Soft Video Delivery for Free Viewpoint Video

Fujihashi, T.; Koike-Akino, T.; Watanabe, T.; Orlik, P.V.

TR2017-041 May 2017

Abstract

Wireless video streaming for multi-view plus depth (MVD) frames enables free viewpoint video on wireless devices, where a viewer can freely synthesize any preferred virtual viewpoint from the received MVD frames. To transmit MVD frames over wireless links, existing schemes use digital-based compression to achieve better compression efficiency. However, the digital-based schemes have an issue called cliff effect, where the video quality is a step function in terms of wireless channel quality. In addition, parameter optimization to allocate quantization levels and transmission power across MVD frames is cumbersome. To facilitate the MVD video streaming, we propose an analog-based transmission scheme. Our scheme directly transmits linear-transformed signals based on five-dimensional discrete cosine transform (5D-DCT), without relying on quantization and entropy coding. The proposed scheme achieves graceful video quality with the improvement of wireless channel quality. In addition, the parameter optimization to achieve highest video quality can be simplified by only controlling transmission power assignment. It is demonstrated with test MVD video sequences that the analog-based scheme offers a great advantage over the conventional digital-based scheme. For instance, the video quality of one intermediate virtual viewpoint is approximately 2.1 dB higher than digital-based schemes at a wireless channel quality, i.e., signal-to-noise ratio (SNR), of 15 dB.

IEEE International Conference on Communications (ICC)

Soft Video Delivery for Free Viewpoint Video

Takuya Fujihashi†*, Toshiaki Koike-Akino†, Takashi Watanabe‡, and Philip V. Orlik†
†Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
*Graduate School of Science and Engineering, Ehime University, Ehime, Japan
‡Graduate School of Information and Science, Osaka University, Osaka, Japan

***Abstract*—Wireless video streaming for multi-view plus depth (MVD) frames enables free viewpoint video on wireless devices, where a viewer can freely synthesize any preferred virtual viewpoint from the received MVD frames. To transmit MVD frames over wireless links, existing schemes use digital-based compression to achieve better compression efficiency. However, the digital-based schemes have an issue called cliff effect, where the video quality is a step function in terms of wireless channel quality. In addition, parameter optimization to allocate quantization levels and transmission power across MVD frames is cumbersome. To facilitate the MVD video streaming, we propose an analog-based transmission scheme. Our scheme directly transmits linear-transformed signals based on five-dimensional discrete cosine transform (5D-DCT), without relying on quantization and entropy coding. The proposed scheme achieves graceful video quality with the improvement of wireless channel quality. In addition, the parameter optimization to achieve highest video quality can be simplified by only controlling transmission power assignment. It is demonstrated with test MVD video sequences that the analog-based scheme offers a great advantage over the conventional digital-based scheme. For instance, the video quality of one intermediate virtual viewpoint is approximately 2.1 dB higher than digital-based schemes at a wireless channel quality, i.e., signal-to-noise ratio (SNR), of 15 dB.**

I. Introduction

Free viewpoint video [1]–[3] is an emerging and attractive technique to observe a three-dimensional (3D) scene from freely switchable angles. For free viewpoint video, a large array of closely spaced cameras capture texture and depth frames of the same 3D scene. The sender encodes and transmits the texture and depth frames of two or more adjacent viewpoints, whose format is known as multi-view plus depth (MVD) [4], based on viewer's preferred viewpoint. The viewer synthesizes intermediate virtual viewpoint using depth image-based rendering (DIBR) [5] from the received MVD frames.

For conventional MVD video streaming over wireless links, the digital video compression and transmission parts operate separately. For example, the digital video compression may be based on MVC+D [6] or 3D-advanced video coding (AVC) [7] to generate a compressed bit stream using linear transform, quantization, and entropy coding. The compression rate of the bit stream is adaptively selected according to the wireless channel quality. The transmission part uses a channel coding and digital modulation scheme to reliably transmit the compressed bit stream over wireless channels.

