Graceful Quality Improvement in Wireless 360-Degree Video Delivery

Takuya Fujihashi*, Makoto Kobayashi†, Keiichi Endo*, Shunsuke Saruwatari†,
Shinya Kobayashi*, and Takashi Watanabe†
*Graduate School of Science and Engineering, Ehime University, Ehime, Japan
†Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

***Abstract*—In 360-degree video streaming, a user may watch a part of video frames, namely, viewport, through an interactive display, e.g., head-mounted display, for the immersive experience. One of major issues for high-quality 360-degree video delivery is how to reduce perceptual redundancy in video frames according to user's viewing viewport. To this end, an existing scheme divides video frames into multiple tiles and adaptively encodes each tile based on user's viewing viewport using digital-based video compression. However, the video compression causes an issue in wireless channel called cliff effect, which causes sudden quality degradation in user's viewport and poor immersion in immersive applications. To reduce perceptual redundancy without cliff effect in wireless 360-degree video delivery, we propose a novel transmission scheme. Our scheme skips quantization and entropy coding, instead, directly transmits linear-transformed signals based on three-dimensional discrete cosine transform (3D-DCT) or the combination of one-dimensional DCT (1D-DCT) and spherical wavelet transform (SWT). By skipping both operations, the quality of viewport can be simply enhanced by its power allocation since it is directly applied at the pixel level instead of the bit level. It is demonstrated with 360-degree video sequences that the proposed scheme offers a great advantage over the conventional digital-based schemes.**

I. Introduction

360-degree video delivery is one of the attractive techniques to provide an immersive experience for users. Each user can watch the 360-degree video using an interactive display such as the head-mounted display (HMD). When the user watches the 360-degree video on the interactive display, a sender provides 360-degree video frames for a user and the user may play a part of the 360-degree video, which is referred to as *viewport*, through the user's interactive display. Here, 360-degree video frames are captured by an omnidirectional camera and mapped onto an equirectangular format. Since the equirectangular video frames can be regarded as panoramic video frames, the sender can take conventional video coding for 360-degree video contents.

For a good immersion, one of major issues in 360-degree video delivery is to achieve better video quality in user's viewport by reducing perceptual redundancy in 360-degree video frames. Since each user only watches viewport via the interactive display in each time instance, it causes a large video traffic if a sender transmits full panoramic video frames with an identical quantization parameter. To reduce perceptual redundancy, one of the simplest methods is viewport-only streaming [1]. In 360-degree video playback, the user may move a viewing viewport according to the user's head movement. The user requests the new viewport to the sender, and then the sender sends back the viewport. Since the sender transmits one viewport at each time instant, viewport-only streaming can mitigate video traffic. However, the user needs to receive a new viewport from the sender in every viewport switching, and thus it causes a long switching delay. Long switching delay, i.e., around 10 ms, may cause *simulator sickness* [2]. Due to a long transmission delay in the standard Internet, the viewport-only streaming schemes are difficult to satisfy switching delay requirement. To prevent *simulator sickness*, conventional schemes of 360-degree video delivery divide 360-degree video frames into multiple tiles and independently encode them with different quantization parameters. For example, [3] adaptively encode each tile based on user's viewport to achieve better video quality under a limited bandwidth.

In this study, we focus on future 360-degree video delivery, i.e., wireless 360-degree video streaming. Conventional studies on 360-degree video streaming consider that user's interactive display is connected to the sender via wire, and it makes it inherently difficult to realize outside immersive applications, e.g., entertainment and education. To overcome this limitation, our study exploits wireless links and a wireless-enabled interactive display for 360-degree video streaming.

If conventional schemes of 360-degree video delivery are simply extended for wireless links, the digital video compression and digital wireless transmission are carried out in sequence [4]–[6] for 360-degree video frames. For example, the video compression part uses H.264/Advanced Video Coding (AVC) [7] or H.265/High-Efficiency Video Coding (HEVC) [8] standard to generate a compressed bit stream using quantization and entropy coding. The wireless transmission part uses a channel coding and a digital modulation scheme to reliably transmit the encoded bit stream.

However, the conventional schemes have the following problems due to the wireless channel unreliability. First, the encoded bit stream is highly vulnerable to bit errors. Even when the video quality of tiles corresponding to user's viewing viewport is enhanced in encoding, the received video quality drops significantly when the channel signal-to-noise ratio (SNR) falls under a certain threshold. This phenomenon is referred to as cliff effect. Second, the video quality of user's viewing viewport is constant even when the wireless channel quality is improved. This is because quantization is a lossy process, whose distortion cannot be recovered at the receiver.

