degrading the viewing experience. Fig. 2 shows the flowchart of the proposed fuzzy logic-based point cloud video streaming.

Next, we explain each component in detail.

### *B. POINT CLOUD VIDEO TILING*

A point cloud video streaming system should provide a smooth viewing angle switch and high-quality received video. For the first point, we propose to encode the whole point cloud video into a base layer at a low quality. The base layer covers all the viewing angles and thus smooths the viewing angle switch. The size of the base layer for time $t$ is $S_t$. Note that uncompressed version of the base layer is not provided to reduce the bandwidth consumption. Regarding the second research issue, the point cloud video is partitioned into tiles to efficiently utilize the precious and limited channel bandwidth. This allows transmitting the tiles inside the user FoV at a high-quality level with limited bandwidth consumption and thus enables a high-quality received video.

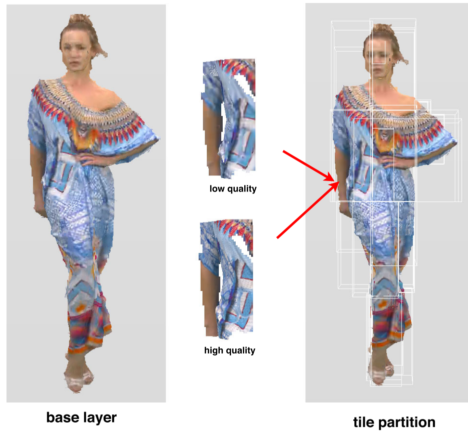



**FIGURE 3.** **Illustration of the base layer and the enhancement layers at two selected quality levels. The empty parts are omitted in each tile.**

Specifically, we first calculate the length, width, and height of the smallest enclosing cuboid that surrounds the point cloud video frame and determine which side of the cuboid is "height" according to the situation of the object. For example, when the point cloud video is about a character, the direction in which the character stands is the "height" of the cuboid. Then, we perform $N \times M$ partitioning on the plane perpendicular to the height and dividing them into $H$ layers in the "height" direction, and finally, $N \times M \times H$ tiles are obtained. Fig. 3 shows a base layer and typical tile partitioning.

Each partitioned tile at time $t$ is compressed into $L$ quality levels. Given that the previous tile may disappear in the next frame, tiling makes the interframe prediction more difficult, and we can adjust the interframe prediction accordingly to avoid this issue. Compressed tile $(n, m, h, t)$, $1 \le n \le N$, $1 \le m \le M$, $1 \le h \le H$ with quality level $l$, $1 \le l \le L$ has source encoding size $s_{n,m,h,l,t}$, point-to-point PSNR $q_{n,m,h,l,t}$, and computational requirement $d_{n,m,h,l,t}$ for decoding (in terms of how many compressed frames to decode). The high-quality level provides a high-quality video with a larger source encoding size. We also decode each compressed tile and thus have uncompressed tiles at different quality levels. Uncompressed tile $(n, m, h, t)$ with quality level $l$ has source encoding size $s'_{n,m,h,l,t}$, point-to-point PSNR $q_{n,m,h,l,t}$ (which is the same as the compressed tile at the same quality level), and zero computational requirement for decoding. The uncompressed tile has a larger bandwidth requirement in comparison with the corresponding compressed tile, i.e., $s'_{n,m,h,l,t} > s_{n,m,h,l,t}$. By using the uncompressed tiles for selection, we can reduce the computational load of the user device by sacrificing more bandwidth consumption. Note that both the base layer and each tile use the group of pictures (GoP) as the encoding unit, which is the same as in traditional video systems.

### *C. VIEWING ANGLE AND BANDWIDTH PREDICTION*

Viewing angle prediction is an important component in point cloud video streaming. Accurate viewing angle prediction

[1]Different from the VR/360 video system, the number of tiles covered by one particular FoV can be very different due to the different selected viewing distances from the scene.

[2]Note that different numbers of tiles are transmitted for different frames. We count the computational capability in the frame level to show its impact on the point cloud video streaming.

could avoid wasting precious network bandwidth resource by only delivering the video parts watched by the user. This paper focuses on fuzzy logic-based point cloud video transmission and associated video processes, and we assume that the viewing angles are known based on the prediction and feedback from user devices.[3]

For the future network bandwidth prediction, we consider the gated recurrent unit (GRU), long short-term memory (LSTM) and stacked autoencoders (SAEs) and heavily reuse the code provided by [29]. Because the page is limited and we focus on fuzzy logic-based quality level selection, we omit the explicit explanations, and refer the reader to [29] for more details. The prediction results are shown in Section V-B.

