IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 21, 2021, accepted March 8, 2021, date of publication March 18, 2021, date of current version April 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3067018

360Cast+: Viewport Adaptive Soft Delivery for 360-Degree Videos

LU YUJUN, TAKUYA FUJIHASHI, (Member, IEEE), SHUNSUKE SARUWATARI, (Member, IEEE), AND TAKASHI WATANABE, (Member, IEEE)
Graduate School of Information Science and Technology, Osaka University, Osaka 5650871, Japan

Corresponding authors: Lu Yujun (lu.yujun@ist.osaka-u.ac.jp) and Takuya Fujihashi (fujihashi.takuya@ist.osaka-u.ac.jp)

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP19H01101 and Grant JP20K19783.

ABSTRACT The existing viewport-adaptive 360-degree video streaming schemes encode tiled 360-degree videos with digital-based compression. However, these schemes cause a cliff effect wherein the headset video quality drops when the channel signal-to-noise ratio (SNR) falls below a certain threshold. To realize high-quality wireless 360-degree video streaming, we propose a novel viewport-adaptive soft delivery scheme for 360-degree videos, called 360Cast+. 360Cast+ skips the non-linear operations in digital-based streaming and adopts power allocation and analog modulation to achieve graceful video quality improvement in unstable wireless links. In particular, 360Cast+ integrates the human vision system (HVS) and projection distortion as a perceptual weight in power allocation operations. A near-optimal low-complexity subcarrier matching algorithm was also adopted to extend 360Cast+ to fading channel environments. To reduce the effect of the prediction error, 360Cast+ uses dynamic linear regression (DLR) to predict the future orientation and future prediction error to extract the extension area around a viewport. The evaluation results demonstrated that the proposed 360Cast+ provides sophisticated video quality irrespective of the prediction error and channel conditions in orthogonal frequency-division multiplexing (OFDM)-based systems.

INDEX TERMS 360-degree video delivery, soft delivery, viewport adaptive.

I. INTRODUCTION

Virtual reality (VR) is a multimedia technology with a high potential for growth. All types of VR content build a synthetic virtual environment to mimic the real world for participants to interact with. VR can be applied to various applications such as automotive video streaming, virtual live concerts, and six-degrees-of-freedom content streaming for remote operation. However, the poor user experience provided by traditional computer-supported VR headsets or all-in-one headsets (e.g., Oculus Go) limits the imagination and potential of virtual worlds.

The restricted mobility of wired VR headsets and the lack of real-time high-quality content in all-in-one VR headsets are the main problems encountered by VR applications. If these limitations are eliminated, wireless VR headsets can provide immersive experiences for users at any time or place, thereby producing a plethora of novel VR application opportunities.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

To display VR content on wireless VR headsets, the delivery of high-quality 360-degree videos over wireless links is a challenging issue. A 360-degree video, also called an immersive or spherical video, is a new video format for VR content. Each 360-degree video consists of spherical images captured by an omnidirectional camera or camera array. The captured 360-degree video is then projected onto a 2D plane via projection methods, including equirectangular and cubemap projections. Although the full resolution of 360-degree video delivery can immerse users in realistic virtual worlds, the required transmission rate is significantly high; therefore, the transmission rate may not meet the bandwidth requirement in wireless links. To meet the bandwidth requirement in wireless links, viewport-adaptive 360-degree video delivery has been proposed in recent years. A viewport represents a part of each 360-degree video frame that is displayed on the user headset. While playing VR content on a headset, each headset user can freely switch their viewing orientation by moving their head, and only 20% of the pixels of the entire 360-degree video frame need to be rendered and displayed on the headset [1]. Based on this information, each headset

user requests a new viewport from the sender; subsequently, the sender sends back the requested viewport with enhanced quality in viewport-adaptive 360-degree video delivery for traffic reduction.

In conventional viewport-adaptive 360-degree video delivery schemes over wireless links, digital video compression and digital wireless transmission are performed in sequence [2]–[7] for the 360-degree video frames. For example, the video compression part uses the H.265/High-Efficiency Video Coding [8] standard and generates a compressed bitstream by implementing nonlinear quantization and entropy coding operations defined in MPEG-I (MPEG Immersive media) standard. The wireless transmission part uses a channel coding and a digital modulation scheme to reliably transmit the encoded bitstream.

However, the conventional schemes have the following problems due to unreliable wireless channels. First, encoded bitstreams are highly vulnerable to bit errors that occur in wireless channels with low quality. Even when the conventional schemes can assign a large number of bits to the user viewing viewport for quality enhancement, the video quality on the headset drops significantly when the channel signal-to-noise ratio (SNR) falls below a certain threshold. This phenomenon is referred to as the cliff effect. Second, the video quality of the user viewing viewport is constant even in high-quality wireless channels. This is because quantization is a lossy process whose distortion cannot be recovered at the receiver. Although some studies [9], [10] have been proposed to mitigate the cliff effect in digital-based video delivery by introducing layered source coding and layered channel coding, the cliff effect is converted into the so-called staircase effect [11]. In the staircase effect, the video quality discontinuously improves as the wireless channel quality improves. Third, the experienced quality of viewport-adaptive delivery schemes is highly dependent on the viewport prediction accuracy. Even when a high-quality viewport is successfully delivered to a headset user, the experienced quality will drop when the distance between the displayed viewport and predicted viewport is large.

In this study, we propose a novel delivery scheme for 360-degree videos, called 360Cast+, to overcome the aforementioned issues. The proposed 360Cast+ is inspired by soft video delivery schemes proposed in [12], [13]. 360Cast+ integrates viewport prediction, analog modulation, and optimal power allocation to provide better video quality on the wireless headsets of users. In the proposed scheme, we first predict the future viewport and prediction error of each user based on their past head movements via dynamic linear regression (DLR). 360Cast+ then transforms the pixel values within the predicted viewport, whose size is adaptive to the prediction error, into frequency components, and then directly maps the frequency components to transmission symbols, i.e., analog modulation, after transmission power allocation. In particular, the power allocation process assigns transmission power to the frequency components within the predicted viewport considering the distortions in the human vision system (HVS) and sphere-to-plane projections. The evaluation results demonstrated that the proposed 360Cast+ yields better video quality in comparison to the existing digital-based and soft video delivery schemes even in band-limited environments.

In contrast to 360Cast, proposed in [14], 360Cast+ overcomes two major issues of wireless 360-degree video delivery, namely the viewport prediction error and frequency-dependent wireless channel variations. In [14], a future viewport was predicted based on the recent head movements of a user. Although it achieved better video quality in the predicted viewport, the headset users do not always conform to the predicted head movements. The prediction error causes significant degradation in the video quality on the user headset, especially in high-quality wireless channels. In 360Cast+, we predict a potential prediction error from the recent head movements and the predicted head movement; then, based on the potential prediction error, the size of the viewport is enlarged. We confirmed that an enlarged viewport can decrease the effect of the viewport prediction error, implying low-quality degradation, with limited increment in video traffic.

In addition, 360Cast was designed for broadcast transmission with additive white Gaussian noise; thus, it does not consider the frequency-dependent channel variations in modern wireless systems. In the widely used orthogonal frequency-division multiplexing (OFDM) systems, which decompose a wideband channel into multiple subcarriers, the channel SNRs across subcarriers are usually different, often by approximately 20 dB [15]. To accomplish high-quality 360-degree video delivery in modern wireless systems, we utilized diversity for channel selectivity. 360Cast+ designs a subcarrier matching algorithm and joint source-channel power allocation technique to minimize the total end-to-end distortion by utilizing the frequency-dependent channel variations.

The contributions of our study are as follows:

- The power allocation process in 360Cast+ integrates the distortions of human perception in HVS and 2D-plane projection for better visual quality on the user headset.
- The joint source-channel power allocation process with a subcarrier matching algorithm utilizes the frequency-dependent differences across the available subcarriers for high-quality 360-degree video delivery in modern OFDM systems.
- We used the head movements of ten headset users for the evaluations. We demonstrated that the prediction of viewports in 360Cast+ reduces the prediction error effect and realizes better and efficient video quality on each user headset in comparison to the existing digital-based and soft video delivery schemes.

II. RELATED WORK

A. MPEG-I STANDARD ACTIVITIES

MPEG-I groups have discussed various 360-degree video streaming standards for VR systems. They defined the common test conditions (CTC) and evaluation procedures for the

video coding of 360-degree videos. In the CTC, the digital-based video coding standard of H.265/HEVC, i.e., HEVC Test Model (HM) [16], is used for the video coding. The HM uses the test 360-degree video sequences such as ClassroomVideo and TechnicolorPainter for video encoding to evaluate the coding efficiency. Since the resolution of the 360-degree video is even high, the user's viewport is one of the key factors to reduce the required bandwidth for VR streaming. MPEG-I groups have discussed the possibility of motion-constrained tile sets (MCTS) [17] for traffic reduction considering the user's viewport. The MCTS limits the temporal prediction range within the same region, enabling the independent extraction and decoding of each encoded tile at the receiver. In this case, the server selectively determines each encoded tile will/will not be delivered to the headset user over networks based on the user's viewport, i.e., viewport-based streaming, using the dynamic adaptive streaming over HTTP (DASH) [18].