Soft video delivery [9], [10] has been proposed to prevent cliff effect and to improve video quality in proportion to the improvement of the wireless channel quality. In soft video delivery, a sender first transforms pixel values of video frames into frequency components with power allocation, and then directly maps the frequency components to transmission symbols, i.e., near-analog modulation, to ensure the received video quality is proportional to the wireless channel quality. Since the soft video delivery skips nonlinear quantization operations and the power allocation is directly applied for each pixel, the perceptual redundancy can be simply minimized by optimizing the power allocation according to user's viewport. To the best of our knowledge, there are no existing studies on soft video delivery for this purpose.

As mentioned above, conventional schemes of wireless 360-degree video delivery have three challenging issues in wireless links: 1) cliff effect, 2) constant quality, and 3) user's viewport-based quality adaptation. In this paper, we propose a new transmission scheme to overcome these issues, motivated by the studies on soft video delivery. The contribution of this paper is two-fold. First, pixel-wise power allocation is applied for quality improvement in user's viewing viewport. Second, three-dimensional discrete cosine transform (3D-DCT) and the combination of one-dimensional DCT (1D-DCT) and spherical wavelet transform (SWT) [11] are used for transformation from pixel to frequency domain to exploit spatial correlation of the 360-degree video frames for quality improvement.

Evaluations show that the proposed scheme gracefully improves video quality with the improvement of wireless channel quality. Even in low channel quality regimes, the proposed scheme prevents cliff effect and keeps better video quality. In addition, the power allocation in the proposed scheme realizes fine-grained quality adaptation according to users' viewport preference.

II. Related Works

A. Viewport-only Streaming

The simplest idea to realize 360-degree video streaming through an interactive display is to send only the part of 360-degree video frames that corresponds to user's viewing viewport [1]. The viewport-only streaming can reduce video traffic while it does not enable smooth viewport switching. If the user moves his/her head for viewport switching, user's interactive display has to immediately show the corresponding viewport. However, since the interactive display in the viewport-only streaming scheme does not have the outside of the viewing viewport. It means the receiver side needs to send a viewport request to the sender, and it will cause playback stalls until the receiver completes the reception of the viewport. As seen in other interactive multimedia systems, this solution cannot meet the switching delay requirement in the standard Internet.

In contrast to the viewport-only streaming, the proposed scheme sends full 360-degree video frames to user's interactive display. Since the display has full 360-degree video frames, the display only extracts a required viewport from the received frames, and thus the proposed scheme can satisfy switching delay requirement. In addition, the proposed scheme integrates orthogonal transform techniques, analog modulation, and power allocation to facilitate wireless 360-degree video streaming. This integration can bring graceful quality improvement in the viewing viewport with the improvement of wireless channel quality.

B. Panoramic Transmission

Another solution of 360-degree video delivery is to send full 360-degree video frames to a user, namely, panoramic transmission. However, a simple panoramic transmission, i.e., equal video quality across whole video frame, causes a large perceptual redundancy and low video quality. To improve the video quality of panoramic transmission, there are mainly two solutions have been proposed: tiling-based streaming and Quality Emphasized Region (QER)-based streaming.

In tiling-based streaming, the sender spatially cuts each 360-degree video frame into independent tiles. Each tile is encoded into multiple representations, i.e., multiple bit-rates. A user downloads adequate representation for each tile based on the user's viewing viewport. In [3], they sketched a tiling-based streaming system for 360-degree videos. The 360-degree video frames are mapped onto equirectangular video frames and cut into 8×8 tiles. In [12], they proposed hexaface sphere-based tiling of 360-degree video to enhance projection distortion compared to equirectangular projection. In addition, they integrate the proposed approach into MPEG-DASH Spatial Relationship Description (SRD) to realize adaptive 360-degree video streaming based on user's viewing viewport.

In QER-based streaming, each 360-degree video frame is divided into QER and the other parts. QER allocates high bit-rate while the other parts allocate lower bit-rate. Facebook implements QER-based 360-degree video streaming by using pyramid projection [13]. In [14], they proposed QER-based adaptive video streaming under the consideration of MPEG-DASH. Specifically, a sender encodes each 360-degree video frame into multiple representations with different QERs. Here, a user downloads the corresponding representation, whose QER is the closest to user's viewing viewport, to achieve better video quality in the viewing viewport.

The proposed scheme is one of QER-based streaming schemes for wireless links. To implement the concept of QER into the proposed scheme, a sender adaptively assigns transmission power to each pixel based on user's viewport preference. In contrast to conventional QER-based streaming schemes, the proposed scheme skips quantization and entropy coding in video coding, instead, uses analog modulation to achieve viewport-aware quality improvement without cliff effect.