## IV. FUZZY LOGIC EMPOWERED POINT CLOUD VIDEO TRANSMISSION

This section introduces the adopted fuzzy logic for user $i$'s tile quality level selection. We first explain the constraints and objective and then introduce the fuzzy logic empowered point cloud video transmission. Finally, we discuss some feasibility issues to implement this system.

### A. CONSTRAINTS AND OBJECTIVE

We first define two indicator notations: $I^{i,c}_{n,m,h,l,t} = 1, 1 \leq n \leq N, 1 \leq m \leq M, 1 \leq h \leq H, 1 \leq l \leq L$ indicates that the compressed tile $(n, m, h, t)$ at quality level $l$ is transmitted at time $t$ for user $i$, and 0 otherwise. Similarly, $I^{i,u}_{n,m,h,l,t} = 1$ indicates that the uncompressed tile $(n, m, h, t)$ at quality level $l$ is transmitted at time $t$ for user $i$, and 0 otherwise.

For each tile, at most one quality level is selected and delivered, then we have the following constraint:

$$\sum_{l=1}^{L}(I^{i,c}_{n,m,h,l,t} + I^{i,u}_{n,m,h,l,t}) \leq 1, \forall n, m, h, t \tag{1}$$

The communication resource constraint for a particular time period $t$ is shown as follows:

$$\begin{aligned} S_t + \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{h=1}^{H}\sum_{l=1}^{L}(I^{i,c}_{n,m,h,l,t} \times s_{n,m,h,l,t} \\ + I^{i,u}_{n,m,h,l,t} \times s'_{n,m,h,l,t}) \leq B_{i,t}, \end{aligned} \tag{2}$$

which guarantees that the bandwidth consumed by user $i$ does not exceed his/her available bandwidth.

The computational resource constraint requires that consumed computations should not exceed the available computational capacity, which is shown as follows:

$$\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{h=1}^{H}\sum_{l=1}^{L} I^{i,c}_{n,m,h,l,t} \times d_{n,m,h,l,t} \leq C_{i,t} \tag{3}$$

We also guarantee the minimal quality level requirement, which is shown as follows:

$$\begin{aligned} &\sum_{l=1}^{L}(I^{i,c}_{n,m,h,l,t} + I^{i,u}_{n,m,h,l,t}) \times q_{n,m,h,l,t} \\ &\geq \sum_{l=1}^{L}(I^{i,c}_{n,m,h,l,t} + I^{i,u}_{n,m,h,l,t}) \times Q_i, \forall n, m, h, t \end{aligned} \tag{4}$$

The objective is to maximize the received video quality by the user during the time period $1..., T$; thus, the optimization problem can be formulated as follows:

*Problem 1 (Quality Maximization):*

$$\max_{\bar{q}} \quad \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{h=1}^{H}\sum_{l=1}^{L}\sum_{t=1}^{T}(I_{n,m,h,l,t} + I'_{n,m,h,l,t}) \times q_{n,m,h,l,t}$$
$$\text{s.t.} \quad (1), (2), (3), (4),$$

where $\bar{q}$ denotes the set of the quality level selection parameters $q_{n,m,h,l,t}$. Note that we priority the base layer in the transmission to guarantee the smooth viewing angle switch and omit the base layer in the objective function for simplicity. To handle the large variable space and fulfill the requirement of real-time control for video streaming, we introduce a fuzzy logic empowered solution for this problem in the next section.

### B. FUZZY LOGIC EMPOWERED QUALITY LEVEL SELECTION

The compressed tiles have different features compared with the uncompressed tiles. On the one hand, the compressed tiles are very small in size in comparison with the uncompressed tiles. On the other hand, the compressed tiles require very high computational capacity for their decoding before playback, which exceeds the computational capacity of user device. Thus, we prioritize transmitting uncompressed tiles over compressed tiles.

Fig. 4 describes the logic of this the fuzzy logic empowered quality level selection. In particular, the user first calculates the highest quality level of the uncompressed tiles allowed according to the predicted bandwidth at the next time instance. If this quality level is below the predefined threshold, and the available computational capability is beyond the predefined threshold (which is different from the threshold for quality levels), the client requests a compressed tile instead. When the compressed tiles are chosen, the highest quality level is selected. By doing so, we provide users a better received video quality with small bandwidth overhead. The threshold values are determined based on the computational capability of the user device and bandwidth conditions. If the quality level is larger than the predefined threshold or the available computational capability is smaller than the predefined threshold, we use fuzzy logic to adjust the quality level, which is the core of this paper and introduced in Section IV-C. Then, the transmission begins, and the buffer and available computational resource are updated.