In addition, the weighted to spherically uniform peak SNR (WS-PSNR) [19] and immersive video PSNR (IV-PSNR) [20] are defined as the quality metrics to evaluate the perceptual quality in VR systems. WS-PSNR considers the pixels errors based on the projection distortion from the sphere to the 2D plane while IV-PSNR considers the pixel shift and global color shift into the conventional PSNR to deal with rounding errors and color characteristics in VR systems.

In this paper, the proposed 360Cast+ skips nonlinear operations of the quantization and entropy coding from the digital-based video coding to prevent cliff effect and constant quality irrespective of wireless channel fluctuation. In addition, we used both WS-PSNR and IV-PSNR as the quality metrics to discuss the perceptual quality of the proposed 360Cast+.

B. VIEWPORT-ADAPTIVE 360-DEGREE VIDEO STREAMING

The streaming schemes of viewport-adaptive 360-degree videos have been widely studied for traffic reduction. The main approach of viewport-adaptive 360-degree video streaming includes tile-based schemes [2]–[7], [21]–[34]. The tile-based schemes divide each 360-degree video frame into sub-frames called tiles. To deliver the optimal quality of 360-degree video over wired/wireless links, each tile is encoded into different quality levels. References [2], [3] encoded the tiles into two layers, i.e., the base layer and the enhancement layer. The base layer provided the entire 360-degree video frame, while the enhancement layer improved the quality of the tiles corresponding to the predicted field of view (FoV). A study by [22] determined an optimal bit allocation technique for each tile based on the user viewport, and [24] used quantization parameters for bit allocation across the tiles. References [23] proposed an optimal hypertext transfer protocol (HTTP) streaming for the encoded tiles, and OpTile [4] adaptively determined the size of the tiles to reduce the storage requirement of HTTP streaming. References [21] evaluated the existing tile-based methods to discuss the benefits of tile-based methods over no tile-based methods. Some studies compensated for the viewport prediction error by extending the predicted viewport [5]–[7]. References [6] extended k-tiles around the predicted viewport, while [5] extended the viewport area based on the user head movements. Other studies considered the characteristics of 360-degree videos for bit allocation. References [28] used 2D projection distortion as the weighted distortion of the bit allocation algorithm, and [29] developed a spherical bit-rate equalization technique for rate-distortion optimization.

The existing tile-based methods optimize the quality of each tile to meet the requirements of the wired/wireless bandwidth. However, in a realistic wireless channel, digital-based methods induce the cliff effect owing to the fluctuations in the wireless channel quality. The proposed 360Cast+ adopts analog modulation for viewport-adaptive 360-degree video delivery to make efficient improvements in the viewport quality by improving the instantaneous wireless channel quality. In addition, 360Cast+ considers the characteristics of 360-degree videos, i.e., joint distortions of 2D projection and human perception, for power allocation to ensure better viewport quality on the user headsets.

C. VIEWPORT PREDICTION

The quality of the viewport-based streaming highly depends on the accuracy of the viewport prediction. For accurate viewport prediction, there are non-learning-based [1], [2], [35]–[37] and learning-based [33], [38]–[40] prediction methods. The most existing studies on the non-learning-based viewport prediction utilize a regression curve obtained from the past head movement, such as linear regression (LR) [1], [35], [37] and weighted LR (WLR) [2], [36]. In the learning-based prediction methods, [38] used saliency map prediction based on orientation data of multiple users and [39] further adopted the recurrent neural network (RNN)-based long short-term memory (LSTM) model to predict future viewport movement from both the saliency map and past head orientation. Some studies [33], [40] discuss the effect of the LR-based and deep-learning-based methods on the viewport prediction. They found that the difference of the prediction accuracy between both methods is slight.

Our study designs the non-learning-based viewport prediction. In contrast to the existing LR-based viewport prediction, we utilize DLR for future head orientation and potential prediction error to reduce the effect of the prediction error. We found the viewport considering the potential prediction error can decrease the quality degradation owing to the prediction error with a slight increment in the bandwidth requirement.

D. SOFT IMAGE/VIDEO DELIVERY

Soft image/video delivery schemes [12], [13], [41]–[55] have been recently proposed to ensure that the received video quality is proportional to the instantaneous wireless channel quality. For example, SoftCast [12] skips quantization and entropy coding and uses analog modulation, which maps

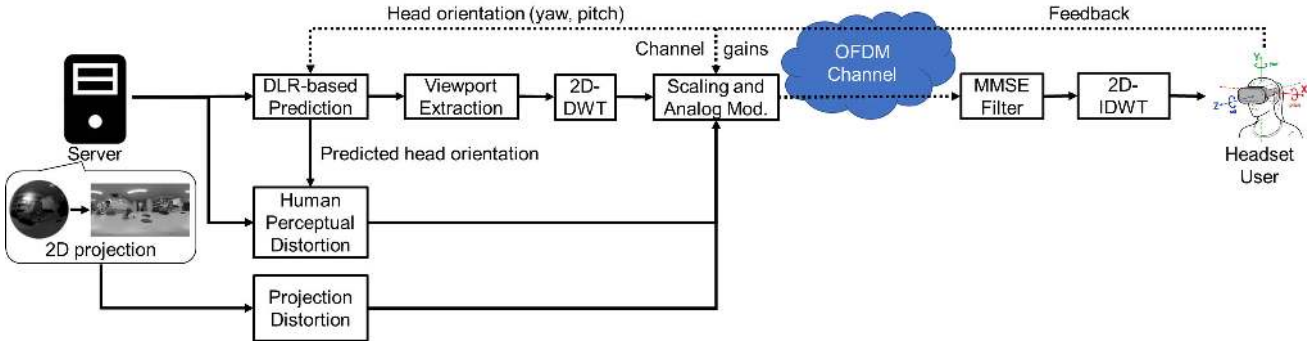


FIGURE 1. Overview of the proposed 360Cast+.

the discrete cosine transform (DCT) coefficients directly to the transmission signals. Some researchers, motivated by the concept that both the source and channel components have non-uniform energy distributions, utilized the distributions for joint source-channel coding. ParCast [13] extended SoftCast to a multiple-input multiple-output OFDM link and assigned high-energy source components to high-gain subchannels that could be utilized based on the non-uniform energy distributions. To accommodate multiple users with diverse channel conditions, ECast [50] proposed a joint source–subcarrier matching and power-allocation scheme to minimize the mean square error (MSE). Additionally, spatial scalability-enabled robust video broadcast (SSRVB) was used to address the multi-user scenario of both heterogeneous device resolutions and channel conditions by iterative joint subcarrier matching and power-allocation methods [42]. Other studies extended SoftCast to immersive video contents. FoveaCast [53] utilized the perceptual distortions in the HVS for soft image delivery to achieve a higher visual quality for users. FreeCast [45] adopted 5D-DCT for multi-view plus depth and exploited a fitting function based on a Gaussian Markov random field model for metadata overhead reduction. References [51] first considered soft delivery for 360-degree videos by using a combination of 1D-DCT and spherical wavelet transform. OmniCast [41] proposed two algorithms to find the block partition with the minimum 2D projection distortion for different sphere-to-plane projections.

Although the existing studies [41], [51] have designed soft delivery schemes for 360-degree videos, they require a large amount of video traffic because they send the full resolution of the 360-degree videos. To reduce the traffic in the soft delivery of 360-degree videos, 360Cast+ only sends the predicted and extended viewports considering the potential viewport prediction error. In addition, we designed our 360Cast+ for modern OFDM-based systems, i.e., frequency-dependent channel variations, by utilizing joint subcarrier matching and power allocation to minimize the end-to-end distortions in wireless 360-degree video delivery.

III. 360Cast+

A. OVERVIEW

This study proposes a novel soft and viewport-adaptive delivery scheme for 360-degree videos. Fig. 1 shows an

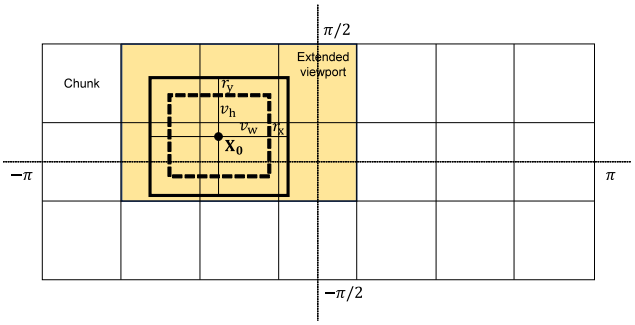


FIGURE 2. Viewport prediction based on the predicted head orientation and potential viewport prediction error.

end-to-end system of the proposed 360Cast+. We consider the past orientations of the user head, i.e., the pitch, yaw, and roll, that are transmitted from the user headset. Based on the past orientations and DLR, 360Cast+ predicts the future orientation and potential prediction error. Here, we consider the foveation point of the headset user as the predicted future orientation and extract the viewport from the full resolution of the 360-degree video based on the foveation point. The extracted viewport is then transformed into discrete wavelet transform (DWT) coefficients using 2D-DWT through the Daubechies 9-tap/7-tap filter. The DWT coefficients are divided into multiple chunks and scaled by chunk-wise power allocation before transmission. In this case, 360Cast+ finds an optimal match between the chunks and subcarriers for optimal power allocation. This optimal match is based on the variance of each chunk and the channel gain of each subcarrier. 360Cast+ then assigns a transmission power to each chunk to optimize the perceptual quality of the headset user by considering the joint distortions in the sphere for 2D plane projection and human perception in the HVS. The power-assigned DWT coefficients in each chunk are sequentially mapped to the I and Q components symbols and transmitted over each subcarrier based on the matching result.