However, the conventional scheme has the following problems due to the wireless channel unreliability. First, the encoded bit stream is highly vulnerable for bit errors. When the channel signal-to-noise ratio (SNR) falls under a certain threshold and bit errors occur in the bit stream during communications, the errors cause the collapse of texture and depth decoding. The decoding failure disables rendering operation. As a result, the video quality of virtual viewpoints degrades significantly. This phenomenon is called cliff effect [8]. Second, the video quality does not improve even when the wireless channel quality is improved. Finally, quantization is a lossy process and its distortion cannot be recovered at the receiver.

T. Fujihashi conducted this research while he was an intern at MERL.

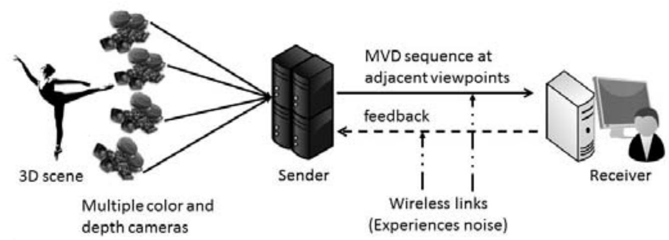


Fig. 1. Wireless MVD video streaming systems for free viewpoint rendering.

In addition, the digital-based wireless MVD scheme needs to solve a complicated parameter optimization to achieve the best quality for virtual viewpoint. The video quality of a certain virtual viewpoint highly depends on bit allocation and transmission power assignment across all the multi-view texture and depth frames. The bit allocation issue is referred to as *view synthesis optimization* [9], [10], which is often cumbersome to derive the solution because it is a combinatorial problem with nonlinear quantization.

As mentioned above, digital-based wireless MVD transmissions have three challenging issues: 1) cliff effect, 2) constant quality, and 3) complicated bit and power assignments. There are no existing studies resolving all of these issues in MVD video streaming to the best of our knowledge. In this paper, we propose a new wireless MVD transmission scheme to overcome these issues, motivated by the studies on soft video delivery [11]–[20]. The key idea of our scheme is skipping quantization and entropy coding at the encoder. Specifically, the proposed scheme jointly transforms texture and depth frames using 5D-discrete cosine transform (DCT), whose output are then scaled and directly mapped to transmission signals. The advantage of this modification lies in a fact that the pixel distortion due to communication noise is proportional to the magnitude of the noise, resulting into a graceful video

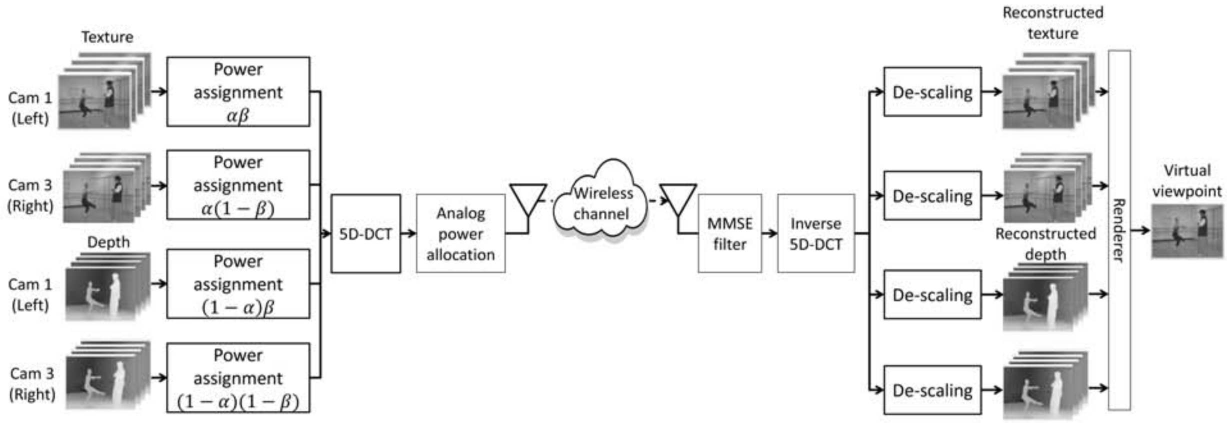


Fig. 2. Overview of proposed soft video delivery scheme for wireless MVD video streaming.

quality according to the wireless channel quality, without any cliff effect. In addition, our scheme simplifies the optimization problems by reformulating into a simple power assignment problem because bit allocation for quantization is not required in our scheme. We assign transmission power to each texture and depth frame before the transformation based on the viewer's preferred viewpoint.