[3]Note that viewing angle prediction in VR/360 video streaming is mature, and similar methods can be used for more challenging point cloud video streaming, where 6DoF is allowed [28].

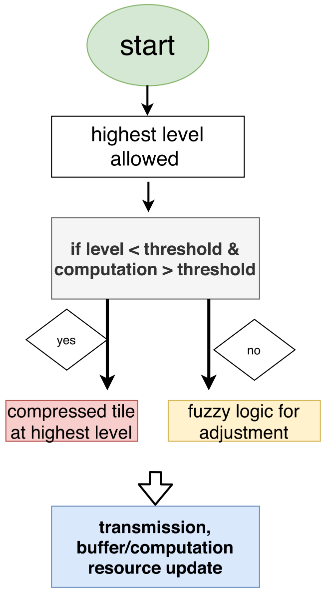



**FIGURE 4.** **Flowchart of the fuzzy logic empowered quality level selection.**

### *C. FUZZY LOGIC-BASED QUALITY LEVEL ADJUSTMENT*

Fuzzy logic is a form of many-valued logic in which the true values of variables may be any real number between 0 and 1, both inclusive. To properly select the quality levels, we consider three input variables, i.e., buffer status, bandwidth changes and available computational resource. The buffer status records the number of point cloud video frames cached in the client buffer. Bandwidth changes show the changes between the bandwidth of the next time instances and the current bandwidth. The available computational resource determines how many frames the client can decode when the decision is made. The client aims to maximize the received video quality by selecting the proper quality level without violating the constraints (1), (2), (3), (4).

We next explain the details from the perspectives of *fuzzification, fuzzy rules* and *defuzzification*.

#### 1) FUZZIFICATION

We first explain the membership functions of the three variables, i.e., buffer status, bandwidth changes and available computational resource. Then, we explain the membership function of the output.

The buffer status records the number of frames cached in the client buffer. In this system, only a very small buffer is allowed to guarantee the real-time requirement. Thus, we consider a buffer up to 15 frames in this paper, which equals a 0.5-second delay and is tolerable for point cloud video users. Note that we can further reduce the number of frames in the buffer, which does not affect the effectiveness of the proposed scheme. Three linguistic variables *[empty, fair, full]* are adopted for the buffer status. A trapezoidal membership function, as shown in Fig. 5(a), is adopted to describe the buffer status. *Empty, fair, full* indicate that the number of frames in the buffer is small, normal, and large, respectively.

Bandwidth changes are calculated as the ratio of the predicted bandwidth to the current bandwidth. In the simulation, we compare the current bandwidth with the predicted bandwidth at the next 5 time instances. The bandwidth changes are classified into three categories: *[decreasing, similar, increasing]*, which represent that the bandwidth requirement will decrease, remain approximately the same, and increase, respectively. We use the trapezoidal membership function as the membership function for bandwidth changes, and this function is shown in Fig. 5(b).

Point cloud video decoding requires much more time than 2D video decoding. The available computational resource records the status of the computational capability. When the client selects the video quality, the available computations are counted in terms of how many point cloud video frames the client can decode. Three linguistic variables *[limited, normal, substantial]* are adopted for the available computational resource. A trapezoidal membership function, as shown in Fig. 5(c), is adopted to describe the status of the available computation resource. *[Limited, normal, substantial]* indicate that the number of frames the client can decode is small, normal and large, respectively. With a higher available decoding capability, the client can select to transmit more compressed video tiles to save bandwidth consumption; otherwise, the uncompressed video is transmitted to guarantee the decoding can be completed in time. This allows a tradeoff between the bandwidth consumption and computational power consumption.