At the decoder, the minimum MSE (MMSE) filter can provide an optimal linear estimate for the received DWT coefficients. AT the decoder, the minimum MSE (MMSE) filter can

B. VIEWPORT PREDICTION

360Cast+ first predicts the future head orientation of the headset user based on the past head orientation received from the headset to estimate the user foveation point. Considering the 2D-projected 360-degree video frames, the user viewport can be determined from two attributes of the orientation, i.e., the yaw and pitch. Let $\boldsymbol{P} = \{P_0, \ldots, P_t\}$ be a set of pitch/yaw attributes from an initial time, P_0, to the present time, P_t, and the predicted pitch/yaw attribute, $\widehat{P}_{t+t_p}$, at the future time, $t + t_p$, can be obtained using DLR as follows:

$$\widehat{P}_{t+t_p} = f_w(\boldsymbol{P}), \tag{1}$$

where w is the dynamic window size for LR and $f_w(*)$ is a LR function that uses the past pitch/yaw attribute from P_{t-w} to P_t. In contrast to the standard LR model, DLR adaptively sets the window size, w, based on the inflection point, $\hat{j}$, obtained from the past head orientations. To find the inflection point from past orientations, 360Cast+ uses the three previous orientations for an arbitrary index, j; for example, P_{t-j-1}, P_{t-j-2}, P_{t-j-3}, and decides the inflection point, $\hat{j}$, satisfying $(P_{t-\hat{j}-1} - P_{t-\hat{j}-2}) \cdot (P_{t-\hat{j}-3} - P_{t-\hat{j}-2}) > 0$.

360Cast+ also predicts the potential prediction error of the head orientation to reduce the effect of the prediction error on the displayed video quality on the user headset. The past prediction error, E_t^P, at time t can be derived between the predicted and actual head orientations as follows:

$$E_t^P = \left|\widehat{P}_t - P_t\right|.$$

Let $\boldsymbol{E}^P = \{E_0^P, \ldots, E_t^P\}$ denotes the set of the past prediction error. The potential prediction error corresponding to the predicted pitch/yaw attribute $\widehat{P}_{t+t_p}$ can be obtained from the DLR-based prediction as follows:

$$\widehat{E}_{t+t_p}^P = f_w\left(\boldsymbol{E}^P\right),$$

The predicted head orientation and potential prediction error were used to obtain the size of the transmission viewport. Fig. 2 displays an overview of viewport prediction based on the predicted head orientation and potential viewport prediction error. Let the predicted head orientation and potential prediction errors transformed to the 2D plane be denoted by $\boldsymbol{X_0} = (x_0, y_0)$ and $\boldsymbol{R} = (r_x, r_y)$, respectively. In addition, we assume that the half-width and half-height of the user headset viewport corresponding to the FoV are v_w and v_h, respectively. We first consider a tentative viewport at the center coordinate of (x_0, y_0) with half-width, $v_w + r_x$, and half-height, $v_h + r_y$. We then consider the chunks included in a tentative viewport as the extended viewport. 360Cast+ only sends extended viewports to the user headsets for traffic reduction.

C. ENCODING

1) PROBLEM FORMULATION

The extracted viewport is then transformed into frequency components via 2D-DWT and divided into N chunks with a resolution of $c_h \times c_w$ pixels. Let $x_i[j]$ denote the i-th analog-modulated symbol, which is the i-th chunk, c_i, scaled by a factor of g_i for noise reduction, as follows:

$$x_i = g_i \cdot c_i. \tag{2}$$

During a transmission time slot, N chunks are assigned to N subcarriers in the OFDM systems. The receiver obtains the received symbol over wireless OFDM links, which is modeled as follows:

$$y_{i,j} = h_j x_i + n_i = h_j g_i c_i + n_i.$$

where $y_{i,j}$ is the received symbol of the i-th chunk over the j-th subcarrier, h_j is the channel fading coefficient of the j-th subcarrier, and n_i is the effective noise with a variance of σ^2. The transmitter performs optimal power control by selecting g_i to obtain the best 360-degree video quality for the headset user. Accordingly, we define a weighted metric called weighted MSE (WMSE) as follows:

$$\mathrm{WMSE}_{i,j} = \frac{S(v, \boldsymbol{X_i}) \cdot D_s(\boldsymbol{\Theta_i}) \cdot \mathrm{MSE}_{i,j}}{\sum_{i=1}^{N} S(v, \boldsymbol{X_i}) \cdot D_s(\boldsymbol{\Theta_i})}, \tag{3}$$

WMSE denotes the weighted mean square error between the original and reconstructed 360-degree video frames considering the sphere-to-2D mapping distortion, $D_s(\boldsymbol{\Theta_i})$, and perceptual distortion, $S(v, \boldsymbol{X_i})$, in the HVS. Both $D_s(\boldsymbol{\Theta_i})$ and $S(v, \boldsymbol{X_i})$ can be calculated based on the current location of the given point in the spherical and pixel domains, i.e., $\boldsymbol{\Theta_i} = (\theta_i, \phi_i)$, $\boldsymbol{X_i} = (x_i, y_i)$, respectively. The best g_i should be obtained by minimizing the WMSE under the power constraint with the total power budget, P. When the transmission symbols of the i-th chunk are assigned to the j-th subcarrier, the square error, $\mathrm{MSE}_{i,j}$, in Eq. 3 can be obtained as follows:

$$\mathrm{MSE}_{i,j} = \mathbb{E}\left[\left(c_{i,j} - \hat{c}_{i,j}\right)^2\right] = \frac{\lambda_i \sigma^2}{g_i^2 h_j^2 \lambda_i + \sigma^2}, \tag{4}$$

Here, the total end-to-end distortion can be calculated when N chunks are assigned to N subcarriers as follows:

$$\begin{aligned}\mathrm{MSE}_{total} &= \sum_{i=1}^{N}\sum_{j=1}^{N} b_{i,j}\mathrm{WMSE}_{i,j} \\ &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} b_{i,j}\frac{W_i \lambda_i \sigma^2}{h_j^2 u_i + \sigma^2}}{\sum_{i=1}^{N} W_i},\end{aligned}$$

where $b_{i,j}$ is a binary value denoting whether the i-th chunk is assigned to the j-th subcarrier. It should be noted that $u_i = g_i^2 \lambda_i$ and $W_i = D_s(\boldsymbol{\Theta_i}) S(v, \boldsymbol{X_i})$. In this case, the optimization problem can be expressed as follows:

$$\begin{aligned}&\min_{\{b_{i,j}, g_i\}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} b_{i,j}\frac{W_i \lambda_i \sigma^2}{h_j^2 u_i + \sigma^2}}{\sum_{i=1}^{N} W_i}, \\ &s.t. \sum_{i=1}^{N} u_i \le P, \\ &b_{i,j} \in \{0, 1\}, \sum_{i=1}^{N} b_{i,j} = \sum_{j=1}^{N} b_{i,j} = 1.\end{aligned} \tag{5}$$

The optimization problem is a mixed binary programming problem, which is an NP-hard problem. We divide the problem into two sub-problems (power allocation and subcarrier matching) to find a near-optimal solution. The power allocation problem can be optimally solved, and the subcarrier matching problem can be reformulated as an assignment problem.

2) POWER ALLOCATION

If the subcarrier matching table, $\{b_{i,j}\}$, is given, the optimization problem (5) can be reformulated as

$$\min_{\{b_{i,j},g_i\}} \frac{\sum_{i=1}^{N} \frac{W_i\lambda_i\sigma^2}{h_i^2u_i+\sigma^2}}{\sum_{i=1}^{N} W_i},$$
$$s.t. \sum_{i=1}^{N} u_i \le P,$$

From [56], this optimization problem is convex and can be solved by the Lagrange multiplier as follows:

$$L = \frac{\sum_{i=1}^{N} \frac{W_i\lambda_i\sigma^2}{h_i^2u_i+\sigma^2}}{\sum_{i=1}^{N} W_i} - \gamma(\sum_i u_i - P),$$
$$= \frac{\sum_{i=1}^{N} \frac{(W_i/h_i^2)\lambda_i\sigma^2}{u_i+\sigma^2/h_i^2}}{\sum_{i=1}^{N} W_i} - \gamma(\sum_i u_i - P),$$

By differentiating with u_i and γ, we obtain:

$$\sqrt{\gamma} = \sum_i \sqrt{\frac{(W_i/h_i^2)\lambda_i\sigma^2}{(P+N\sigma^2/h_i^2)\sum_j W_j}},$$
$$u_i = \sqrt{\frac{(W_i/h_i^2)\lambda_i\sigma^2}{\sum_j W_j}} \sum_j \frac{(P+N\sigma^2/h_j^2)\sqrt{\sum_k W_k}}{\sqrt{(W_j/h_j^2)\lambda_j\sigma^2}} - \frac{\sigma^2}{h_i^2}.$$

When σ is small, u_i can be simplified as follows:

$$u_i = \frac{P\sqrt{(W_i/h_i^2)\lambda_i}}{\sum_j \sqrt{(W_j/h_j^2)\lambda_j}}$$

Thus, an optimal scaling factor g_i can be calculated:

$$g_i = \sqrt{\frac{u_i}{\lambda_i}} = \lambda_i^{-1/4} \sqrt{\frac{P\sqrt{W_i/h_i^2}}{\sum_j \sqrt{\lambda_j W_j/h_j^2}}}.$$

The weight matrix, W_i, which consists of the 2D mapping distortions and perceptual distortions of the HVS, will be introduced in the following subsections.