Evaluations using several test MVD sequences show that our proposed scheme achieves graceful video quality with the improvement of wireless channel quality. In addition, we discuss how to optimize the power assignment for texture and depth frames according to the viewer's preferred viewpoint.

Related Works and Our Contributions: Soft video delivery schemes have been recently proposed for single-view video in [11]–[20]. For example, SoftCast [11] 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. ParCast [12] and AirScale [13] extended soft video delivery for multi-carrier and multi-antenna systems, respectively. Although graceful video delivery for stereo videos (i.e., two-view video) was discussed in [14], the paper does not account for rendering operation and thus how to achieve graceful performance in virtual viewpoints was beyond the scope of the paper.

Our study further extends the existing soft video delivery to wireless MVD video streaming. To achieve soft video delivery in free viewpoint video with high video quality, our scheme has the following major contributions.

- We use 5D-DCT for MVD video frames to exploit inter-view and texture-depth correlations for performance improvement.
- We show how to optimize the power assignment for texture and depth frames to achieve the highest video quality in each virtual viewpoint.

II. Soft Video Delivery for Free Viewpoint Video

The objectives of our study are 1) to prevent cliff effect in virtual viewpoints, 2) to gracefully improve video quality with the improvement of wireless channel quality, and 3) to simplify the issues of bit and power assignments to achieve the best video quality based on viewer's preferred viewpoint.

Fig. 1 shows the system model under consideration. The sender has multiple color (texture) and depth frames, which are video data captured by multiple cameras for the same 3D scene. The receiver sends feedback, which notifies his/her preferred virtual viewpoint, to the sender using a feedback channel with a certain interval. Based on the feedback, the sender transmits adjacent viewpoints of the requested viewpoints to the receiver. The requested viewpoint is then synthesized from the received texture and depth frames. The receiver can freely change the virtual viewpoint around the requested viewpoint for real-time rendering.

Fig. 2 shows the overview of our proposed transmission scheme. The encoder first assigns transmission power to each texture and depth frame, followed by 5D-DCT operation. The DCT coefficients are then scaled and analog-modulated for wireless transmissions. Next, the encoder sends the analog modulated symbols to the receiver over a wireless channel with additive white Gaussian noise (AWGN). At the receiver side, the decoder uses minimum mean-square error (MMSE) filter to obtain the transmitted DCT coefficients. The decoder then performs inverse 5D-DCT to reconstruct pixel values of MVD frames. Finally, the decoder synthesizes one intermediate virtual viewpoint from the MVD frames via DIBR [5].

A. Encoder

At the encoder, 5D-DCT is used for the whole texture and depth frames in one group of picture (GoP), which is a sequence of successive MVD video frames. The DCT coefficients are divided into several chunks. After power assignment for each chunk, the DCT coefficients are mapped to I (in-phase) and Q (quadrature-phase) components.

Let $x_{i,j}$ denote the ith analog-modulated symbol in jth chunk. Each analog-modulated symbol is scaled by g_j for noise reduction:

$$x_{i,j} = g_j \cdot s_{i,j}. \tag{1}$$

Here, $s_{i,j}$ is the ith DCT coefficient in jth chunk and g_j is the scale factor for the chunk. The near-optimal solution g_j to minimize the mean-square error (MSE) is obtained as follows [11]:

$$g_j = \lambda_j^{-1/4}\sqrt{\frac{P}{\sum_k \lambda_k}}, \tag{2}$$

where P denotes a total transmission power budget, and λ_j is the variance of jth chunk.