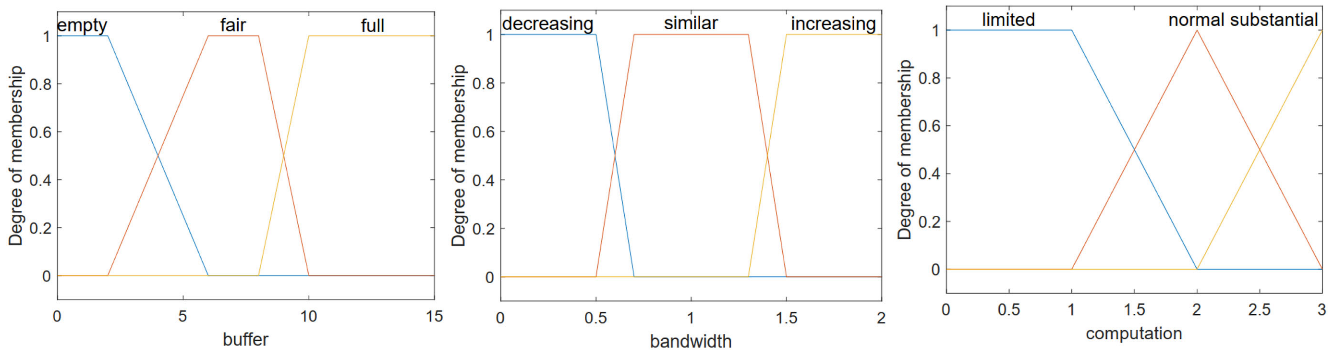
The output is defined in a similar manner to represent the quality level selection factor. As shown in Fig. 6, the linguistic variables of the output are described as *large decrease/mode change (BD), decrease (D), no change (keep), increase (I), and large increase (BI)*. Decrease (D), no change (keep), increase (I), and large increase (BI) show how the quality level should be adjusted, where the quality level decreases by one, remains the same, increases by one, and increase by two, respectively. Large decrease/mode change (BD) means that the client will receive compressed tiles instead due to resource indigence. Similarly, the highest quality level is selected for the compressed tiles.

#### 2) FUZZY RULES

The fuzzy if-then rules for our controller are shown in Table 1, which map the three inputs to an output value. To design the fuzzy if-then rules, parameter tuning is required according to the conditions of the scenarios. However, this algorithm is not complicated and does not require heavy training as in deep reinforcement learning; thus, it is suitable for point cloud video transmission in wireless networks.

#### 3) DEFUZZIFICATION

The last step of the fuzzy logic process is defuzzification, where there are many mature defuzzification methods. In this article, we adopt the *centroid* as the defuzzification method. Finally, we obtain a recommendation value for the system.




(a) (b) (c)

**FIGURE 5. Membership functions of the input, where (a), (b), and (c) are membership functions of buffer, bandwidth changes and available computational resource, respectively. Here we assume a buffer with 15 frames, and the maximum available computational capability can decode three frames within half a second.**

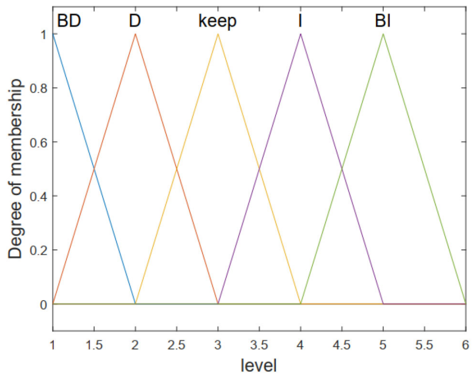



**FIGURE 6. Membership functions of the output.**

**TABLE I. Fuzzy Rules**

| Rules | | | | |
|---|---|---|---|---|
| | buffer | bandwidth | computation | action |
| Rule1 | empty | decreasing | limited | D |
| Rule2 | empty | decreasing | normal | BD |
| Rule3 | empty | decreasing | substantial | BD |
| Rule4 | empty | similar | limited | D |
| Rule5 | empty | similar | normal | keep |
| Rule6 | empty | similar | substantial | BD |
| Rule7 | empty | increasing | limited | D |
| Rule8 | empty | increasing | normal | keep |
| Rule9 | fair | increasing | substantial | BD |
| Rule10 | fair | decreasing | limited | D |
| Rule11 | fair | decreasing | normal | D |
| Rule12 | fair | decreasing | substantial | BD |
| Rule13 | fair | similar | limited | keep |
| Rule14 | fair | similar | normal | I |
| Rule15 | fair | similar | substantial | I |
| Rule16 | fair | increasing | limited | I |
| Rule17 | fair | increasing | normal | BI |
| Rule18 | fair | increasing | substantial | BI |
| Rule19 | full | decreasing | limited | keep |
| Rule20 | full | decreasing | normal | I |
| Rule21 | full | decreasing | substantial | I |
| Rule22 | full | similar | limited | BI |
| Rule23 | full | similar | normal | BI |
| Rule24 | full | similar | substantial | BI |
| Rule25 | full | increasing | limited | BI |
| Rule26 | full | increasing | normal | BI |
| Rule27 | full | increasing | substantial | BI |

### *D. FEASIBILITY*

Video streaming has a stringent delay requirement, and thus, the complexity should not be high. The fuzzy logic-based quality level selection method has low complexity, which can be seen from the aforementioned details. The threshold values, membership functions, and rules in this scheme must be predefined according to the network and user conditions, which is also not difficult. Therefore, the fuzzy logic-based quality level selection method is feasible with good transmission performance, as introduced in the next section.