3) 2D MAPPING DISTORTION

In contrast to the conventional 2D videos, a 360-degree video is captured by an omnidirectional camera and mapped onto a sphere domain. The spherical 360-degree videos are then mapped onto the 2D plane using a linear projection technique. Thus, the spherical distortions create an unequal weight between the pixels of the 2D-projected 360-degree videos. Specifically, the pixels, (θ, ϕ), in the spherical domain are projected to the pixels, (x, y), in the 2D-plane domain. In this case, $d_p(x, y)$ represents the distortions between the original and reconstructed pixel values at the location of (x, y) in the 2D-plane domain. The spherical distortion can be defined as follows [44]:

$$D_s(\theta, \phi) = d_p(x, y)^2 \cos(\phi) J(x, y),$$

where $J(x, y)$ is the Jacobian determinant, that is:

$$J(x, y) = \frac{\partial(\theta, \phi)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial\theta}{\partial x} & \frac{\partial\theta}{\partial y} \\ \frac{\partial\phi}{\partial x} & \frac{\partial\phi}{\partial y} \end{vmatrix}.$$

4) HUMAN PERCEPTUAL DISTORTION

Based on the predicted head orientation of the headset user, 360Cast+ determines the user viewport region that will be displayed on the headset from the full resolution of 360-degree videos. We consider that the user foveation point of the viewport is the same as the predicted head orientation. In this case, the user error sensitivity for the pixels within the viewport decreases as the distance between the foveation point and target pixel increases. 360Cast+ adopts the error sensitivity features [56], [57] of the pixel and wavelet domains into the transmission power allocation process to realize better visual quality in the HVS. To make our description more concise, the values of the variables are listed in Table 1.

TABLE 1. Empirical parameters of human perception model.

Parameters	Description	Parameters	Description
δ	0.1	CT_0	1/64
α	0.106	e_2	2.3
v	1 m	β_1	1
β_2	1	L	3

We first introduce the error sensitivity in the pixel domain. The empirical model of contrast sensitivity as a function of retinal eccentricity can be defined as:

$$CS(f, e) = \frac{1}{CT(f, e)} = \frac{1}{CT_0 e^{\alpha f \frac{e+e_2}{e_2}}},$$

where CT_0, α, and e_2 denote the minimal contrast threshold, spatial frequency decay constant, and half-resolution eccentricity constant, respectively. The retinal eccentricity at location x is calculated as follows:

$$e(v, \boldsymbol{X}) = \tan^{-1}\left(\frac{d(\boldsymbol{X})}{Nv}\right),$$

where N and v denote the resolution of the transmission area and viewing distance, respectively. $d(\boldsymbol{X})$ is the distance between point $\boldsymbol{X} = (x, y)$ and the foveation point, (x_f, y_f).

In this case, the error sensitivity in the pixel domain is defined as the normalization of contrast sensitivity as follows:

$$S_f(v, f, \boldsymbol{X}) = \begin{cases} \dfrac{CS(f, e(v, \boldsymbol{X}))}{CS(f, 0)}, & f \le f_m(v, \boldsymbol{X}) \\ \delta, & \text{otherwise}, \end{cases} \tag{6}$$

where δ is the visual sensitivity when the spatial frequency, f, exceeds the threshold.

The cutoff frequency is the minimum value of the critical invisible frequency, f_c, and the display Nyquist frequency, f_d, which can be defined as follows:

$$f_m(\boldsymbol{X}) = \min(f_c, f_d) = \min\left(\frac{e_2 \ln(\frac{1}{CT_0})}{\alpha(e + e_2)}, \frac{\pi N v}{360}\right).$$

The error sensitivity defined in Eq. (6) is then extended to the wavelet domain [57]. The wavelet coefficients provide different perceptual distortions in the four sub-bands, i.e., LL, HL, LH, and HH. In the wavelet domain, the spatial frequency, f, is determined by the wavelet decomposition level, l, i.e., $f = r2^{-l}$, where r is the display resolution. The weight of the error sensitivity, $S_w(l, m)$, in the wavelet domain related to the sub-band (l,m) is presented in Table 2.

TABLE 2. The error sensitivity $S_w(l, m)$ in subband (l, m).

subbands	1	2	3	4	5	6
LL	0.3842	0.3818	0.2931	0.1804	0.0905	0.0372
HL	0.2700	0.3326	0.3019	0.2129	0.1207	0.0558
HH	0.1316	0.2138	0.2442	0.2098	0.1430	0.0791
LH	0.2700	0.3326	0.3019	0.2129	0.1207	0.0558

Finally, the visual sensitivity in the wavelet domain is defined as follows:

$$S(v, \boldsymbol{X}) = [S_w(l, m)]^{\beta_1} \cdot [S_f(v, f, d_{l,m}(\boldsymbol{X}))]^{\beta_2},$$

where β_1 and β_2 denote the weights of s_w and s_f, respectively.

5) SUBCARRIER MATCHING

We obtained the optimal power scaling factor by assuming an optimal subcarrier and chunk matching. To determine the optimal subcarrier matching table, $\{b_{i,j}\}$, between chunk i and subcarrier j, the subcarrier assignment problem must minimize the end-to-end distortions using a matching table, which can be formulated as follows:

$$\min_{\{b_{i,j}\}} \sum_{i=1}^{N}\sum_{j=1}^{N} b_{i,j}\text{WMSE}_{i,j},\ b_{i,j} \in \{0, 1\}, \sum_{i=1}^{N} b_{i,j} = \sum_{j=1}^{N} b_{i,j} = 1.$$

Thus, according to Eq. (5), a chunk should be appropriately assigned to a subcarrier based on the variance and channel gain to decrease the WMSE. Specifically, the chunks with larger variance should be allocated to subcarriers with higher channel gains. 360Cast+ sorts the chunks and subcarriers in the descending order before power allocation; subsequently, it assigns the chunks corresponding to the subcarriers in this order. Fig. 3 illustrates an overview of the subcarrier matching

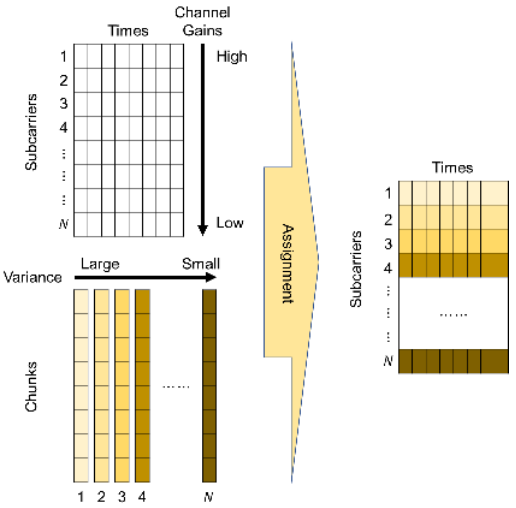


FIGURE 3. Subcarrier matching based on the channel gain of each subcarrier and the variance of each chunk.

operation. 360Cast+ uses a matrix, whose columns and rows correspond to the number of transmission symbols and subcarriers, respectively. The rows are sorted in descending order based on the channel gain, h_j. The encoder also uses vectors of each chunk, y_i, and sorts the vectors in descending order based on the variance. Each vector includes $c_h \times c_w$ elements. The elements of the chunk with higher variance are assigned to OFDM channels with higher channel gain by the encoder in a sequential manner to maximize diversity gain. After the assignment, 360Cast+ allocates the frequency representation of each chunk to OFDM subcarriers based on the matrix. The algorithm determines the average computational complexity of $\mathcal{O}(n^2)$ in comparison to the existing subcarrier assignment algorithm, i.e., an auction algorithm with a computational complexity of $\mathcal{O}(n^2 \log n)$.

D. DECODING

At the receiver side, the received symbols, y_i, of chunk, i, are filtered via an MMSE filter [12] as follows:

$$\widehat{c_i} = \frac{\lambda_i g_i}{\lambda_i g_i^2 + \sigma^2} y_i.$$

360Cast+ utilizes inverse 2D-DWT operations for the filtered symbols to reconstruct the pixel values of the transmitted viewport. Finally, the user headset renders the displayed viewport based on the real head orientation of the user.

E. ANALOG COMPRESSION

In the above designs, we consider that the available baud rate, i.e., bandwidth, is enough to send all the analog-modulated symbols within the viewport. If the available bandwidth and/or time resources are restricted for wireless channel use, the proposed 360Cast+ has to selectively transmit the DWT coefficients to fit the available bandwidth. For such cases, the proposed 360Cast+ discards the chunks in

high-frequency components to fill the bandwidth. When the sender discards a chunk, the receiver regards all coefficients in the chunk as zeros. As a result, data compression can be accomplished even for the proposed 360Cast+.

IV. EVALUATION

A. SIMULATION SETTINGS

1) PERFORMANCE METRIC

We evaluated the performance in terms of the PSNR, structural similarity index measure (SSIM) [58], WS-PSNR [19], IV-PSNR [20], and the proposed weighted PSNR (WPSNR).