C. Soft Video Delivery

Soft video delivery schemes have been recently proposed for wireless video streaming in [9], [15]–[17]. SoftCast [9] is a pioneer work of soft video delivery. SoftCast skips quantization and entropy coding, and uses analog modulation, which maps DCT coefficients directly to transmission signals, to ensure that the received video quality is proportional to wireless channel quality. Since SoftCast was designed for an additive white Gaussian noise (AWGN) channel, another study [15] extended such soft video delivery for fading channels to discuss an impact of soft video delivery in worse case. ParCast [16] and AirScale [17] extended soft video delivery for multi-carrier and multi-antenna systems, respectively.

Recent studies consider the reduction on perceptual redundancy in soft video delivery [18]–[20]. In [18], they consider foveation points in playback video frames. Based on the foveation points, they used two-dimensional discrete wavelet transform (2D-DWT) as an orthogonal transform technique and assigned transmission power to each wavelet coefficient for reducing perceptual redundancy. In addition, [19] considered to reduce perceptual redundancy in free viewpoint video delivery. Specifically, they adaptively assigned transmission power to multi-view plus depth video frames in left and right cameras to achieve the best video quality at user's requested virtual viewpoint.

Our study aims to reduce spatial and perceptual redundancy in 360-degree soft video delivery in order to realize better and graceful video quality in user's viewing viewport. To facilitate such 360-degree video streaming, our scheme makes the following major contributions.

- We use cosine-based orthogonal transform techniques, e.g., 3D-DCT and the combination of 1D-DCT and SWT, for 360-degree video frames to reduce spatial redundancy of spherical images.
- We adaptively assign transmission power to each pixel of 360-degree video frames based on user's viewport preference before the transform techniques to reduce perceptual redundancy.
- We investigate an effect of orthogonal transform techniques in 360-degree soft video delivery. From the evaluations, it is demonstrated that cosine-based transform techniques bring better video quality compared to wavelet-based transform techniques in low channel quality regimes.

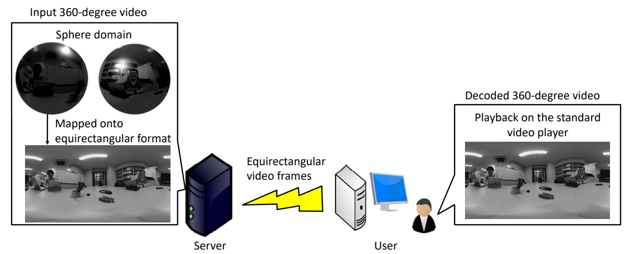
In addition, we consider an AWGN channel as wireless channel environment to demonstrate the baseline performance of the proposed scheme. Detailed evaluations with fading channels will be left as future work.

III. Proposed 360-Degree Soft Video Delivery

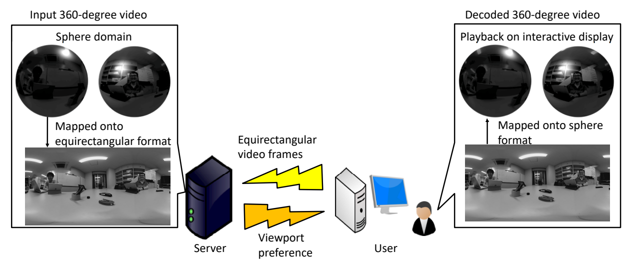
The objectives of our study are 1) to prevent cliff effect, 2) to realize graceful quality improvement with the improvement of wireless channel quality, and 3) to achieve quality adaptation based on user's viewing viewport when the user is watching 360-degree video on the interactive display.

Fig. 1 shows system models under consideration. 360-degree video frames are captured by an omnidirectional camera and mapped onto a sphere. We can consider three models of the end-to-end wireless 360-degree video streaming. The first model is plane-to-plane streaming. In this streaming, the 360-degree video frames are mapped onto the equirectangular format. Here, the resolution of the equirectangular video frames is $2N \times N$ pixels. A sender transmits the equirectangular video frames to the receiver, and the receiver watches the video frames on the standard video player. The second model is plane-to-sphere streaming. A sender also transmits the equirectangular video frames to the receiver. The received video frames are mapped onto a sphere using the geodesic sphere construction [11] and the receiver extracts a viewport from the sphere video frames. When the user moves a viewing viewport, the user transmits the viewport information to the sender, and then the sender sends quality-adapted video frames based on the viewport information. The third model is sphere-to-sphere streaming. In sphere-to-sphere streaming, a sender transmits sphere video frames to the receiver. The receiver watches the sphere video frames on the interactive display.