In addition, the user quality requirement can be satisfied by eliminating improper low-quality levels from the quality level choices, which further lowers the complexity as well.

## V. SIMULATION RESULTS

In this section, we introduce the simulation setup and simulation results to verify the performance of the proposed fuzzy logic-based point cloud video streaming.

### *A. SIMULATION SETUP*

We used two MPEG recommended point cloud video sequences *longdress* and *loot* as the video source. We allowed the view switch at each frame boundary, i.e., the user could switch viewing angles 30 times per second. This is also the highest frequency allowed by a video system with a frame rate of 30 Hz. The encoder and decoder was *VPCC-TMC2-v7*, which is recommended by MPEG. Both videos were divided into $3 \times 4 \times 4$ tiles. The lowest video quality was provided by the compressed base layer. Each tile was encoded into another 5 video quality levels for both compressed versions and uncompressed versions. The FoV was selected randomly, and our scheme works with other types of viewing angle switch models as well. We assumed that each user could decode 3

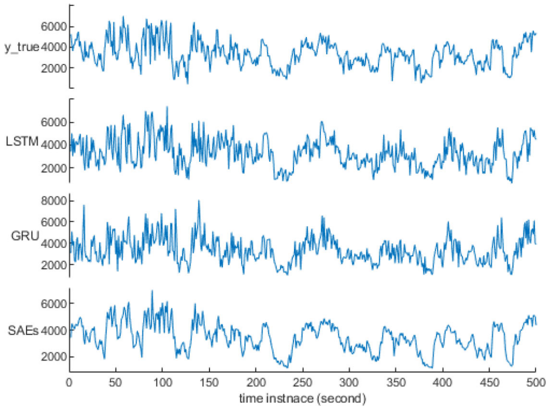
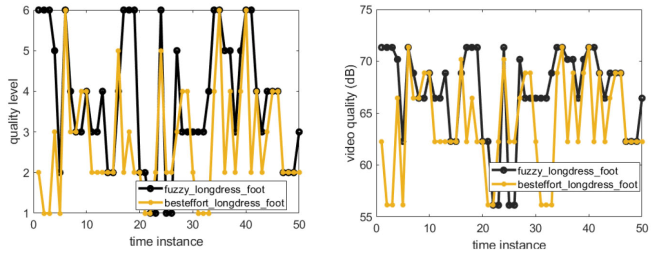
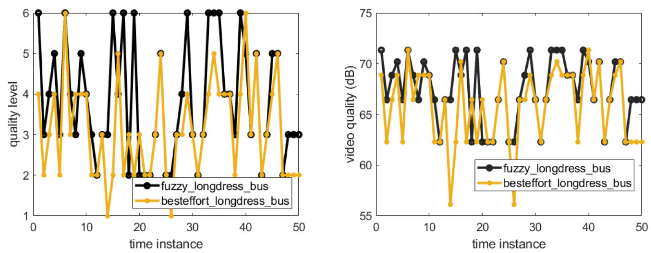



**FIGURE 7.** **Comparison of the bandwidth and the predicted bandwidth at different time instances.**




(a) Quality level. (b) Received video quality (dB).

**FIGURE 8.** **Simulation performance with trace *foot* and video sequence *longdress*.**




(a) Quality level. (b) Received video quality (dB).

**FIGURE 9.** **Simulation performance with trace *bus* and video sequence *longdress*.**

compressed tiles per 0.5 seconds[4] and the base layer could satisfy the user quality requirement.

The network trace we used was the 4G LTE traces [30] in two different scenarios: *foot* and *bus*, with different channel dynamics. We scaled the traces to provide a larger bandwidth and excluded the bandwidth consumed by the base layer, which was the same for all schemes.

### *B. CHANNEL BANDWIDTH PREDICTION*

The dataset [30] provides network traces in different scenarios, including *bus, car, foot, train, bicycle* and *tram*. Here, we first show the prediction result for the *bus* trace. The goal was to predict log02 and log03 of the *bus* trace, and the remaining traces were used as the training data.

The results are shown in Fig. 7. From the simulation results, we can observe that the deep learning algorithms provided very accurate bandwidth prediction. Among the three adopted deep learning schemes, SAEs had the highest accuracy. This was also true for other network traces such as *foot* and *car*. Thus, the results achieved by SAEs were used for the quality level selection.