PSNR is defined as follows:

$$\mathrm{PSNR} = 10\log_{10}\frac{(2^L-1)^2}{\epsilon_{\mathrm{MSE}}}, \tag{7}$$

where L is the number of bits used to encode the pixel luminance (typically, 8 bits) and $\varepsilon_{\mathrm{MSE}}$ is the MSE between all pixels of the decoded and original videos. SSIM can predict the perceived quality of video streaming. A larger SSIM value, close to 1, indicates higher perceptual similarity between the original and decoded 360-degree video frames. As mentioned in Sec. II-A, WS-PSNR and IV-PSNR represent perceptual quality metrics for 360-degree video defined in CTC.

WPSNR represents the 360-degree video quality considering the 2D projection distortions and human perceptual distortions in the HVS as follows:

$$\mathrm{WPSNR} = 10\log_{10}\frac{(2^L-1)^2}{\mathrm{WMSE}}, \tag{8}$$

2) TEST DATASET

We used three different types of standard reference 360-degree videos, namely, *Mega Coaster*, *Shark*, and *Pacman* with a frame rate of 30 fps, 4:2:0 chroma sampling, and resolution of 3840 × 2048 pixels, along with 50 user head orientations derived from the headset sensors provided in [59]. First, we used the *Mega Coaster* reference 360-degree video and the head orientations of 10 users for comparison; subsequently, the other 360-degree videos in Section IV-E were used to discuss the effect of the 360-degree video categories. We assumed the FoV of the user headset to be 90 degrees × 90 degrees. In this case, the resolution of the viewport was set to 960 × 1024 pixels. We set the chunk size to 32 × 32 pixels for all comparative schemes.

3) WIRELESS SETTINGS

We simulated OFDM channels with 128 subcarriers, whose channel gains included *i.i.d.* Rayleigh distributions, i.e., $h_i \sim \mathcal{N}(0, 1)$. Here, $\sim$ implies "distributed as" and $\mathcal{N}(a, b)$ is a Gaussian distribution with a mean of a and variance of b. The effective noise, n_i, follows a white Gaussian distribution with a variance of σ^2, i.e., $n_i \sim \mathcal{N}(0, \sigma^2)$.

We first set the available bandwidth to 24.3 MHz (= 1084 (coefficients in width) × 998 (coefficients in height) × 1.5 (color channels) × 30 (Hz) × 0.5 (symbol/coefficients). The bandwidth is almost enough to send all the analog-modulated symbols within the predicted viewport in the proposed 360Cast+ since the average region of the predicted viewport is 998 × 1084 pixels. In Sec. IV-C, we discuss the effect of the available bandwidth on the video quality.

B. VIDEO QUALITY

To clarify the benefits of the proposed 360Cast+ from the existing video delivery schemes, we compared the proposed 360Cast+ with the existing digital-based and soft delivery schemes. The digital-based schemes use HM 16.20 for video compression and the modulation format of Binary Phase Shift Keying (BPSK) with 1/2-rate and 1/4-rate convolutional codings, respectively. We prepare three existing soft delivery schemes: 360Cast, ParCast, and SoftCast. Especially, 360Cast+ and 360Cast only deliver a part of the 360-degree video frames, while the other schemes transmit the full resolution of the 360-degree video frames constrained by the same transmission power budget. Besides, 360Cast+ and ParCast implement the proposed subcarrier matching algorithm, while 360Cast and SoftCast implement random subcarrier assignment.

Figs. 4 (a) through (e) show the video quality using the five metrics of PSNR, SSIM, WS-PSNR, IV-PSNR, WPSNR, respectively, as a function of wireless channel SNRs. We can see the following points:

- 360Cast+ prevents cliff effect at low SNR regimes and gracefully improves the received video quality with the improvement of the wireless channel quality.
- 360Cast+ yields better performance in comparison to the existing ParCast and SoftCast schemes because it achieves traffic reduction by only delivering the predicted viewport.
- 360Cast+ also yields better video quality in comparison to 360Cast by assigning the power budget within the predicted viewport considering the viewport prediction error and makes adequate subcarrier matches.
- The digital-based schemes cause cliff effect at a certain wireless channel SNR. This is because a large number of bit errors happens in the received bitstream even with a low-rate channel coding and the errors caused decoding error at the receiver.

For example, the average IV-PSNR and WS-PSNR improvements of the proposed 360Cast+ are 13.1 dB and 12.0 dB compared with the BPSK with a 1/4-rate convolutional coding scheme, respectively, across the wireless channel SNRs of 0 dB through 16dB.

C. EFFECT OF AVAILABLE BANDWIDTH

In this section, we discuss the efficiency of the proposed 360Cast+ considering the available bandwidth. We compare the effect of the available bandwidth with the digital-based 360-degree video delivery schemes. We prepare two digital-based schemes: full resolution-based and viewport-based schemes. The full resolution-based scheme encodes and delivers the full resolution of 360-degree video to the

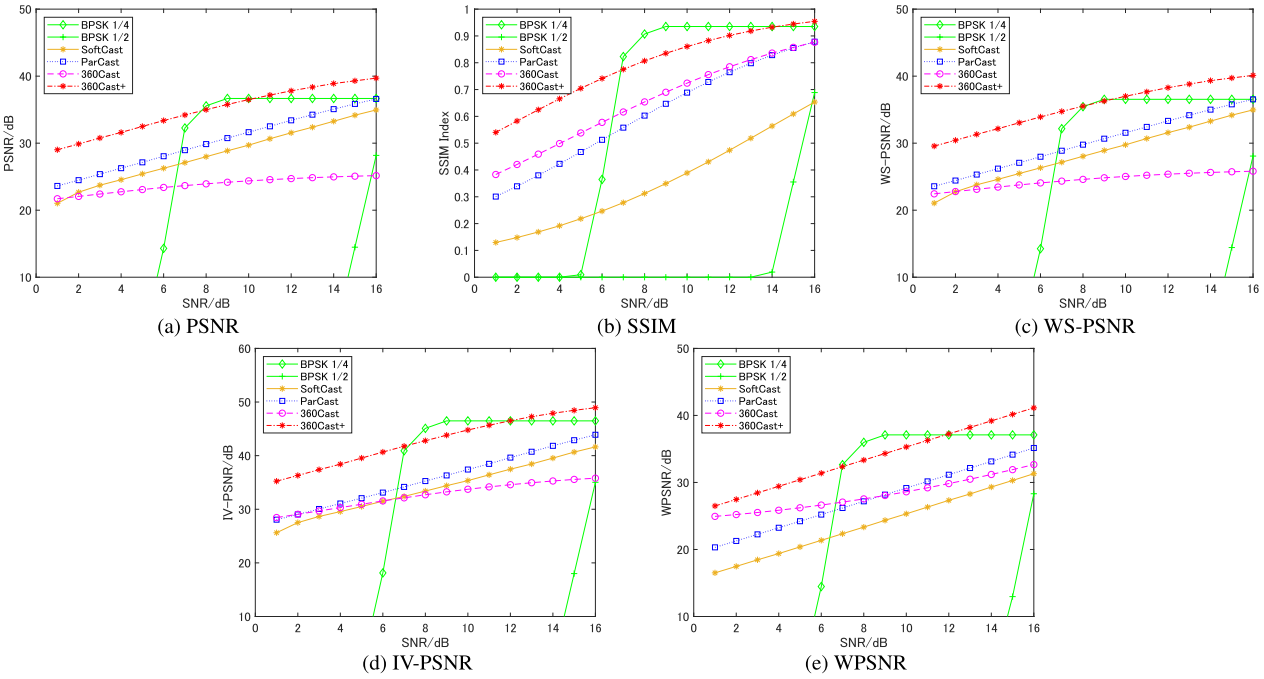


FIGURE 4. Video quality of the comparative schemes as a function of wireless channel SNRs.

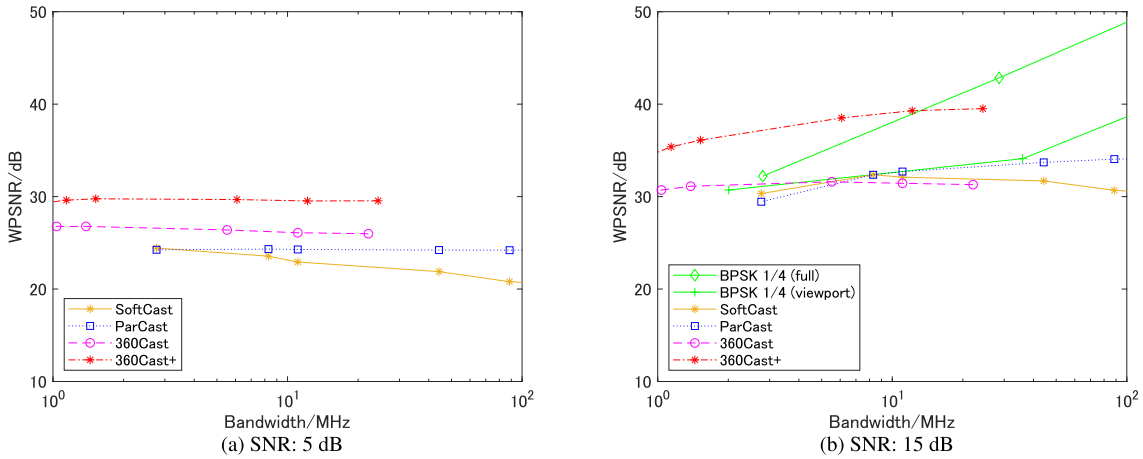


FIGURE 5. Video quality as the function of the available bandwidth at the wireless channel SNRs of 5 and 15 dB, respectively. Here, the digital-based schemes do not reconstruct the 360-degree video at the channel SNR of 5 dB owing to a large number of bit errors.

headset user while the viewport-based scheme only encodes and delivers the user's viewport predicted by the proposed viewport prediction in 360Cast. Here, we use HM 16.20 for video compression and the modulation format of BPSK with 1/4-rate convolutional coding. To discuss the effect of the available bandwidth, the digital-based schemes set QP parameters of 5, 9, 16, 28, and 50 for video coding.

