360Cast: Foveation-Based Wireless Soft Delivery for 360-Degree Video

Lu Yujun*, Takuya Fujihashi*, Shunsuke Saruwatari* and Takashi Watanabe*
*Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

***Abstract*—Wireless 360-degree video delivery provides virtual reality (VR) immersive experience where each user can freely switch his/her viewing orientation. The existing schemes of the wireless 360-degree video streaming use digital-based compression and transmission. However, they have many disadvantages in terms of video traffic and quality: cliff effect due to unstable wireless channels, large video traffic due to the extremely high resolution of the 360-degree video, and the perceptual redundancy within the transmitted video. To solve the above problems, this paper proposes a novel wireless 360-degree video transmission scheme called $360Cast$. 360Cast adopts the analog-based transmission, including power allocation and analog modulation, to achieve graceful video quality improvement even in time-varying wireless channels. Here, the power allocation considers the distortion of human perception and sphere-to-plane projection to maximize the human perceptual quality at the 360-degree video playback. In addition, 360Cast predicts and transmits the user's viewport based on the recent HMD user's orientation by using dynamic linear regression (DLR) and only transmits the viewport for traffic reduction. Evaluation results show that the proposed 360Cast provides better human perceptual quality via HMD compared with the existing analog transmission schemes, i.e., SoftCast and FoveaCast, by using the integration of viewport prediction, power allocation, and analog modulation.**

I. Introduction

With the rapid development of electronics and computer technology, the huge computational power in high-end devices made virtual reality (VR) a possibility. However, the existing head-mounted display (HMD), such as Oculus Rift and HTC VIVE, needs cables to connect to the computing device. The wired connection seriously affects the user's immersive experience since it limits the user's movement. To provide a higher immersive experience for the users, many researchers pay more attention to realize wireless VR.

High-quality wireless 360-degree video delivery is a key component to provide good immersion in wireless VR applications. 360-degree videos, also known as immersive videos or spherical videos, are captured by an omnidirectional camera. During playback on the HMD, the user freely switches his/her viewing orientation. The resolution requirement of the 360-degree video is even high, such as 8K ($7,680 \times 4,320$) with 60 fps and more, to reproduce 3D scenes on human eyes.

The efficient and reliable delivery of such 360-degree videos over a wireless network is a challenging issue. Firstly, the wireless channel state will fluctuate drastically when the HMD user switches his/her viewing angle during a 360-degree video playback. In conventional 360-degree video delivery, they use digital video compression and digital wireless transmission for video frames in a sequence. For example, the video compression part uses H.264/Advanced Video Coding (AVC) [1] or H.265/High-Efficiency Video Coding (HEVC) [2] standards to generate a compressed bitstream using quantization and entropy coding. The wireless transmission part uses channel coding and a digital modulation scheme to reliably transmit the encoded bitstream. However, the reconstructed video quality of conventional digital transmission schemes via unstable wireless channels is significantly low due to the cliff effect. Specifically, when the channel SNR falls beneath a certain threshold, possible few bit errors occurred in the encoded bitstream during communications can cause a synchronization problem in entropy decoding and the reconstructed video quality drops drastically. A novel analog coding scheme called SoftCast has been proposed to solve this problem [3]–[7]. SoftCast skips the non-linear coding operations in the digital coding operations such as quantization, entropy coding, and channel coding. Instead, SoftCast performs only the linear operations including discrete cosine transform (DCT) or discrete wavelet transform (DWT) and power allocation. Owing to the nature of linear operations, the video quality of SoftCast is proportional to instantaneous wireless channel quality.

Secondly, the existing analog schemes are not designed for 360-degree videos. In the conventional analog transmission schemes, the sender sends the full 360-degree videos and it will cause large video traffic. However, each HMD user does not require full 360-degree videos. Specifically, the HMD user only watches a part of the received 360-degree video at any given time. This means only the part of the 360-degree frames, called viewport, should be delivered to the HMD user, which is typically 20% of the full 360-degree videos [8]. Some researchers focus on viewport-based 360-degree video delivery [9]–[11] or tile-based 360-degree video delivery [12]–[16] to balance video traffic and quality.