We also found that more training data (including the traces in different scenarios) resulted in higher accuracy, and the accuracy of predicting the trend (i.e., bandwidth increase or decrease) was higher than that of predicting the exact value.

### *C. TRANSMISSION PERFORMANCE*

To show the performance of the proposed scheme, we compared the fuzzy logic-based scheme with a baseline scheme *best effort*. The *best-effort* scheme selected the highest possible video quality levels of the uncompressed tiles according to the predicted bandwidth at the next time instance.

[4]We found that each compressed tile required 0.5 seconds for decoding using an office laptop; thus, we assumed that each user could decode 3 compressed tiles in 0.5 seconds considering that each machine had multiple CPUs. Our scheme is also effective with other available computational resource, which is reflected in the simulation result as well.

We first show the simulation results using the trace *foot* and video sequence *longdress* in Fig. 8. From Fig. 8, we can observe that the proposed scheme achieves the *best effort* by a large margin. This is because the fuzzy logic-based scheme can adaptively select the proper quality level according to the buffer status, bandwidth changes and available computational capacity and thus can efficiently use these available bandwidth and computing resource. The fuzzy logic-based scheme also reduces the stall (where the advanced tiles are not received), where the stall degrades the viewing experience. This is because the fuzzy logic-based scheme can properly adjust the quality level in accordance with future bandwidth changes and available computing resource.

We then show the simulation results using the trace *bus* and video sequence *longdress* in Fig. 9. In Fig. 9, we can observe similar performance as in Fig. 8. The fuzzy logic-based scheme outperforms the baseline scheme because it can properly select the quality level according to the buffer status, bandwidth changes and available computation capacity.

In Table 2, we show more simulation results in terms of average received video quality of 1,100 time instances with different network traces, different video sequences and different available computational capabilities $C$. From these simulation results, we make the following observations. First, the fuzzy logic-based scheme greatly reduces the number of stalls. Second, the fuzzy logic-based scheme outperforms the baseline scheme in all scenarios, which is similar to the aforementioned results in Figs. 8 and 9. Third, when the available computational capability becomes higher, better received video quality can be obtained, and the stall ratio decreases as

**TABLE II. Simulation Performance**

| | PNSR (dB) | Ratio of stall |
|---|---|---|
| Fuzzy with longdress&foot | 66.25 | 0.026 |
| Best_effort with longdress&foot | 63.57 | 0.231 |
| Fuzzy with longdress&bus | 66.83 | 0 |
| Best_effort with longdress&bus | 65.19 | 0.101 |
| Fuzzy with longdress&foot&$C$=10 | 66.60 | 0.007 |
| Fuzzy with loot&foot | 69.00 | 0.015 |
| Best_effort with loot&foot | 66.58 | 0.171 |
| Fuzzy with loot&bus | 69.45 | 0 |
| Best_effort with loot&bus | 68.39 | 0.060 |

well. This is because a higher computational capability allows the transmission of more compressed tiles, which reflects the tradeoff between the bandwidth and computational capability.

## VI. CONCLUSION

This paper studied fuzzy logic-based adaptive point cloud video streaming. The point cloud video is first partitioned into smaller segments with different quality levels, along with a lower quality base layer covering the entire video. In addition to the compressed tiles, uncompressed tiles are also provided to enable the tradeoff between network bandwidth consumption and computation consumption. Then, based on the viewing angle, predicted network bandwidth and buffer status, a fuzzy logic empowered quality level selection scheme is conducted to maximize the received video quality. Extensive simulations based on real point cloud video sequences and network traces were conducted, and the results reveal the superiority of the proposed scheme over the baseline scheme.

## ACKNOWLEDGMENT

Jie Li is the second author.

## REFERENCES


[1] M. Zink, R. Sitaraman, and K. Nahrstedt, "Scalable 360° video stream delivery: Challenges, solutions, and opportunities," *Proc. IEEE*, vol. 107, no. 4, pp. 639–650, Apr. 2019.

[2] T. S. Mujber, T. Szecsi, and M. S. Hashmi, "Virtual reality applications in manufacturing process simulation," *J. Mater. Process. Technol.*, vol. 155, pp. 1834–1838, 2004.

[3] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 828–842, Apr. 2017.

[4] E. Alexiou, E. Upenik, and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process.*, 2017, pp. 1–6.