Figs. 5 (a) and (b) show the video quality as a function of the available bandwidth at the wireless channel SNRs of 5 and 15 dB, respectively. We can see the following points:

- The proposed 360Cast+ achieves the best video quality in the band-limited environments irrespective of wireless channel quality.
- The digital-based schemes at a wireless channel of 15 dB achieves better video quality in broadband environments.
- At a low wireless channel SNR of 5 dB, the digital-based schemes do not reconstruct the 360-degree video from the received bit stream owing to a large number of bit errors.

D. EFFECT OF SUBCARRIER ASSIGNMENT

The previous evaluations demonstrated that 360Cast+ yielded better performance than the comparative schemes owing to subcarrier matching. To further discuss the effect of subcarrier matching, we evaluated the viewport quality under different subcarrier matching algorithms, i.e., random, SSRVB, and the proposed subcarrier matching algorithm. The random scheme uniformly assigns a chunk within a predicted viewport to a subcarrier. The SSRVB scheme iteratively assigns a chunk to a subcarrier based on an auction algorithm, proposed in [41]. Each scheme delivers the same-sized predicted viewport to the headset user according to DLR-based viewport prediction.

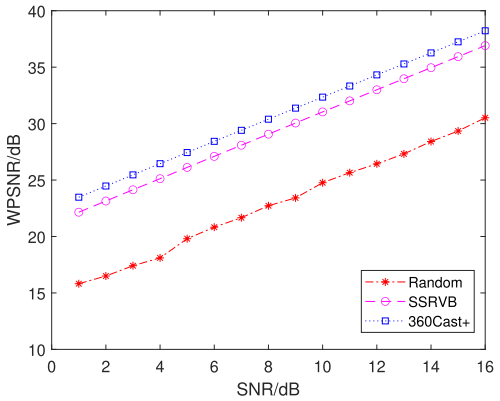
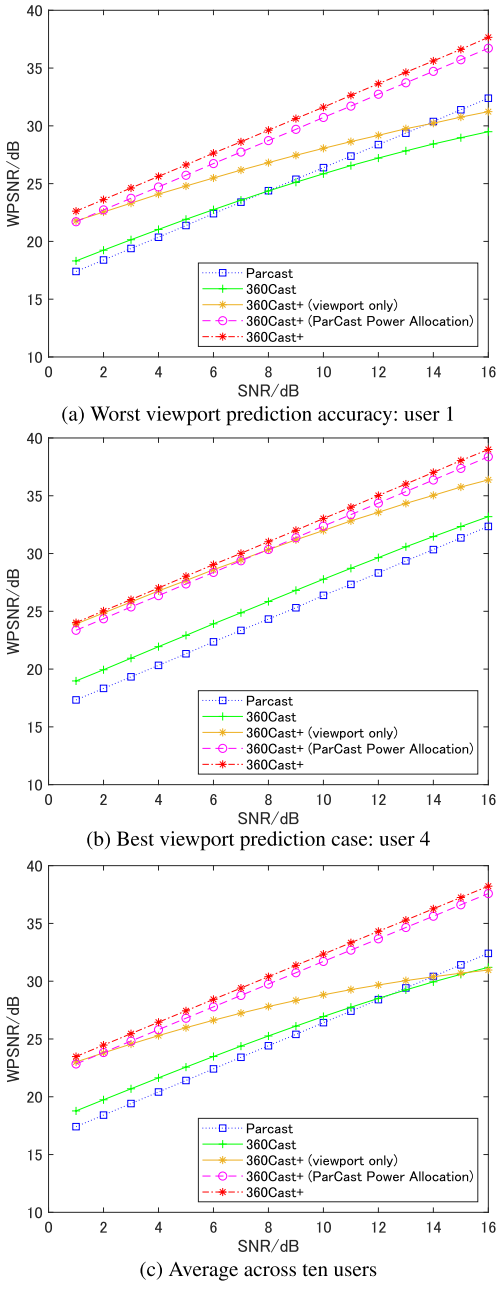


FIGURE 6. Video quality of the proposed 360Cast+ under the different subcarrier matching algorithms as a function of wireless channel SNRs.

Fig. 6 demonstrates the video quality of the proposed 360Cast+ under different subcarrier matching algorithms as a function of wireless channel SNRs. This shows that the proposed subcarrier matching algorithm outperforms other algorithms irrespective of the wireless channel SNRs. For example, the video quality of the proposed subcarrier matching algorithm is 7.6 and 1.3 dB higher than that of random matching and SSRVB at a wireless channel SNR of 10 dB, respectively.

E. EFFECTS OF USER HEAD MOVEMENT AND 360-DEGREE VIDEO SEQUENCES

In this section, we evaluate the performance improvement of the proposed 360Cast+ under the effects of weighted power allocation and viewport prediction error. We first evaluated the video quality of the proposed 360Cast+ and the comparative schemes for 10 different headset users as a function of wireless channel SNRs. Here, 360Cast+ (viewport only) represents our 360Cast+ scheme without considering the viewport prediction error. 360Cast+ (ParCast Power Allocation) performs power allocation without considering the perceptual distortions in the HVS and 2D projection distortions. To compare the effect of user head movements in the comparative schemes, we calculated the average DLR-based viewport prediction error for each headset user during 360-degree video playback. User 4 was observed to be the best headset user



(a) Worst viewport prediction accuracy: user 1

(b) Best viewport prediction case: user 4

(c) Average across ten users

FIGURE 7. Video quality of the comparative schemes across different ten users scenarios as a function of wireless channel SNRs.

with a prediction error of 0.875 pixels, while user 1 was the worst user with a prediction error of 33.375 pixels.

Fig. 7 (a) shows the WPSNR performance of headset user 1 as a function of wireless channel SNRs. The proposed 360Cast+ outperformed the 360Cast+ (viewport only) and 360Cast+ (ParCast Power Allocation) schemes by up to 1.8 and 0.9 dB at a wireless channel SNR of 5 dB, respectively. In addition, the performance difference between the proposed 360Cast+ and 360Cast+ (viewport only) increased as the wireless channel SNRs improved. This is because the viewport prediction error has a higher effect on quality

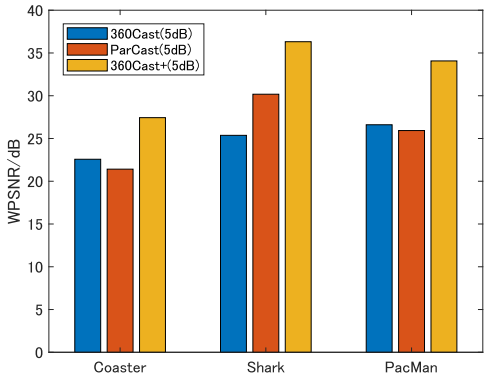


FIGURE 8. Video quality of the comparative schemes for three types of 360-degree videos at wireless channel SNRs of 5 dB.

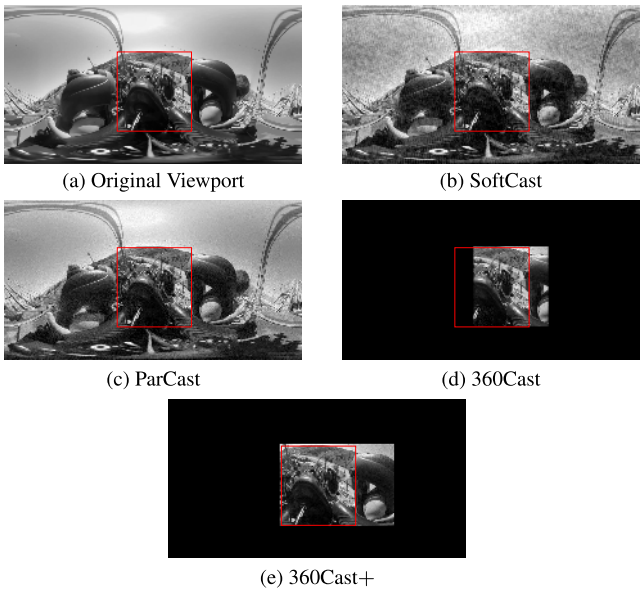


FIGURE 9. Snapshots in each comparative scheme. The black areas in both 360Cast and 360Cast+ represent the un-transmitted regions. Please enlarge the figures to observe details.

degradation in 360Cast, especially in high wireless channel SNR regimes. Fig. 7 (b) shows the WPSNR performance of headset user 4 as a function of wireless channel SNRs and Fig. 7 (c) shows the average WPSNR performance across 10 headset users as a function of wireless channel SNRs. In Fig. 7 (b), it can be observed that 360Cast+ also yields better WPSNR performance than the comparative schemes with DLR-based viewport prediction including 360Cast+ (viewport only) and 360Cast+ (ParCast Power Allocation). This indicates that the proposed 360Cast+ can realize better viewport quality irrespective of the viewport prediction accuracy by extending viewports based on the potential viewport prediction error and scaling them with the perceptual weight.