Thirdly, the conventional digital and analog transmissions suffer from low video quality due to the distortions of 2D projection and human perception. Each 360-degree video is mapped from the spherical surface to a two-dimensional (2D) plane using a certain projection, e.g., Equirectangular Projection (ERP). Each projection generates a non-uniform sampling grid, and thus it will cause unequal distortion across the 2D-projected 360-degree pixels. The unequal distortion due to 2D plane projection should be considered for efficient 360-degree video delivery. In addition, the viewport still remains the perceptual redundancy, which is referred to as the redundant information that cannot be perceived by human vision, and bring low perceptual quality for the HMD user.

For solving the above three problems, this paper proposes a novel analog wireless delivery system called 360Cast for 360-degree video delivery. To cope with time-varying wireless channels, 360Cast adopts analog transmission to achieve graceful video quality. Specifically, 360Cast skips the nonlinear operations of the quantization and entropy coding at the sender's operation, instead, it directly maps the transformed coefficients to the transmission symbols. In addition, 360Cast predicts the viewport based on the past HMD's user orientation data by using dynamic linear regression (DLR) and only delivers the viewport for traffic reduction. In this case, the proposed scheme minimizes the perceptual redundancy and distortion of 2D projection in the predicted viewport by optimizing the power allocation to provide the highest visual perceptual quality for 360-degree video playback. From the evaluations, the proposed 360Cast achieves the best video quality under the consideration of 2D mapping distortion and perceptual redundancy. The contributions of our work are summarized as follows:

- We integrate the Human Vision System (HVS) and distortion of 2D plane projection for the power allocation of 360-degree videos to achieve better perceptual quality for HMD user.
- Our scheme uses DLR to only deliver the predicted viewport for traffic reduction. Since the DLR can reduce the prediction error, the proposed 360Cast keeps better reconstruction quality compared with the conventional analog transmission scheme with the conventional linear regression (LR)-based viewport prediction.

II. Related Works

A. 360-Degree Delivery

The adaptive 360-degree video streaming has been widely studied in recent years to efficiently deliver the 360-degree videos over networks. The 360-degree videos are firstly projected into a 2D plane by using projection methods, such as equirectangular, cube-map, and pyramid projection. The delivery of full 360-degree may waste a mass of bandwidth due to large video traffic. The existing methods on 360-degree video delivery divide the projected video into multiple tiles and transmit only a subset of the tiles to a user. Here, the existing adaptive 360-degree video streaming studies are mainly classified into two aspects: tiling-based and viewport-dependent streaming.

In tiling-based 360-degree video streaming, the tiles within the user's viewport are encoded into higher quality while the tiles out of the user's viewport are encoded into lower quality. To this end, [9] uses Scalable Video Coding (SVC) to encode the tiles into two layers to improve the quality of the viewport: base layer and enhancement layer. For quality optimization of tiling-based video streaming, [10] designs a tile-rate optimization and [11] considers storage costs for each tile optimization.

In viewport-dependent streaming, they only send the predicted user's future viewport for traffic reduction. For example, a head movement-aware streaming [12] and gaze-aware streaming with an in-built eye tracker [13] have been proposed under the consideration of viewport prediction. Furthermore, by leveraging sensor- and content-related features, a fixation prediction network has been proposed in [14], [15] to predict the viewer fixation in the future. In addition, [16] observes the user's motion patterns when multiple users watch the same 360-degree video and proposes a multicast strategy for multi-user 360-degree video delivery based on motion prediction.

The proposed 360Cast is one of the viewport-dependent 360-degree video streaming. For the viewport prediction with high accuracy, we design the DLR-based viewport prediction. Different from the existing studies, the proposed 360Cast adopt an analog transmission for solving the cliff effect due to channel quality fluctuation. In addition, 360Cast considers both visual perceptual quality and 2D projection distortion to optimize the human perceptual quality of the 360-degree video playback.

B. HVS-based Delivery

Human vision covers a field of view of 135×160 degrees while the highest-resolution foveal vision covers only the central 1.5×2 degrees. In the HMD, only 4% of the pixels are estimated to be mapped onto the foveation [17]. Under the consideration of the HMD features, the delivery of the full 360-degree videos causes large traffic due to the perceptual redundancy. The modern video compression schemes consider the perceptual redundancy to improve the perceptual coding efficiency. A human perception-based video compression [17] introduces four human perception models: region-of-interest (ROI), visual attention, visual sensitivity, and cross-modal attention. [18] designed foveation-based digital video delivery with scalable video coding. FoveaCast [7] utilized the HVS model of the contrast sensitivity [19] and analog transmission to achieve a better human perceptual quality of wireless image delivery.