[5] S. Schwarz *et al.*, "Emerging MPEG standards for point cloud compression," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 133–148, Mar. 2019.

[6] Global Industry Analysts, "Global holography for industrial applications industry," 2019. [Online]. Available: https://www.reportlinker.com/p080443/World-Holography-for-Industrial-Applications-Market.html?utm_source=GNW, Accessed on: Jul. 7, 2020.

[7] H. Liao *et al.*, "Learning-based context-aware resource allocation for edge computing-empowered industrial IoT," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4260–4277, May 2020.

[8] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[9] H. Hu, Y. Jin, Y. Wen, and C. Westphal, "Orchestrating caching, transcoding and request routing for adaptive video streaming over icn," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 1, pp. 1–23, 2019.

[10] J. Park, P. A. Chou, and J.-N. Hwang, "Rate-utility optimized streaming of volumetric media for augmented reality," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 149–162, Mar. 2019.

[11] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, "Towards 6dof http adaptive streaming through point cloud compression," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 2405–2413.

[12] F. Qian, B. Han, J. Pair, and V. Gopalakrishnan, "Toward practical volumetric video streaming on commodity smartphones," in *Proc. 20th Int. Workshop Mobile Comput. Syst. Appl.*, 2019, pp. 135–140.

[13] 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 *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–6.

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

[15] Y. Huang, J. Peng, C.-C. J. Kuo, and M. Gopi, "A generic scheme for progressive point cloud coding," *IEEE Trans. Visualization Comput. Graph.*, vol. 14, no. 2, pp. 440–453, Mar./Apr. 2008.

[16] C. Cao, M. Preda, and T. Zaharia, "3D point cloud compression: A survey," in *Proc. 24th Int. Conf. 3D Web Technol.*, 2019, pp. 1–9.

[17] A. Ali *et al.*, "Adaptive bitrate video transmission over cognitive radio networks using cross layer routing approach," *IEEE Trans. Cogn. Commun. Netw.*, early access, Apr. 28, 2020, doi: 10.1109/TCCN.2020.2990673.

[18] E. dEon, B. Harrison, T. Myers, and P. Chou, "8i voxelized full bodies-a voxelized point cloud dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006, Geneva*, 2017.

[19] C. Guo, Y. Cui, and Z. Liu, "Optimal multicast of tiled 360 VR video," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 145–148, Dec. 2018.

[20] L. A. Zadeh, "Fuzzy logic," *Computer*, vol. 21, no. 4, pp. 83–93, 1988.

[21] C.-C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller. I," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 2, pp. 404–418, Mar./Apr. 1990.

[22] X. Chen *et al.*, "Age of information aware radio resource management in vehicular networks: A proactive deep reinforcement learning perspective," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2268–2281, Apr. 2020.

[23] D. J. Vergados, A. Michalas, A. Sgora, D. D. Vergados, and P. Chatzimisios, "FDASH: A fuzzy-based MPEG/DASH adaptation algorithm," *IEEE Syst. J.*, vol. 10, no. 2, pp. 859–868, Jun. 2016.

[24] H. J. Kim, Y. S. Son, and J. T. Kim, "A modification of the fuzzy logic based dash adaptation scheme for performance improvement," *Wireless Commun. Mobile Comput.*, vol. 2018, 2018, Art. no. 8078710.

[25] C. Zhang, H.-H. Cho, C.-Y. Chen, T. K. Shih, and H.-C. Chao, "Fuzzy-based 3-D stream traffic lightweighting over mobile P2P network," *IEEE Syst. J.*, vol. 14, no. 2, pp. 1840–1851, Jun. 2020.

[26] W. Lemkens, P. Kaur, K. Buys, P. Slaets, T. Tuytelaars, and J. De Schutter, "Multi RGB-D camera setup for generating large 3d point clouds," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 1092–1099.

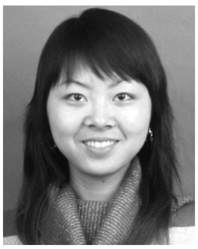
[27] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the internet," *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[28] X. Hou, J. Zhang, M. Budagavi, and S. Dey, "Head and body motion prediction to enable mobile VR experiences with low latency," in *Proc. IEEE Global Commun. Conf.*, 2019, pp. 1–7.

[29] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[30] J. van der Hooft *et al.*, "HTTP/2-based adaptive streaming of HEVC video over 4G/LTE networks," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2177–2180, Nov. 2016.