Fig. 8 evaluates the average video quality of the comparative schemes across 10 headset users for the test 360-degree video sequences of *Mega Coaster*, *Shark*, and *PacMan* at a wireless channel SNR of 5 dB. Each video belongs to three categories, i.e., (i) natural image, fast-paced head movement; (ii) natural image, slow-paced head movement; and (iii) computer graphic (CG) and fast-paced head movement. Accordingly, the following two observations were obtained:

- The proposed 360Cast+ achieves the best performance irrespective of the 360-degree video categories.
- As the speed of the user head movements increases, the received video quality decreases in all comparative schemes owing to the increase in the viewport prediction error.

F. VISUAL QUALITY

Finally, we discuss the visual quality of the comparative schemes. Fig. 9 shows a received frame of each comparative scheme using the test 360-degree video sequence of *Mega Coaster* at a wireless channel SNR of 10 dB. The area within the red rectangle represents the displayed viewport on the user headset. The conventional SoftCast scheme yielded the worst visual quality with numerous blurs; the ParCast scheme achieved better visual quality by utilizing the channel diversity gains. In comparison to the ParCast scheme, 360Cast only delivered the DWT coefficients within the predicted viewport. It demonstrated less visual degradation due to the channel noise by assigning more transmission power to the limited coefficients. However, in this case, the displayed viewport contained a large untransmitted area, i.e., the black rectangle caused by the viewport prediction error. Because the proposed 360Cast+ delivers an extended viewport to reduce the effect of the viewport prediction error, the transmitted viewport covers the red rectangle by only sending the extended viewport.

G. DISCUSSION ON DELAY ISSUE IN VR SYSTEMS

We finally discuss the delay issue on the proposed 360Cast+ and the existing digital-based schemes. The delay issue in VR video streaming is known as event-to-eye delay [60]. The event-to-eye delay mainly consists of the video capturing and stitching delays on the camera, encoding and transmission delays on the server, and decoding and rendering delays on the headset. The main differences of the event-to-eye delay between the proposed 360Cast+ and the digital-based 360-degree video delivery are encoding and decoding delays. We note that as discussed in Sec. IV-C, the proposed 360Cast+ can realize compression according to an available bandwidth by discarding high-frequency coefficients. In this case, the transmission delay between the proposed 360Cast+ and the digital-based schemes can be regarded as the same.

We evaluate the encoding and decoding delays in the proposed 360Cast+ and the digital-based schemes to discuss the effect on the event-to-eye delay. Specifically, we used a 360-degree video of *Mega Coaster* with the resolution of 2048×3840 pixels. Here, we consider the resolution of the viewport is 960×1024 pixels. The specifications of the operating environment are Windows 10 64-bit operating system

TABLE 3. Total delay of encoding and decoding (s) for full resolution and viewport of the 360-degree video using HM 16.20.

QP	5	9	16	28	50
Full resolution	10443.4	9818.6	8991.6	7416.9	6905.5
Viewport	4886.0	4653.1	4208.4	2990.6	1244.1

with Intel (R) Core (TM) i7-8750H CPU and 16 GB memory. The proposed 360Cast+ used MATLAB encoder and decoder while the digital-based schemes used HM 16.20 for 360-degree video encoding and decoding.

Table 3 shows the total delay of encoding and decoding for full resolution and viewport of eight 360-degree video frames using HM 16.20 considering different QP parameters, respectively. Here, the full resolution-based scheme and viewport-based scheme are the same as Sec IV-C. The digital-based schemes integrate nonlinear operations including intra and inter predictions, quantization, and entropy coding for efficient coding while such integration may cause long encoding and decoding delays. On the other hand, the total delay of encoding and decoding for eight 360-degree video frames in the proposed 360Cast+ is approximately 44.5s. The proposed 360Cast+ skips such nonlinear operations, instead, only performs 2D-DWT with power allocation for coding. The proposed 360Cast+ realizes at least a hundred-fold and twenty-fold improvement compared with the full resolution-based and viewport-based digital-based schemes, respectively, and may contribute to reducing the event-to-eye delay in wireless VR video delivery.

V. CONCLUSION

This study proposed 360Cast+ to realize the viewport-adaptive and efficient video delivery of 360-degree videos. 360Cast+ overcame the issues involved in digital-based wireless 360-degree video delivery, i.e., the cliff effect, constant quality, and large perceptual redundancy, by integrating analog modulation, DLR-based viewport prediction, and optimal power allocation considering the joint distortions between 2D projection and human perception. In addition, 360Cast+ reduced the effect of the viewport prediction error and frequency-dependent channel environment in modern OFDM systems by integrating viewport extraction considering the potential prediction error and chunk-subcarrier matching algorithm. The evaluations demonstrated that the proposed 360Cast+ can yield better viewport quality in comparison to the existing digital-based and soft delivery schemes, irrespective of the viewport prediction error, through modern OFDM channels.

One of potential issues in the proposed 360Cast+ is to require sophisticated modulator and demodulator to realize the analog modulation. To carry out the analog modulation in practical scenarios, we will discuss the integration with the System on Chip (SoC) [61], [62] for the proposed 360Cast+ as a future work.

REFERENCES

[1] Y. Bao, H. Wu, T. Zhang, A. A. Ramli, and X. Liu, "Shooting a moving target: Motion-prediction-based transmission for 360-degree videos," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 1161–1170.

[2] A. T. Nasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using scalable video coding," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 1689–1697.

[3] L. Sun, F. Duanmu, Y. Liu, Y. Wang, Y. Ye, H. Shi, and D. Dai, "A two-tier system for on-demand streaming of 360 degree video over dynamic networks," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 43–57, Mar. 2019.

[4] M. Xiao, C. Zhou, Y. Liu, and S. Chen, "OpTile: Toward optimal tiling in 360-degree video streaming," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 708–716.

[5] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "An optimal tile-based approach for viewport-adaptive 360-degree video streaming," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 29–42, Mar. 2019.

[6] D. V. Nguyen, H. T. T. Tran, and T. C. Thang, "Impact of delays on 360-degree video communications," in *Proc. TRON Symp. (TRONSHOW)*, Dec. 2017, pp. 1–6.

[7] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, "Optimizing 360 video delivery over cellular networks," in *Proc. 5th Workshop All Things Cellular, Oper., Appl. Challenges*, Oct. 2016, pp. 1–6.

[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] M. Li, Z. Chen, and Y.-P. Tan, "Scalable resource allocation for SVC video streaming over multiuser MIMO-OFDM networks," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1519–1531, Nov. 2013.

[10] M. M. Ghandi and M. Ghanbari, "Layered H.264 video transmission with hierarchical QAM," *J. Vis. Commun. Image Represent.*, vol. 17, no. 2, pp. 451–466, Apr. 2006.

[11] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 331–345, Feb. 2014.

[12] S. Jakubczak and D. Katabi, "SoftCast: One-size-fits-all wireless video," in *Proc. ACM SIGCOMM Conf.*, 2010, pp. 449–450.

[13] X. L. Liu, W. Hu, Q. Pu, F. Wu, and Y. Zhang, "ParCast: Soft video delivery in MIMO-OFDM WLANs," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw. (Mobicom)*, 2012, pp. 233–244.

[14] Y. Lu, T. Fujihashi, S. Saruwatari, and T. Watanabe, "360Cast: Foveation-based wireless soft delivery for 360-degree video," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–7.

[15] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 159–170, Aug. 2010.

[16] Heinrich-Hertz Institut, Fraunhofer Institute for Telecommunications. *High Efficiency Video Coding (HEVC) Reference Software HM*. Accessed: Feb. 21, 2021. [Online]. Available: https://hevc.hhi.fraunhofer.de/

[17] G. K. S. Y. M. Hannuksela and K. M. Vadakital, *Extractor Design for HEVC Files*, document MPEG2016/M38147, 114th MPEG Meeting of ISO/IEC JTC1/SC29/WG11, 2016, pp. 1–10.

[18] Z. E. G. Y. Y. Y. Gu and P. Higgs, *Multiple Angle VR Streaming*, document MPEG2017/M39994, 117th MPEG Meeting of ISO/IEC JTC1/SC29/WG11, 2016, pp. 1–2.

[19] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1408–1412, Sep. 2017.

[20] A. Dziembowski, *Software Manual of IV-PSNR for Immersive Video*, document mPEG127/n18709, 128th MPEG meeting of ISO/IEC JTC1/SC29/WG11, 2019.

[21] D. V. Nguyen, H. T. T. Tran, and T. C. Thang, "An evaluation of tile selection methods for viewport-adaptive streaming of 360-degree video," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 1, pp. 1–24, Apr. 2020.

[22] C. Ozcinar, A. De Abreu, and A. Smolic, "Viewport-aware adaptive 360° video streaming using tiles for virtual reality," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2174–2178.

[23] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 261–271.