B. Decoder

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

$$y_{i,j} = x_{i,j} + n_{i,j}, \tag{3}$$

where $y_{i,j}$ is the ith received symbol in jth chunk and $n_{i,j}$ is an effective AWGN with a variance of σ^2. The DCT coefficients are extracted from I and Q components via an MMSE filter [11]:

$$\hat{s}_{i,j} = \frac{g_j \lambda_j^2}{g_j^2 \lambda_j^2 + \sigma^2} \cdot y_{i,j}. \tag{4}$$

The decoder then obtains corresponding video sequence by taking the inverse 5D-DCT for the filter output $\hat{s}_{i,j}$. Finally, the decoder synthesizes a preferred virtual viewpoint from the received texture and depth frames using DIBR [5].

C. Power Assignment

The video quality of virtual viewpoint is determined by the distortion of each texture and depth frame. In digital-based MVD schemes, the distortion depends on bit and power assignments for the frames. The parameter optimization is typically complicated to achieve the best quality at a target virtual viewpoint. In particular, finding the best quantization parameters across all texture and depth frames is not straightforward as the total number of possible combinations for quantization parameters can scale up to 52^4. This is because the interval of quantization parameters is $[0, 51]$ in the MVD encoder [21].

Our scheme simplifies the parameter optimization by removing quantization and entropy coding. Specifically, the distortion of texture and depth frames are reduced to a simple function of the assigned transmission power. Our scheme assigns transmission power for the frames under a certain power budget before 5D-DCT operation. To control the power assignment, we use two parameters: α and β. α is the power ratio for texture and depth frames. β is the ratio for frames at adjacent viewpoints. Specifically, adjacent viewpoints in texture and depth frames are scaled as shown in Fig. 2. We will discuss how to optimize α and β parameters in next section using a real MVD sequence. These two parameters, i.e., α and β, are sent from the transmitter to the receiver for de-scaling operations.

D. Analog Compression for Limited Bandwidth

The previous designs assume that the sender has enough bandwidth to transmit all the DCT coefficients 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 chunks in descending order of the variance and picks higher-variance chunks to fill the bandwidth. When the sender discards a chunk, the receiver regards all coefficients in the chunk as zeros. As a result, a sort of data compression can be accomplished even for analog-based video delivery. Even when some chunks 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.

III. Performance Evaluations

A. Simulation Settings

Performance Metric: We evaluate the performance in terms of the peak SNR (PSNR) defined as follows:

$$\mathsf{PSNR} = 10\log_{10}\frac{(2^L-1)^2}{\epsilon_{\mathsf{MSE}}}, \tag{5}$$

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. The original video is generated by DIBR given distortion-less adjacent MVD frames. We obtain the average PSNR across whole video frames in each video sequence.

Test Video: We use three standard reference MVD videos, namely, *balloons*, *kendo*, and *pantomime*, with 30 fps from Fujii Laboratory at Nagoya University [22]. In *balloons* and *kendo*, we use two cameras, i.e., camera 1 and 3, with a resolution of 1024×768 pixels for texture and depth frames. In *pantomime*, we also use two cameras, i.e., camera 37 and 39, with a resolution of 1280×960 pixels for both frames. The distance between two cameras is 10 cm long. We first focus on *balloons*, and then compare the video quality for the other video sequences.

Digital Video Encoder: We set the GoP size for all transmission schemes to 8 video frames. For digital schemes, we use 3D video AVC-based test model (3DV-ATM) v14.0 [21] video encoder/decoder to generate a bit stream from the test MVD video. We encode video frames of viewpoint 1 in one GoP into one I-frame and subsequent seven P-frames. Video frames of viewpoint 3 are encoded into eight P-frames by using both motion compensation and disparity compensation. We set the channel symbol rate of reference schemes to the half number of DCT coefficients transmitted in one second. Specifically, the channel symbol rate of each MVD video sequence is approximately 70.3 ($= 1280\times960\times30\times2\times2\times\frac{1}{2}$) Msymbols/s. We set quantization parameters to hold the number of digital-modulated symbols within the channel symbol rate. As an example, we consider the identical quantization parameter for texture and depth frames. When bit errors occur in the bit stream of a video frame, we regard the video frame and the