However, none of the existing studies integrate the HVS model into 360-degree video delivery. The proposed 360Cast adopts the human perceptual distortion for the power allocation of the viewport delivery to optimize the human perception during 360-degree video playback.

C. Soft Delivery

SoftCast [3] is an analog transmission scheme to provide graceful quality improvement with the improvement of wireless channel quality. In recent years, many analog transmission schemes [4]–[7] have been proposed for different wireless video delivery scenarios.

For example, ParCast [4] considers both source and MIMO-OFDM channel components exhibit non-uniform energy distribution and allocates power weights with joint consideration to the source and the channel modulation for soft delivery. FlexCast [5] modifies the SoftCast and transmits a continuous stream of encoded video bits for binary representation of DCT coefficients. They divide video bits into the bit groups with

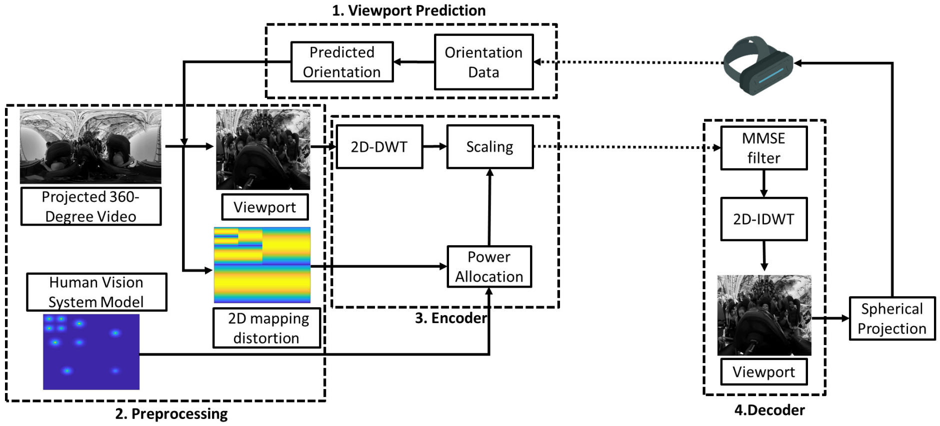


Fig. 1. Overview of the proposed 360Cast.

equal importance and add protection to each group according to their relative importance by adopting rateless coding. FoveaCast [7] considers HVS for wireless image delivery to improve human perceptual quality. OmniCast [6] is the analog transmission scheme under the consideration of wireless 360-degree video delivery. They propose two algorithms to find the block partition with the least projection distortion for different omnidirectional video projections.

Although OmniCast also proposed the analog transmission scheme for the 360-degree videos, they have large traffic since the sender sends the full resolution of 360-degree videos. The proposed 360Cast only delivers the predicted viewport for traffic reduction and uses power allocation under the consideration of human perception and 2D mapping distortion to maximize the human perceptual quality in 360-degree video playback via wireless HMD.

III. Proposed 360Cast

A. Overview

This paper proposes a novel analog wireless transmission scheme, namely, 360Cast for 360-degree video delivery. Fig. 1 shows an overview of the proposed 360Cast scheme.

360Cast first uses the DLR-based viewport prediction to extract the future viewport of the HMD user from the full 360-degree videos. For the prediction, HMD can record the three features of the user's HMD orientation, i.e., pitch, yaw, and roll. The orientation data are transmitted from the user's HMD to the video server during video playback. The video server predicts the future orientation of the HMD user and decides the foveation point within the full 360-degree video. We consider the foveation point is the center of the viewport. Based on the predicted foveation point, the corresponding viewport is extracted from the full 360-degree video which is already mapped from the spherical surface to the 2D plane. At the same time, the 2D projection and human perceptual distortion of the viewport is calculated based on the predicted foveation point for the following power allocation.

At the video server, the extracted viewport is transformed into the DWT coefficients by using 2D-DWT with Daubechies 9-tap/7-tap filter. The DWT coefficients are then divided into multiple chunks with a smaller size and scaled by a chunk-wise power allocation before the transmissions. The power allocation algorithm of 360Cast optimizes the human perception quality via HMD under the consideration of the distortion of the sphere to 2D plane projection and human perception.

At the decoder in the user's HMD, the minimum mean square error (MMSE) can be provided an optimal linear estimate of the scaled DWT coefficients. The user's HMD reconstructs the viewport from the filtered DWT coefficients by using inverse 2D-DWT and converts the reconstructed viewport into the spherical plane according to the HMD user's orientation.