[24] R. Skupin, Y. Sanchez, D. Podborski, C. Hellge, and T. Schierl, "HEVC tile based streaming to head mounted displays," in *Proc. 14th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2017, pp. 613–615.
[25] S. Lee, D. Jang, J. Jeong, and E.-S. Ryu, "Motion-constrained tile set based 360-degree video streaming using saliency map prediction," in *Proc. 29th ACM Workshop Netw. Operating Syst. Support Digit. Audio Video (NOSSDAV)*, 2019, pp. 20–24.
[26] C.-L. Fan, J. Lee, W.-C. Lo, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "Fixation prediction for 360° video streaming in head-mounted virtual reality," in *Proc. 27th Workshop Netw. Operating Syst. Support Digit. Audio Video*, Jun. 2017, pp. 67–72.
[27] Y. Zhang, P. Zhao, K. Bian, Y. Liu, L. Song, and X. Li, "DRL360: 360-degree video streaming with deep reinforcement learning," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2019, pp. 1252–1260.
[28] B. Li, L. Song, R. Xie, and W. Zhang, "Weight-based bit allocation scheme for VR videos in HEVC," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.
[29] Y. Zhou, L. Tian, C. Zhu, X. Jin, and Y. Sun, "Video coding optimization for virtual reality 360-degree source," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 118–129, Jan. 2020.
[30] L. Sun, Y. Mao, T. Zong, Y. Liu, and Y. Wang, "Flocking-based live streaming of 360-degree video," in *Proc. 11th ACM Multimedia Syst. Conf.*, May 2020, pp. 26–37.
[31] H. Cao, J. Lu, and N. Zong, "A multi-user 360-video streaming system for wireless network," in *Proc. 17th Int. Conf. Virtual-Reality Continuum Appl. Ind.*, Nov. 2019, pp. 1–5.
[32] C. Liu, N. Kan, J. Zou, Q. Yang, and H. Xiong, "Server-side rate adaptation for multi-user 360-degree video streaming," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3264–3268.
[33] X. Jiang, S. A. Naas, Y.-H. Chiang, S. Sigg, and Y. Ji, "SVP: Sinusoidal viewport prediction for 360-degree video streaming," *IEEE Access*, vol. 8, pp. 164471–164481, 2020.
[34] H. Yuan, S. Zhao, J. Hou, X. Wei, and S. Kwong, "Spatial and temporal consistency-aware dynamic adaptive streaming for 360-degree videos," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 177–193, Jan. 2020.
[35] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo, "360ProbDASH: Improving QoE of 360 video streaming using tile-based HTTP adaptive streaming," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 315–323.
[36] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, "Optimizing 360 video delivery over cellular networks," in *Proc. 5th Workshop All Things Cellular, Oper., Appl. Challenges*, Oct. 2016, pp. 1–6.
[37] F. Duanmu, E. Kurdoglu, S. A. Hosseini, Y. Liu, and Y. Wang, "Prioritized buffer control in two-tier 360 video streaming," in *Proc. Workshop Virtual Reality Augmented Reality Netw.*, Aug. 2017, pp. 13–18.
[38] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, "How do people explore virtual environments?" 2016, *arXiv:1612.04335*. [Online]. Available: http://arxiv.org/abs/1612.04335
[39] A. Nguyen, Z. Yan, and K. Nahrstedt, "Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1190–1198.
[40] B. Han, Y. Liu, and F. Qian, "ViVo: Visibility-aware mobile volumetric video streaming," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, Apr. 2020, pp. 1–13.
[41] Y. Gui, H. Lu, F. Wu, and C. W. Chen, "Robust video broadcast for users with heterogeneous resolution in mobile networks," *IEEE Trans. Mobile Comput.*, early access, Jun. 1, 2020, doi: 10.1109/TMC.2020.2999195.
[42] J. Zhao, R. Xiong, and J. Xu, "OmniCast: Wireless pseudo-analog transmission for omnidirectional video," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 58–70, Mar. 2019.
[43] S. Aditya and S. Katti, "FlexCast: Graceful wireless video streaming," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2011, pp. 277–288.
[44] X. Fan, F. Wu, and D. Zhao, "D-cast: DSC based soft mobile video broadcast," in *Proc. 10th Int. Conf. Mobile Ubiquitous Multimedia (MUM)*, 2011, pp. 226–235.
[45] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "FreeCast: Graceful free-viewpoint video delivery," *IEEE Trans. Multimedia*, vol. 21, no. 4, pp. 1000–1010, Apr. 2019.
[46] T. Zhang and S. Mao, "Joint power and channel resource optimization in soft multi-view video delivery," *IEEE Access*, vol. 7, pp. 148084–148097, 2019.
[47] J. Wu, B. Tan, J. Wu, and M. Wang, "Video multicast: Integrating scalability of soft video delivery systems into NOMA," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1722–1726, Dec. 2019.
[48] X. Jiang, H. Lu, C. W. Chen, and F. Wu, "Receiver-driven video multicast over NOMA systems in heterogeneous environments," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2019, pp. 982–990.
[49] Y. Li, Z. Li, Y. Liu, and Y. Wang, "SCAST: Wireless video multicast scheme based on segmentation and softcast," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
[50] Z. Zhang, D. Liu, X. Ma, and X. Wang, "ECast: An enhanced video transmission design for wireless multicast systems over fading channels," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2566–2577, Dec. 2017.
[51] T. Fuiihashi, M. Kobavashi, K. Endo, S. Saruwatari, S. Kobayashi, and T. Watanabe, "Graceful quality improvement in wireless 360-degree video delivery," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.
[52] G. Angelopoulos, M. Medard, and A. P. Chandrakasan, "AdaptCast: An integrated source to transmission scheme for wireless sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2894–2899.
[53] J. Shen, L. Yu, L. Li, and H. Li, "Foveation-based wireless soft image delivery," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2788–2800, Oct. 2018.
[54] R. Xiong, H. Liu, S. Ma, X. Fan, F. Wu, and W. Gao, "G-CAST: Gradient based image SoftCast for perception-friendly wireless visual communication," in *Proc. Data Compress. Conf.*, Mar. 2014, pp. 133–142.
[55] H. Liu, R. Xiong, X. Fan, D. Zhao, Y. Zhang, and W. Gao, "CG-cast: Scalable wireless image Softcast using compressive gradient," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 6, pp. 1832–1843, Jun. 2019.
[56] J. Shen, L. Yu, L. Li, and H. Li, "Foveation-based wireless soft image delivery," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2788–2800, Oct. 2018.
[57] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1397–1410, 2001.
[58] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[59] W.-C. Lo, C.-L. Fan, J. Lee, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "360 video viewing dataset in head-mounted virtual reality," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 211–216.
[60] J. Yi, M. R. Islam, S. Aggarwal, D. Koutsonikolas, Y. C. Hu, and Z. Yan, "An analysis of delay in live 360° video streaming systems," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 982–990.
[61] F. Gao, H. Gao, and J. Wu, "A reconfigurable SoC for SoftCast wireless video transmission," in *Proc. IEEE Int. Conf. Ind. Internet (ICII)*, Oct. 2018, pp. 169–170.
[62] F. Gao, H. Ren, Z. Zhang, J. Wu, W. Yu, and J. Shen, "A heterogeneous SoC for soft cast wireless video transmission," in *Proc. IEEE 18th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2018, pp. 688–694.

LU YUJUN received the B.E. degree from Zhejiang University, China, in 2018. He is currently pursuing the M.E. degree with the Graduate School of Information Science and Technology, Osaka University, Japan. His research interests include the areas of wireless networks and video transmission.

TAKUYA FUJIHASHI (Member, IEEE) received the B.E. and M.S. degrees from Shizuoka University, Japan, in 2012 and 2013, respectively, and the Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan, in 2016. From 2014 to 2015, he was an Intern with the Mitsubishi Electric Research Laboratories (MERL), Electronics and Communications Group. He is currently an Assistant Professor with the Graduate School of Information Science and Technology, Osaka University. His research interests include video compression and communications, with a focus on multi-view video coding and streaming. From 2014 to 2016, he was a Research Fellow (DC1) of the Japan Society for the Promotion of Science. He was Research Fellow (PD) of the Japan Society for the Promotion of Science, in 2016. He selected one of the Best Paper candidates in IEEE International Conference on Multimedia and Expo (ICME) 2012.

SHUNSUKE SARUWATARI (Member, IEEE) received the B.E. degree from The University of Electro-Communications, Japan, in 2002, and the M.S. and Ph.D. degrees from The University of Tokyo, Japan, in 2004 and 2007, respectively. In 2007, he was a Visiting Researcher with the Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana–Champaign. From 2008 to 2011, he was a Research Associate with the Research Center for Advanced Science and Technology, The University of Tokyo. From 2012 to 2015, he was a Tenure-Track Assistant Professor with the Graduate School of Informatics, Shizuoka University, Japan. He is currently an Associate Professor with the Graduate School of Information Science and Technology, Osaka University, Japan. His research interests include the areas of wireless networks, sensor networks, and systems software. He is a member of ACM, IPSJ, and IEICE.

TAKASHI WATANABE (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Osaka University, Japan, in 1982, 1984, and 1987, respectively. He joined the Faculty of Engineering, Tokushima University, in 1987, and moved to the Faculty of Engineering, Shizuoka University, in 1990. He was a Visiting Researcher with the University of California, Irvine, from 1995 to 1996. He has been a Professor with the Graduate School of Information Science and Technology, Osaka University, since 2013. His research interests include mobile networking, ad hoc sensor networks, the Internet of Things/M2M networks, and intelligent transport systems, esspecially medium access control and routing. He is a member of IPSJ and IEICE. He has served on program committees for many networking conferences, such as the IEEE, ACM, IPSJ, and IEICE.