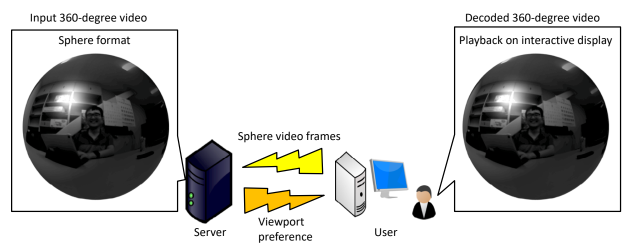
Fig. 2 shows an overview of our proposed scheme. The proposed scheme will switch an orthogonal transformation based on the input format of the 360-degree video. When



(a) Plane-to-Plane Streaming: send equirectangular video frames and watch the video frames on the standard video player.



(b) Plane-to-Sphere Streaming: send equirectangular video frames and map onto sphere to watch the video frames on the interactive display.



(c) Sphere-to-Sphere Streaming: send sphere video frames and watch the video frames on the interactive display.

Fig. 1. Three models of wireless 360-degree video streaming. (a) plain-to-plain streaming, (b) plain-to-sphere streaming, and (c) sphere-to-sphere streaming.

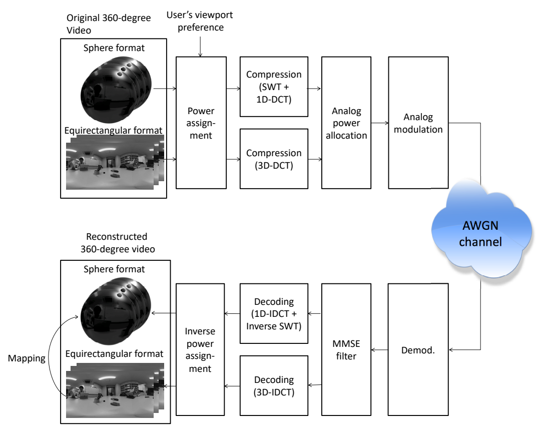


Fig. 2. Overview of the proposed encoding/decoding operations.

the input format is a sphere, the proposed scheme uses a combination of SWT and 1D-DCT for compression. On the other hand, the proposed scheme takes 3D-DCT operation for 360-degree video frames when the input format is equirectangular. After the compression, the transformed coefficients are then scaled and analog-modulated for wireless transmissions. Next, the encoder sends the analog modulated symbols to the user over a wireless channel with AWGN. At the user side, the decoder uses minimum mean-square error (MMSE) filter to obtain the transmitted coefficients. The decoder then performs the inverse orthogonal transformation to reconstruct the transmitted 360-degree video frames. Finally, the decoder can playback the received video frames on the standard video player and interactive display.

A. Encoding

1) Compression: A sender transforms pixel values of 360-degree video frames in one Group of Pictures (GOP) into frequency domain coefficients. GOP is a unit representing a set of video frames to be encoded at one time. In our study, we consider one GOP represents eight 360-degree video frames. In plane-to-plane and plane-to-sphere streaming, the sender uses 3D-DCT operation for 360-degree video frames because of its strong energy compaction property and simplicity of computing. On the other hand, SWT can be used in sphere-to-sphere streaming because SWT brings better compression performance in the sphere domain. However, since SWT is designed for each sphere video frame, it does not utilize time-domain correlation for compression. To utilize such correlation for quality improvement, the proposed scheme combines SWT and 1D-DCT for compression. More specifically, the sender first obtains a vector of spherical wavelet coefficients $\boldsymbol{w}_i$ by using SWT for each sphere video frame i and then constructs a matrix of spherical wavelet coefficients in one GOP W by horizontally aligned $\boldsymbol{w}_i$. Finally, 1D-DCT is used for W in each row to exploit time-domain correlation.

2) Power Allocation: Let $\check{s}_i \in \mathbb{R}$ denote a power allocated symbol of i-th analog-modulated symbol. Each coefficient in frequency domain is scaled by a scale factor g_i for noise reduction:

$$\check{s}_i = g_i \cdot s_i. \tag{1}$$

Here, $s_i \in \mathbb{R}$ is the i-th DCT coefficient. The near-optimal solution g_i to minimize the mean-square error (MSE) is obtained as follows [9]:

$$g_i = \lambda_i^{-1/4} \sqrt{\frac{P}{\sum_{k=1}^{N_c} \sqrt{\lambda_k}}}, \tag{2}$$

where P denotes a total transmission power budget for wireless 360-degree video streaming, λ_i is the power of i-th coefficient, and N_c is the number of coefficients in one GOP Finally, a transmission symbol $x_i \in \mathbb{C}$ is formed by superposing two power allocated symbols of $\check{s}_i$ and $\check{s}_j$ as follows:

$$x_i = \check{s}_i + \jmath\, \check{s}_j, \tag{3}$$

where $\jmath = \sqrt{-1}$ denotes the imaginary unit.