**ZHI LIU** (Senior Member, IEEE) received the B.E., from the University of Science and Technology of China, China and Ph.D. degree in informatics in National Institute of Informatics. He is currently an Assistant Professor at Shizuoka University. His research interest includes video network transmission, vehicular networks and mobile edge computing. He is now an Editorial Board Member of *Springer Wireless Networks* and has been a Guest Editor of *ACM/Springer Mobile Networks & Applications*, *Springer Wireless Networks*, and *IEICE Transactions on Information and Systems*.



**JIE LI** (Member, IEEE) received the B.E. degree in electronic engineering in 2006 and Ph.D. degree in 2011 from University of Science and Technology, China. She is currently an Associate Professor at the Hefei University of Technology, Anhui, China. From Mar. 2012 to Sep. 2012, she was a Research Assistant at the National Institute of Informatics, Japan. Her research interest includes cognitive radio network and wireless video streaming.

**XIANFU CHEN** (Member, IEEE) received the Ph.D. degree in signal and information processing from the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China, in March 2012. He is currently a Senior Scientist with the VTT Technical Research Centre of Finland, Oulu, Finland. His research interests include various aspects of wireless communications and networking, with the emphasis on software-defined networking, green communications, centralized and decentralized resource allocation, dynamic spectrum access, and the application of artificial intelligence to wireless communications.

**CELIMUGE WU** (Senior Member, IEEE) received the M.E. degree from the Beijing Institute of Technology, China, in 2006, and the Ph.D. degree from The University of Electro-Communications, Japan, in 2010. He is currently an Associate Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications. His current research interests include vehicular ad hoc networks, sensor networks, intelligent transport systems, IoT, and mobile cloud computing. He is/has been a TPC Co-Chair of Wireless Days 2019, ICT-DM 2018, a track Co-Chair of many international conferences including ICCCN 2019 and IEEE PIMRC 2016. He is/has been serving as an Associate Editor of *IEEE Access*, *IEICE Transactions on Communications*, *International Journal of Distributed Sensor Networks*, *MDPI Sensors*, and a Guest Editor of the *IEEE Transactions on Emerging Topics in Computational Intelligence*, *IEEE Computational Intelligence Magazine*, *ACM/Springer MONET*, etc.

**SUSUMU ISHIHARA** (Member, IEEE) received the B.E., M.E., and Dr. Eng. degrees in electronic engineering from Nagoya University, Nagoya, Japan, in 1994, 1996, and 1999, respectively. From 1998 to 1999, he was a JSPS Special Researcher. He joined Shizuoka University in 1999 and is now a Professor of the College of Engineering, Academic Institute, Shizuoka University, Japan. He was a Visiting Researcher at the University of California, Irvine, in 2008 and at the University of California, Los Angeles, in 2014–2015. His current research interests include design and implementation of communication protocols and services, especially for vehicular ad hoc networks, and wireless sensor networks. He is a Member of the ACM, IPSJ, and IEICE, and a Senior Member of IPSJ.

**YUSHENG JI** (Senior Member, IEEE) received the B.E., M.E., and D.E. degrees in electrical engineering from the University of Tokyo. She joined the National Center for Science Information Systems (NACSIS), Japan, in 1990. Currently, she is a Professor at the National Institute of Informatics (NII), and SOKENDAI (the Graduate University for Advanced Studies). Her research interests include network architecture, resource management, and quality of service provisioning in wired and wireless communication networks. She is/has been an Editor of the *IEEE Transactions on Vehicular Technology*, a Symposium Co-Chair of IEEE GLOBECOM 2012, 2014, and a Track Co-Chair of IEEE VTC2016- Fall, VTC2017-Fall, etc.

**JIE LI** (Senior Member, IEEE) received the B.Eng. degree in computer science from Zhejiang University, Hangzhou, China, the M.Eng. degree in electronic engineering and communication systems from the China Academy of Posts and Telecommunications, Beijing, China, and the Dr.Eng. degree from the University of Electro-Communications, Tokyo, Japan. He is currently with the Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai, China, where he is also a Chair Professor. He was a Full Professor with the Department of Computer Science, University of Tsukuba, Japan. He was a Visiting Professor with Yale University, USA, Inria Sophia Antipolis, and Inria Grenoble-Rhone-Alpes, France. His current research interests include big data and AI, cloud and Edge computing, machine learning and networking, network security, OS, and information system architectures. He has also served on the program committees for several international conferences. He is a Co-Chair of the IEEE Technical Community on Big Data, and the Founding Chair of the IEEE ComSoc Technical Committee on Big Data. He serves as an Associate Editor for many IEEE journals and Transactions.