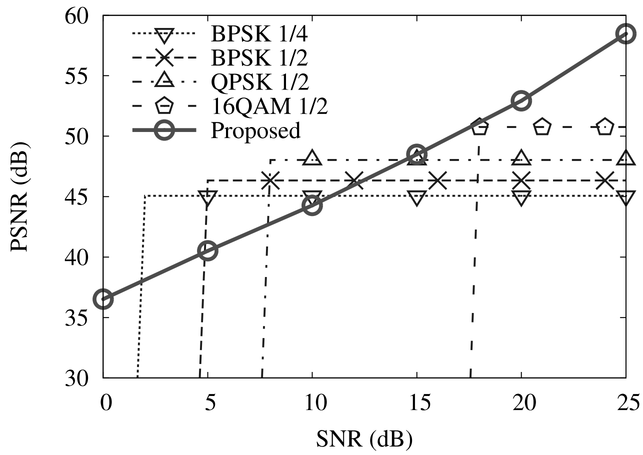


Fig. 3. PSNR vs. SNR at a virtual viewpoint of 2.

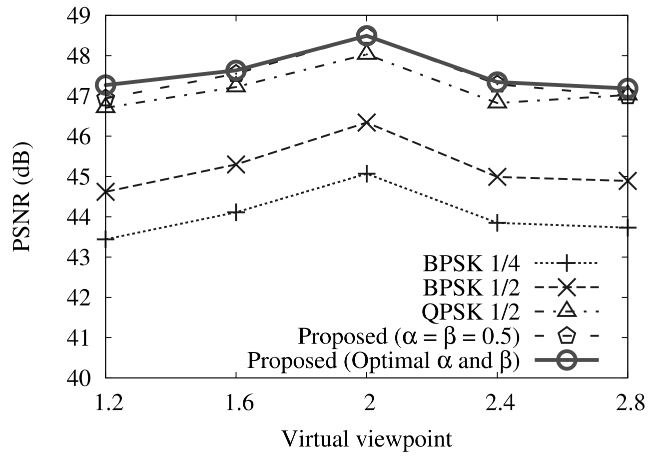


Fig. 4. PSNR vs. positions of virtual viewpoint at an SNR of 15 dB.

subsequent video frames as lost. In this case, we duplicate the nearest video frame, which is decoded successfully, to the lost video frame as an error concealment operation. For the analog scheme, we used MATLAB video encoder/decoder [23]. The chunk size is set to 64×32 pixels.

Rendering Software: To synthesize a virtual viewpoint from the received texture and depth frames, we used 3D HEVC test model (HTM) software renderer v13.0 [24]. The renderer requires two texture and depth frames as the input to produce video frames at a virtual viewpoint.

Wireless Settings: The received symbols are impaired by an AWGN channel. For digital-based schemes, we use a rate-$1/2$ and $1/4$ convolutional codes with a constraint length of 8. The digital modulation formats are either binary phase-shift keying (BPSK), quadrature PSK (QPSK), or 16-ary quadrature-amplitude modulation (16QAM).

B. Video Quality vs. Different Channel Quality

We first show the performance of the proposed analog transmission scheme in comparison to four digital-based schemes: BPSK with rate-$1/4$, BPSK with rate-$1/2$, QPSK with rate-$1/2$, and 16QAM with rate-$1/2$ convolutional codes. Fig. 3 shows the video quality at virtual viewpoint 2 (center between left viewpoint 1 and right viewpoint 3) as a function of the channel SNR. From this figure, we can observe two points:

- Our proposed scheme gracefully improves video quality without cliff effect.
- The video quality of digital-based scheme is a step function of SNR.