B. Viewport Prediction

After receiving three features of the orientation from the HMD user, the server predicts the future foveation based on the received features by using DLR. The proposed 360Cast needs to predict the X and Y rotation angles from the features of the yaw and the pitch because 360Cast is designed for 2D-projected 360-degree video frames. Let x_{t_0} represents the pitch or yaw at the current time t_0. The predicted foveation point at the time $t_0 + T_p$ can be defined as follows:

$$\widehat{x}_{t_0+T_p} = f\left(\overrightarrow{x}_{t_0-T_w}\right), \tag{1}$$

where T_p and T_w are the predicted time and the dynamic window size for prediction, respectively. Furthermore, $f(*)$ is a linear regression prediction. Since the HMD user may suddenly move the viewing orientation according to the content of the video, the accuracy of the conventional LR-based prediction will decrease due to the sudden movement. The proposed 360Cast finds the nearest inflection point from the past yaw and pitch values and only uses the values following the inflection point for the prediction. To find the inflection point from the past yaw and pitch values, the proposed 360Cast takes recent three values x_{t_0-j-1}, x_{t_0-j-2}, x_{t_0-j-3}, in each orientation feature. Here, j is the candidate of the inflection point. We find the index of inflection point j until satifying

$(x_{t_0-j-1} - x_{t_0-j-2}) \times (x_{t_0-j-3} - x_{t_0-j-2}) > 0$. Here, the dynamic window T_w is the same as the distance from the inflection point.

C. Encoder

In 360Cast, the viewport is extracted from each 2D-projected 360-degree video frame based on the predicted foveation point. 360Cast can obtain its wavelet coefficients by using 2D-DWT with a 9-tap/7-tap filter. The sender then divides the DWT coefficients into the equal-sized chunks. Each chunk c_i are scaled according to g_i, where g_i is the scaling factor of i-th chunk. Finally, the sender transmits each scaled chunk $x_i = g_i c_i$ through the wireless channel. Here, 360Cast needs to find the optimal scaling factor for each chunk to achieve the best human perceptual quality via HMD. For finding such optimal scaling factor, we define a weighted metric called weighted peak signal-to-noise ratio (WPSNR) as follows:

$$\mathrm{WPSNR} = 10 \log_{10} \frac{255^2}{\mathrm{WMSE}},$$

where weighted mean square error (WMSE) is the weighted mean square error between the original and reconstructed 360-degree video frames under the consideration of 2D mapping distortion D_s and human perceptual distortion S as follows:

$$\mathrm{WMSE} = \frac{\sum_{i=1}^{M} S(v, x_i) \cdot D_s(x_i) \cdot E(x_i)}{\sum_{i=1}^{M} S(v, x_i) \cdot D_s(x_i)}.$$

Here, the square error $E(x_i)$ of i-th chunk can be obtained by g_i [7] as follows:

$$E(x_i) = \frac{\lambda_i \sigma^2}{g_i^2 \lambda_i + \sigma^2},$$

where λ_i is the variance of i-th chunk and σ^2 is the variance of the Gaussian noise, respectively. Finally, optimizing the scaling factor g_i can be turned into the following optimization problem:

$$\begin{aligned} &\min \ \mathrm{WMSE} = \frac{\sum_{i=1}^{M} \frac{S_i D_i \lambda_i \sigma^2}{g_i^2 \lambda_i + \sigma^2}}{\sum_{i=1}^{M} S_i D_i} \\ &s.t. \ \sum_i u_i \leq P, u_i \geq 0, \end{aligned}$$

where u_i is the power of all the DWT coefficients after power allocation, that is, $u_i = \sum x_i^2$ and P is the total transmission power.

From [7], this problem can be solved by the Lagrange multiplier and g_i can be formulated as follows:

$$g_i = \sqrt{\frac{\sqrt{\frac{s_i D_i \sigma^2}{\lambda_i \sum_j s_j D_j}}(P + M\sigma^2)}{\sum_j \sqrt{\frac{s_j D_j \sigma^2}{\lambda_j \sum_k s_k D_k}}} - \frac{\sigma^2}{\lambda_i}}.$$

For a high SNR case, the above equation can be simplified as

$$g_i = \lambda_i^{-1/4} \sqrt{\frac{P\sqrt{s_i D_i}}{\sum_j \sqrt{\lambda_j s_j D_j}}}.$$

D. Decoder

At the receiver side, the received symbols can be modeled as follows:

$$z_i = x_i + n_i,$$

where n_i is an effective noise having a variance of σ^2. The received DWT coefficients are filtered via a MMSE filter as follows:

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

Finally, the reconstructed pixel values can be obtained from the filtered DWT coefficients by using inverse 2D-DWT operation. The HMD reconstructs the viewport according to the predicted foveation, which is transmitted from the sender as the metadata, and the reconstructed pixel values. After computing the distance between the real orientation (x_r, y_r) and the predicted orientation (x_p, y_p) received from the server, the HMD renders the adjusted viewport on the display, which is a slight translation from the reconstructed viewport.