B. Decoding

Over the wireless links, the receiver obtains the received symbol, which is modeled as follows:

$$y_i = x_i + n_i, \tag{4}$$

where $y_i \in \mathbb{C}$ is the i-th received symbol and $n_i \in \mathbb{C}$ is an effective AWGN with a variance of σ^2. The received coefficients are extracted from each component of the received symbols via an MMSE filter [9]:

$$\hat{s}_i = \frac{g_i \lambda_i^2}{g_i^2 \lambda_i^2 + \sigma^2} \cdot \Re(y_i),\ \hat{s}_j = \frac{g_j \lambda_j^2}{g_j^2 \lambda_j^2 + \sigma^2} \cdot \Im(y_i). \tag{5}$$

The filtered signals are then reconstructed to pixel values by using inverse operation of compression at the sender.

C. Viewport-based Quality Adaptation

In 360-degree video playback, a user may watch a part of the 360-degree video frames, i.e., viewport, through an interactive display and moves the viewing viewport during video playback. It means a sender should emphasize quality of pixels corresponding to the viewing viewport to realize a good immersion.

For such quality adaptation, the proposed scheme can control transmission power assignment for pixels of viewing viewport and the other parts. For simplicity, the proposed scheme assigns $p_{\text{viewport}}\%$ of transmission power to the viewing viewport and $(1 - p_{\text{viewport}})\%$ of transmission power to the other parts. If the display probability of each pixel in 360-degree video frames was obtained from past users' viewing log, pixel-wise power assignment can be applied for fine-grained quality adaptation. When the display probability of pixel $f_{i,j}$ is $p_{i,j}$, the pixel-wise power assignment can be achieved as follows:

$$\hat{f}_{i,j} = f_{i,j} \cdot p_{i,j}, \tag{6}$$

where $\hat{f}_{i,j}$ is the corresponding pixel value after power assignment.

D. Analog Compression for Limited Bandwidth

The previous designs assume that the sender has enough bandwidth to transmit all the coefficients in the spectral domain over the wireless medium. If the available bandwidth and/or time resources are restricted for wireless channel use, it has to selectively transmit the coefficients to fit the available bandwidth. For such cases, our scheme sorts the coefficients in descending order of the power and picks higher-power coefficients to fill the bandwidth. When the sender discards a coefficient, the receiver regards the discarded coefficient as zero. As a result, a sort of data compression can be accomplished even for analog-based video delivery. Even when some coefficients are discarded to reduce the amount of data, the receiver can still achieve a graceful video quality until reaching the distortion limit due to the compression.

IV. Evaluation

A. Settings

Performance Metric: We evaluate the performance in terms of the peak SNR (PSNR) and SSIM [21]. 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 pixel luminance (typically eight bits), and $\varepsilon_{\mathrm{MSE}}$ is the MSE between all pixels of the decoded and the original video. SSIM can predict the perceived quality of video streaming. Larger values of SSIM close to 1 indicates higher perceptual similarity between original and decoded images. We obtain the average PSNR and SSIM across 80 video frames in the test video sequence.

Test Video: We use one test 360-degree video sequence, namely, *meeting*, at a resolution of 1920×960 pixels with 30 fps for comparison. The test sequence is captured by the Ricoh Theta camera. In this test sequence, two people are face-to-face talking and working on the desk and the camera is located on the middle point of two people. In future work, we will use multiple 360-degree test sequences with high and low motions to evaluate an impact of content features.

Wireless Channel Environment: The effective noise in AWGN channel n_i follows white Gaussian distribution with a variance of σ^2, i.e., $n_i \sim \mathcal{CN}(0, \sigma^2)$. In addition, we consider the channel symbol rate of the wireless channel is 0.6 Msymbols/s.

B. Video Quality in Plane-to-Plane Streaming

We first show the performance of the proposed scheme in plane-to-plane streaming. We consider two digital-based schemes and two analog-based schemes as reference schemes. The digital-based schemes use BPSK and QPSK as a modulation format, respectively. The analog-based schemes compress 360-degree video frames using two-dimensional DCT (2D-DCT) and two-dimensional discrete wavelet transform (2D-DWT), which is based on [18], with subdivision level of 3, respectively.

Fig. 3 shows the received video quality of five reference schemes in plane-to-plane streaming as a function of wireless channel SNRs. From the result, we can observe the video quality of the proposed scheme is proportional to wireless channel quality. On the other hand, the video quality in digital-based schemes is a step function of channel SNRs. For example, the proposed scheme achieves the PSNR improvement by 13.5 and 18.7 dB over BPSK and QPSK schemes across channel SNRs of 0 to 25 dB.