In digital-based schemes, bit errors induce the decoding failure for texture and depth frames of adjacent viewpoints in low SNRs. As a result, the renderer does not properly synthesize the virtual viewpoint. In addition, quantization operation in adjacent viewpoints causes the constant video quality in virtual viewpoint in high SNR regimes.

The proposed scheme prevents the cliff and constant quality by skipping quantization and entropy coding, and achieves better video quality than the digital-based schemes at some SNR regimes. For example, our scheme improves video quality by 2.1 dB compared to QPSK with rate-$1/2$ convolutional code at an SNR of 15 dB. Note that the video quality of the digital-based schemes can be potentially improved by jointly optimizing quantization parameters and power assignments for MVD frames. However, this optimization is much more complicated to solve, compared to the proposed method.

TABLE I
PARAMETERS FOR FITTING FUNCTION

	a	b	c	d	e	h
α	0.2	0.0004	0.0005	-0.82	-0.017	1.43
β	0.047	0	0.0031	-0.46	-0.0059	1.23

C. Impact of Different Virtual Viewpoint

We next evaluate the effect of the virtual viewpoint positions. For the evaluations, the renderer synthesizes the virtual viewpoint with five positions: 1.2, 1.6, 2.0, 2.4, and 2.8. Fig. 4 shows the video quality with different positions of virtual viewpoint at an SNR of 15 dB. Note that the result of 16QAM with $1/2$-rate convolutional code is not present because the collapse of rendering operation occurs frequently. This figure shows the following key observations:

- The proposed scheme yields the best quality regardless of the positions.
- Optimal transmission power assignment achieves better video quality especially at edge viewpoints.

The proposed scheme with optimal power assignment achieves at most 0.3 dB higher video quality compared to equal power assignment scheme.

D. Optimal Power Assignments

Previous evaluations revealed that the power assignments for 1) texture and depth 2) left and right frames are important to achieve the best video quality in each virtual viewpoint. This section investigates the optimal power assignments depending

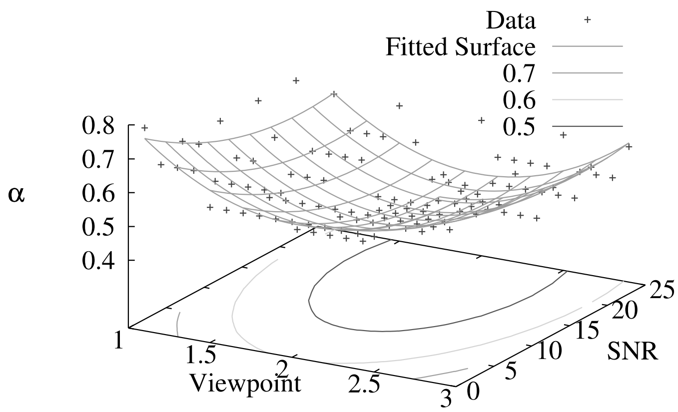


Fig. 5. Optimal points of α.

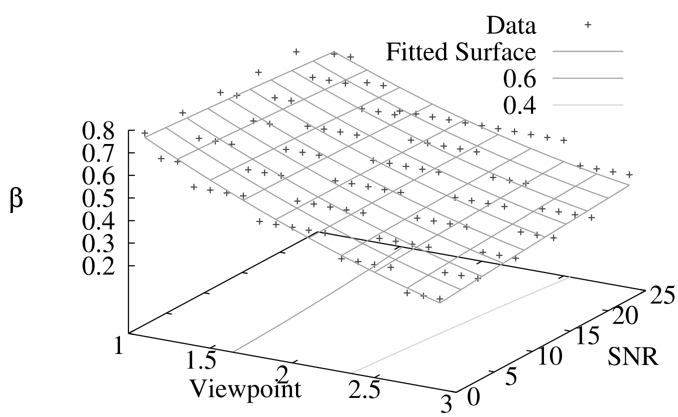


Fig. 6. Optimal points of β.