E. 2D Mapping Distortion

360-degree video frames are captured by an omnidirectional camera and mapped onto a sphere. The sphere 360-degree videos are then mapped onto the 2D plane using a certain projection technique. Specifically, the point (θ, ϕ) in the spherical domain is projected to the point (x, y) in the 2D plane domain. Let $d_s(\theta, \phi)$ and $d_p(x, y)$ represent the distortion between the original and reconstructed pixel values at the location of (θ, ϕ) in the spherical domain and (x, y) in the 2D plane domain, respectively. The spherical distortion can be defined as follows [6]:

$$\begin{aligned} D_s &= \iint_{\theta,\phi} d_s(\theta, \phi)^2 \cos(\phi) d\theta d\phi \\ &= \iint_{x,y} d_p(x, y)^2 \cos(\phi) J(x, y) dx dy, \end{aligned} \tag{2}$$

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

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

According to Eq.(2), the spherical distortion $d_s(\theta, \phi)$ at the point of (θ, ϕ) can be defined by the product of distortion at the corresponding point (x, y) in the 2D plane domain and a weight determined by θ and ϕ.

After mapping the 360-degree video frame onto the 2D plane, the proposed scheme extracts the viewport from the 2D-projected 360-degree video frame and divides the viewport into the equal-sized chunks based on the predicted foveation calculated in (1).

F. Human Perceptual Distortion

Within the viewport, we consider the human perception in the pixel and wavelet domains [7], [20]. In the pixel domain,

TABLE I
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

the error sensitivity is the normalization of contrast sensitivity as follows:

$$S_f(v,f,x) = \begin{cases} \frac{CS(f,e(v,x))}{CS(f,0)}, & f \leq f_m(v,x) \\ \delta, & \text{otherwise}, \end{cases} \tag{3}$$

Visual sensitivity is set to $\delta = 0.01$ when spatial frequency f exceeds threshold.

The empirical model of the contrast sensitivity as a function of the retinal eccentricity is given as follows:

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

where CT_0, α, and e_2 are minimal contrast threshold, spatial frequency decay constant, and half-resolution eccentricity constant, respectively. The best fitting parameter values are $CT_0 = 1/64$, $\alpha = 0.106$, $e_2 = 2.3$. The retinal eccentricity at location x is calculated as follows:

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

where N and v are the size of the viewport in pixel domain and viewing distance, respectively. In addtion, $d(x)$ is the distance between the given point $x = (x_1, y_1)$ and the foveation point (x_f, y_f). Considering the display of HMD is closer to the user, v is set to 1 m.

The cutoff frequency is obtained by minimizing the critical invisible frequency f_c and the display Nyquist frequency f_d as follows:

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

360Cast extends the error sensitivity defined in Eq. (3) to the wavelet domain. As introduced in [20], the wavelet coefficients provide the different perceptual distortion in the different subbands of LL, HL, LH, HH. In addition, the spatial frequency f and the perceptual weight are affected by the wavelet decomposition level l, that is, $f = r2^{-l}$ where r is the display resolution. In this case, the error sensitivity in the wavelet domain $S_w(l,m)$ related to subband (l,m) is shown in Table. I.

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

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

where β_1 and β_2 are used for the weight of s_w and s_f. Here, β_1 and β_2 are set to 1.

IV. EVALUATION

A. Settings

Performance Metric: We evaluate the performance in terms of the PSNR and WPSNR, which is defined as Eq. 2. PSNR is defined as follows:

$$\text{PSNR} = 10\log_{10}\frac{(2^L-1)^2}{\epsilon_{\text{MSE}}}, \tag{4}$$

where L is the number of bits used to encode pixel luminance (typically eight bits), and ε_{MSE} is the MSE between all pixels of the decoded and the original video.