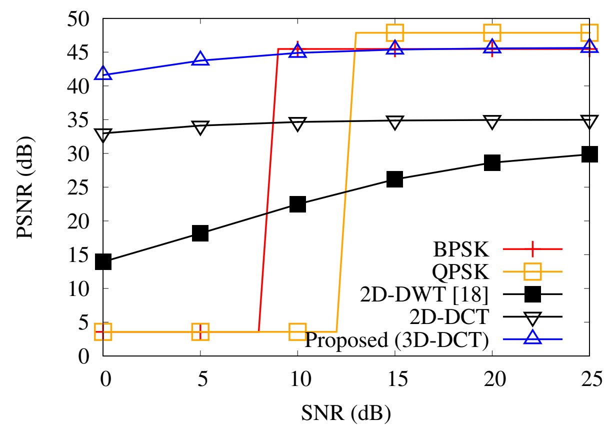


Fig. 3. Video Quality vs. wireless channel SNRs in plane-to-plane streaming at a channel symbol rate of 0.6 Msymbols/s.

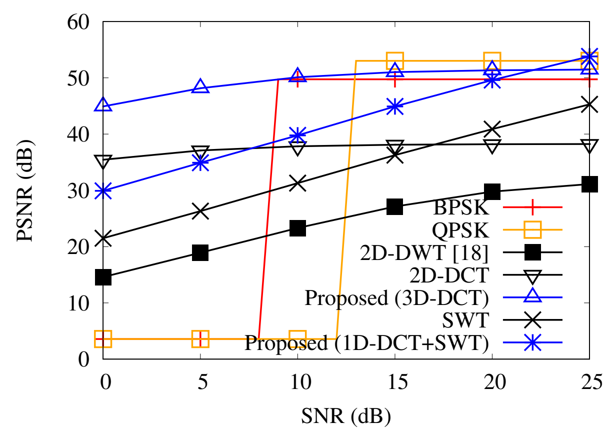


Fig. 4. Video Quality vs. wireless channel SNRs in plane-to-sphere and sphere-to-sphere streaming at a channel symbol rate of 0.6 Msymbols/s.

In view of the analog-based scheme, the proposed scheme yields better video quality compared to both analog schemes with 2D-DCT and 2D-DWT. This is because 3D-DCT can compact the energy of 360-degree video frames by utilizing time-domain correlations. In addition, we can see that cosine-based schemes achieve better performance at low channel SNR regimes compared to the wavelet-based scheme.

C. Video Quality in Plane-to-Sphere and Sphere-to-Sphere Streaming

When a user watches the 360-degree video using an interactive display, the 360-degree video frames are mapped onto the sphere before playback. In this section, we evaluate 1) video quality of reference schemes in the sphere and 2) an effect of

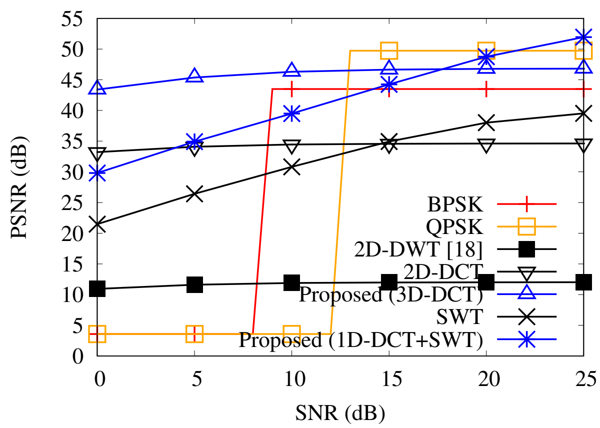


Fig. 5. Video Quality vs. wireless channel SNRs in plane-to-sphere and sphere-to-sphere streaming at a channel symbol rate of 0.3 Msymbols/s.

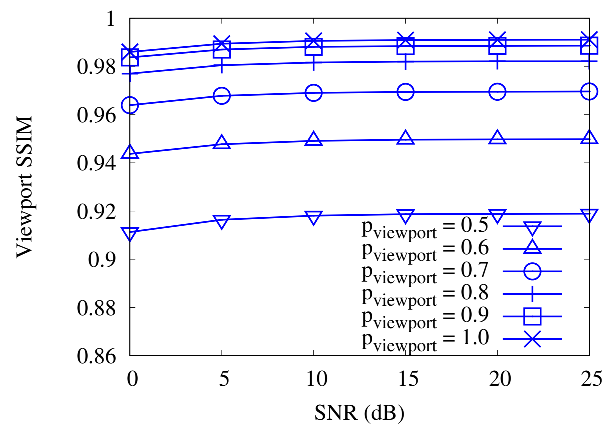


Fig. 7. Viewport Quality vs. wireless channel SNRs in equirectangular format at a channel symbol rate of 0.6 Msymbols/s.