Test dataset: We use a standard reference 360-degree video, namely, Mega Coaster with 30 fps and a resolution of 3840×2048 as well as the orientation features derived from the HMD sensors offered in [21]. We assume that a Field of View (FoV) of the 360-degree video is 90 degree $\times$ 90 degree, thus, the resolution of each viewport is set to 960×1024.

Encoder: We set the GoP size to eight video frames for all reference schemes. Each GoP contains one I-frame and the subsequent seven P-frames. We set the chunk size to 32×32 pixels across the proposed 360Cast and other reference schemes.

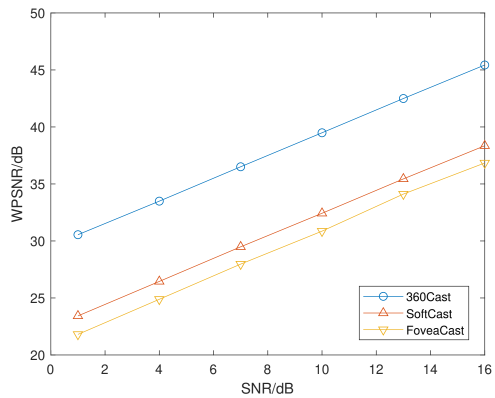
B. Video Quality

We first evaluate the performance of the proposed 360Cast in comparison to two existing analog transmission schemes: SoftCast [3] and FoveaCast [7]. Here, the existing schemes transmit the full 360-degree video frames while 360Cast only delivers the predicted viewport. Fig. 2 shows the average WPSNR and PSNR performance across the video frames as a function of wireless channel SNRs. In this case, we assume that the prediction error between the predicted and actual viewports is zero. We can see that the proposed 360Cast yields better WPSNR and PSNR performance compared with the other reference schemes. For example, the proposed 360Cast improves WPSNR performance by 7.1 dB and 8.7 dB compared to the FoveaCast and SoftCast, respectively, at the wireless channel SNR of 1dB. In view of FoveaCast and SoftCast, FoveaCast has a higher performance of WPSNR but the lower performance of PSNR compared with SoftCast.

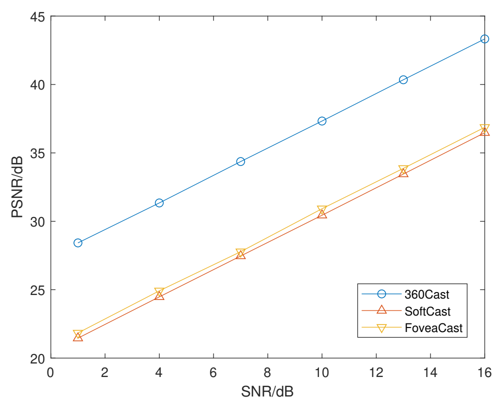
C. Effect of Viewport Prediction Error

In the previous section, the reconstruction quality of the proposed 360Cast outperforms the existing schemes at the perfect prediction case. However, the proposed DLR-based prediction is difficult to predict the user's foveation with 100% accuracy and the prediction error may cause quality degradation. In the preliminary evaluations, we observe that the average prediction error of LR and DLR are (10.3, 9.1) and (7.8, 7.1), respectively, in the horizontal and vertical direction.

We then evaluate the video quality under the consideration of the prediction error. Here, the proposed 360Cast sends the extracted viewport based on the predicted foveation point (x_p, y_p) and the HMD user will see the viewport based on the correct foveation point (x_r, y_r). When some pixels of the requested viewport do not contain within the predicted



(a) WPSNR



(b) PSNR

Fig. 2. Video quality of the proposed 360Cast and the conventional SoftCast and FoveaCast as a function of wireless channel SNRs. Here, we consider the viewport prediction is error-free.

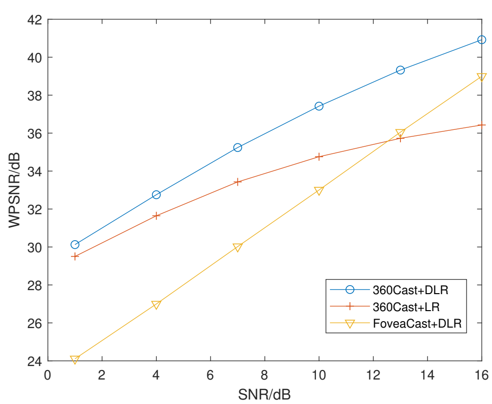


Fig. 3. Video quality of the proposed 360Cast and the conventional SoftCast and FoveaCast as a function of wireless channel SNRs. Here, we consider the DLR-based viewport prediction contains some errors.

viewport, we set the pixel values to zeros. We also add a reference scheme of 360Cast with the LR-based viewport prediction to discuss the effect of the prediction methods on the reconstruction quality. All the schemes use the correct

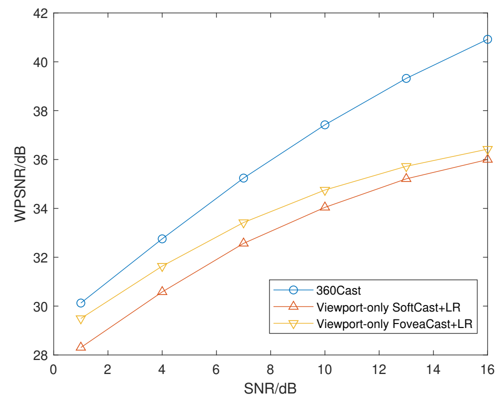


Fig. 4. Video quality of the proposed 360Cast and the viewport-only SoftCast and FoveaCast as a function of wireless channel SNRs.

foveation point (x_r, y_r) for WPSNR calculation.

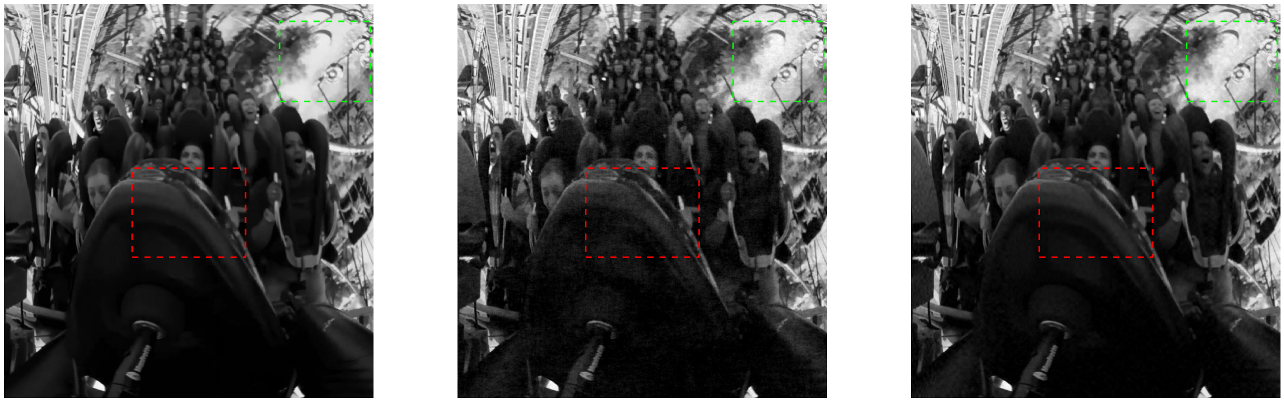
Fig. 3 shows the video quality as a function of wireless channel SNRs under the consideration of the prediction error. The proposed 360Cast with the DLR-based viewport prediction outperforms the other reference schemes because of high prediction accuracy. Especially, the proposed DLR-based viewport prediction achieves a lower quality degradation compared with the LR-based viewport prediction at high wireless channel SNR regimes. It is demonstrated that 360Cast with LR-based viewport prediction is higher than FoveaCast below wireless channel SNRs of 12 dB.

D. Discussion on Viewport Only Delivery

In the previous sections, we consider the conventional FoveaCast send the full 360-degree video frames. In this section, we evaluate the performance of the proposed 360Cast and the viewport-only SoftCast and FoveaCast to discuss the video quality under the same amount of traffic. Here, viewport-only SoftCast and FoveaCast adopt the LR-based viewport prediction to realize viewport-only delivery. Fig. 4 shows the video quality of the viewport-only reference schemes as a function of wireless channel SNRs. We can see the proposed 360Cast still outperforms the other viewport-only schemes irrespective of wireless channel SNRs. For example, the video quality of the proposed 360Cast is 1.8 dB higher than the viewport-only FoveaCast and 2.3 dB higher than the viewport-only SoftCast at the wireless channel SNR of 10 dB.