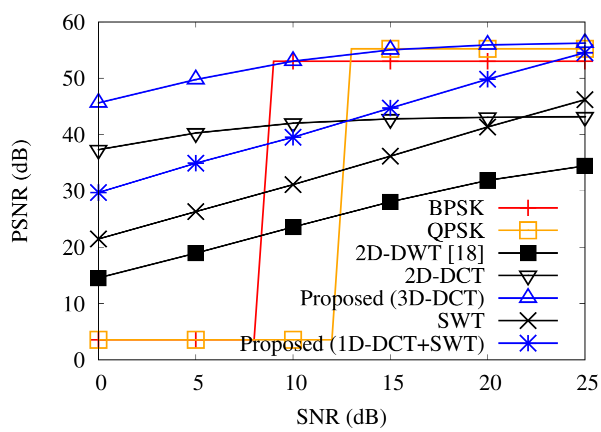


Fig. 6. Video Quality vs. wireless channel SNRs in plane-to-sphere and sphere-to-sphere streaming at a channel symbol rate of 1.2 Msymbols/s.

sphere-to-sphere streaming in comparison with plane-to-sphere streaming.

For comparison in plane-to-sphere streaming, two digital-based schemes, i.e., BPSK and QPSK, and two analog-based schemes with 2D-DCT and 2D-DWT are transmitted 360-degree video frames in the equirectangular format to a receiver, and then the received video frames are mapped onto the sphere. For sphere-to-sphere streaming, one analog-based scheme uses SWT for each sphere video frame and transmits the wavelet coefficients to the receiver.

To construct sphere video frames from the equirectangular video frames for comparison, we use the geodesic sphere construction starting with icosahedron in 7 subdivision levels. Here, the number of surfaces in each sphere video frame is 163,842. We obtain the average PSNR across all the surfaces in sphere video frames.

Fig. 4 shows the video quality of reference schemes in plane-to-sphere and sphere-to-sphere streaming as a function of channel SNRs. We can see the following key points:

- Even in the sphere video frames, the proposed schemes gracefully improve video quality with the improvement of wireless channel SNRs.
- Our scheme of SWT and 1D-DCT compression achieves better video quality compared to the analog-based scheme of SWT compression by utilizing time-domain correlations for compression.
- The proposed scheme of 3D-DCT compression achieves better PSNR performance compared to that of SWT and 1D-DCT compression in low channel SNR regimes while the proposed scheme of SWT and 1D-DCT compression, i.e., sphere-to-sphere streaming, yields higher video quality in high channel SNR regimes.

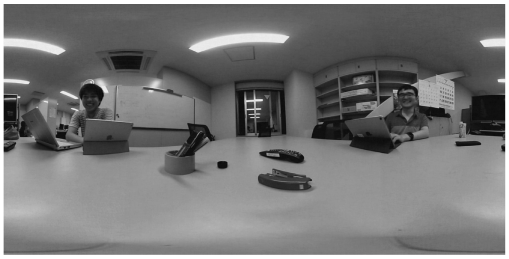
D. Discussion on Bandwidth Limitation

Previous evaluations assumed that the proposed scheme transmits linear-transformed signals to a receiver according to the channel symbol rate of 0.6 Msymbols/s. We can adaptively increase and reduce the number of transmitting coefficients if the channel symbol rate is high and low. This section discusses the video quality in narrow-band and broadband environments.

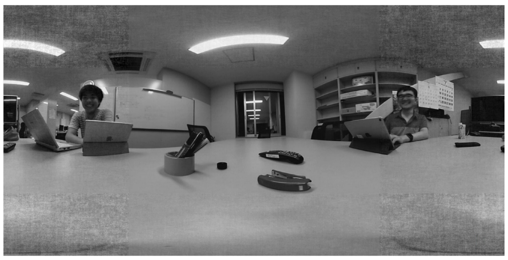
Figs. 5 and 6 show the video quality of reference schemes in plane-to-sphere and sphere-to-sphere streaming as a function of channel SNRs at channel symbol rates of 0.3 Msymbols/s and 1.2 Msymbols/s, respectively. The proposed scheme with 3D-DCT compression still achieves graceful quality improvement irrespective of channel symbol rates. On the other hand, the video quality of 2D-DWT scheme is even low at a channel symbol rate of 0.3 Msymbols/s. This is because a sender does not transmit most of the low frequency components due to the narrow bandwidth.

E. Discussion on Quality Adaptation

Finally, we discuss an effect of our proposed power assignment on quality adaptation based on the user's watching viewport. We consider the resolution of the viewport is 960 × 480 pixels and the viewport is set in the center of the equirectangular video frames. Here, we use 3D-DCT for



(a) p_{viewport} is 0.5



(b) p_{viewport} is 0.9

Fig. 8. Snapshot of *meeting* (frame #1) in different power assignment ratios at an SNR of 0 dB and channel symbol rate of 0.6 Msymbols/s.

compression and the power assignment ratio p_{viewport} between 0.5 to 1 for comparison.