E. Visual Quality

Finally, we discuss the visual quality of the proposed 360Cast and conventional SoftCast. Fig. 5 shows the extracted viewport of each scheme at the wireless channel SNR of 10 dB. Since the proposed 360Cast uses the power allocation based on 2D mapping distortion and human perceptual distortion, the center of the viewport, i.e., dotted red region, is less visual degradation by channel noise in the proposed 360Cast in comparison to the conventional SoftCast. On the other hand, the green region in the proposed 360Cast seems the same visual quality as the conventional SoftCast.



(a) Original Viewport　(b) SoftCast　(c) Proposed 360Cast

Fig. 5. Snapshots in each reference scheme.

V. Conclusion

This paper proposed 360Cast to provide a graceful video quality improvement for wireless 360-degree video delivery. 360Cast overcomes the issues of wireless 360-degree video delivery, i.e., cliff effect, large traffic, and perceptual redundancy, by integrating analog modulation, DLR-based viewport prediction, and optimal power allocation. Evaluations demonstrated that the proposed 360Cast can yield better video quality compared with the conventional SoftCast and FoveaCast at low wireless channel SNR regimes even with the prediction error in the DLR-based viewport prediction. It means the proposed scheme can provide better visual experience via HMD even in unstable wireless channel environments.

In future works, we will consider 1) arbitrary shape of viewport for scalability and 2) an intelligent viewport prediction for further performance improvement.

References

[1] W. Thomas, S. G. J, B. Gisle, and L. Ajay, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions of Circuirts And Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[2] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," in *IEEE PCS*, 2013, pp. 394–397.

[3] S. Jakubczak and D. Katabi, "Softcast: one-size-fits-all wireless video," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 449–450, 2011.

[4] X. L. Liu, W. Hu, Q. Pu, F. Wu, and Y. Zhang, "Parcast: Soft video delivery in mimo-ofdm wlans," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 233–244.

[5] S. Aditya and S. Katti, "Flexcast: Graceful wireless video streaming," in *Proceedings of the 17th annual international conference on Mobile computing and networking*. ACM, 2011, pp. 277–288.

[6] J. Zhao, R. Xiong, and J. Xu, "Omnicast: Wireless pseudo-analog transmission for omnidirectional video," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 58–70, 2019.

[7] J. Shen, L. Yu, L. Li, and H. Li, "Foveation-based wireless soft image delivery," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2788–2800, 2018.

[8] 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 *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016, pp. 1161–1170.

[9] A. T. Nasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using scalable video coding," in *Proceedings of the 25th ACM International Conference on Multimedia*. ACM, 2017, pp. 1689–1697.

[10] S. Rossi and L. Toni, "Navigation-aware adaptive streaming strategies for omnidirectional video," in *IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2017, pp. 1–6.

[11] M. Xiao, C. Zhou, Y. Liu, and S. Chen, "Optile: Toward optimal tiling in 360-degree video streaming," in *Proceedings of the 25th ACM International Conference on Multimedia*. ACM, 2017, pp. 708–716.

[12] D. V. Nguyen, H. T. Tran, A. T. Pham, and T. C. Thang, "A new adaptation approach for viewport-adaptive 360-degree video streaming," in *IEEE International Symposium on Multimedia (ISM)*. IEEE, 2017, pp. 38–44.

[13] P. Lungaro, R. Sjöberg, A. J. F. Valero, A. Mittal, and K. Tollmar, "Gaze-aware streaming solutions for the next generation of mobile vr experiences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1535–1544, 2018.

[14] 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 *Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video*. ACM, 2017, pp. 67–72.

[15] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li, "Salient object detection: A survey," *arXiv preprint arXiv:1411.5878*, 2014.

[16] Y. Bao, T. Zhang, A. Pande, H. Wu, and X. Liu, "Motion-prediction-based multicast for 360-degree video transmissions," in *14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2017, pp. 1–9.

[17] X. Meng, "Foveated rendering techniques in modern computer graphics," Ph.D. dissertation, 2018.

[18] J. Ryoo, K. Yun, D. Samaras, S. R. Das, and G. Zelinsky, "Design and evaluation of a foveated video streaming service for commodity client devices," in *Proceedings of the 7th International Conference on Multimedia Systems*. ACM, 2016, p. 6.

[19] J.-S. Lee and T. Ebrahimi, "Perceptual video compression: A survey," *IEEE Journal of selected topics in signal processing*, vol. 6, no. 6, pp. 684–697, 2012.

[20] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Transactions on image processing*, vol. 10, no. 10, pp. 1397–1410, 2001.

[21] 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 *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, 2017, pp. 211–216.