Fig. 7 shows SSIM performance of the viewport with different power assignment ratios as a function of channel SNRs at a channel symbol rate of 0.6 Msymbols/s. Based on the power assignment ratio, the proposed scheme adaptively enhances viewport quality. For example, the proposed scheme with power assignment ratio of 0.9 achieves the SSIM improvement by 0.07 over the proposed scheme with power assignment ratio of 0.5 across channel SNRs of 0 to 25 dB.

On the other hand, the quality of the other parts is degraded at a large power assignment ratio. Fig. 8 compares the visual quality of the proposed scheme with power assignment ratio of 0.9 and 0.5, respectively, at a channel SNR of 0 dB and the channel symbol rate of 0.6 Msymbols/s. From the snapshots, we can clearly see that the proposed scheme with power assignment ratio of 0.9 achieves a clean viewport while the quality of the other parts is low.

V. Conclusion

In this paper, we propose a new 360-degree video transmission scheme for wireless VR/AR clients. The proposed scheme directly sends frequency domain coefficients to ensure the received video quality is proportional to wireless channel quality. In addition, the quality adaptation considering user's viewing viewport can be solved by a simple power assignment issue. Evaluations show that the proposed scheme gracefully improves video quality according to wireless channel quality and achieves better video quality even in low channel SNR regimes.

Acknowledgment

This work was partly supported by JSPS KAKENHI Grant Numbers JP17K12672 and JP17J02859.

References

[1] F. Qian, B. Han, L. Ji, and V. Gopalakrishnan, "Optimizing 360 video delivery over cellular networks," in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, 2016, pp. 1–6.

[2] J. Moss and E. Muth, "Characteristics of head-mounted displays and their effects on simulator sickness," *The Journal of the Human Factors and Ergonomics Society*, vol. 53, no. 3, pp. 308–319, 2011.

[3] D. Ochi, Y. Kunita, A. Kameda, A. Kojima, and S. Iwaki, "Live streaming system for omnidirectional video," in *IEEE VR*, 2015, pp. 349–350.

[4] T. Stockhammer, H. Jenkac, and G. Kuhn, "Streaming video over variable bit-rate wireless channels," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 268–277, 2004.

[5] Z. Guo, Y. Wang, E. Erkip, and S. Panwar, "Wireless video multicast with cooperative and incremental transmission of parity packets," *IEEE Transactions on Multimedia*, vol. 17, no. 8, pp. 1135–1346, 2015.

[6] S. Almowuena, M. M. Rahman, C. H. Hsu, A. A. Hassan, and M. Hafeeda, "Energy-aware and bandwidth-efficient hybrid video streaming over mobile networks," *IEEE Transactions on Multimedia*, vol. 18, no. 1, pp. 102–115, 2016.

[7] 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.

[8] 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.

[9] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *ACM Annual International Conference on Mobile Computing and Networking*, 2011, pp. 289–300.

[10] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "High-quality soft video delivery with gmrf-based overhead reduction," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 473–483, 2018.

[11] P. Schrodder and W. Sweldens, "Spherical Wavelets: efficiently representing functions on the sphere," in *ACM SIGGRAPH*, 1995, pp. 161–172.

[12] M. Hosseini and V. Swaminathan, "Adaptive 360 vr video streaming: Divide and conquer!" in *IEEE International Symposium on Multimedia*, 2016, pp. 107–110.

[13] E. Kuzyakov and D. Pio, "Next-generation video encoding techniques for 360 video and vr," 2016. [Online]. Available: https://code.facebook.com/posts/1126354007399553/

[14] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *IEEE ICC*, 2017, pp. 1–5.

[15] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust linear video transmission over rayleigh fading channel," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2790–2801, 2014.

[16] X. L. Liu, W. Hu, C. Luo, Q. Pu, F. Wu, and Y. Zhang, "ParCast+: Parallel video unicast in MIMO-OFDM WLANs," *IEEE Transactions on Multimedia*, vol. 16, no. 7, pp. 2038–2051, 2014.

[17] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Scalable video multicast for mu-mimo systems with antenna heterogeneity," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 992–1003, 2016.

[18] J. Shen, L. Yu, L. Li, and H. Li, "Foveation based wireless soft image delivery," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–13, 2018.

[19] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Soft video delivery for free viewpoint video," in *IEEE ICC*, 2017, pp. 1–7.

[20] J. Zhao, R. Xiong, C. Luo, F. Wu, and W. Gao, "Wireless image and video soft transmission via perception-inspired power distortion optimization," in *IEEE Visual Communications and Image Processing*, 2017, pp. 1–4.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Imge quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.