on the different viewpoints and SNRs. Figs. 5 and 6 show the optimal assignment values of α and β, respectively. The optimal values were obtained by sweeping α and β in a finite grid. In these figures, we also present a best-fit quadratic function of $f(p,q) = ap^2 + bq^2 + cpq + dp + eq + h$ by means of least-squares method, where p and q are viewpoint and SNR in dB, respectively. Table I lists the coefficients of the best-fit function for α and β. From the results, we can see the following aspects:

- Texture frames require higher transmission power (i.e., larger α) at low wireless channel quality.
- Depth frames shall use more transmission power as the wireless channel quality improves and the distance between adjacent viewpoints and virtual viewpoint increases.
- The optimal power ratio β for left and right frames is almost linear according to the distance between adjacent viewpoints and virtual viewpoint.

E. Discussion on Multi-Dimensional DCT

Our proposed scheme uses 5D-DCT to utilize correlations among pixels of texture and depth frames. This section

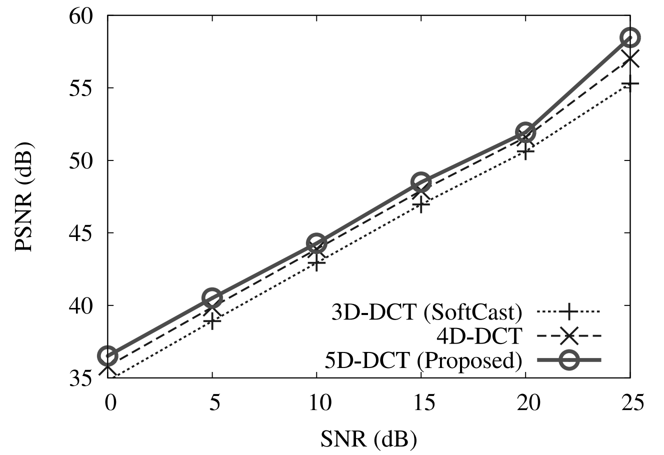


Fig. 7. PSNR vs. SNR with different DCT operations at a virtual viewpoint of 2.

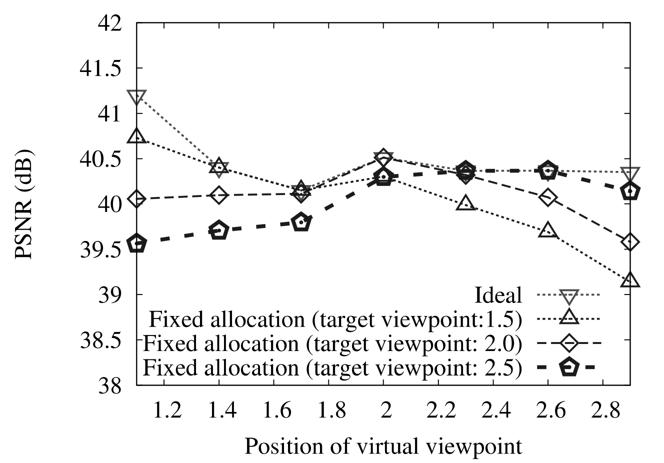


Fig. 8. PSNR vs. positions with different target viewpoints at an SNR of 5 dB.

compares the video quality of 5D-DCT with that of 4D-DCT and 3D-DCT schemes to evaluate the effect of different transforms. More specifically, 4D-DCT scheme independently carries out DCT operations for texture and depth frames, and 3D-DCT scheme (which is based on SoftCast [11]) takes DCT operations for each viewpoint in texture and depth frames.

Fig. 7 plots video quality at virtual viewpoint 2 as a function of wireless channel quality. It is observed that 5D-DCT scheme, i.e., our proposed scheme, outperforms the other schemes for the whole range of SNRs. This is because our scheme can efficiently exploit both the correlations between left and right viewpoints and between texture and depth frames. For example, 5D-DCT scheme achieves 1.6 dB higher video quality compared to the 3D-DCT scheme at an SNR of 5 dB.

F. Effect of Feedback Accuracy

Previous evaluations assume that a feedback channel has a sufficient capacity and a receiver sends feedback information with short interval to notify his/her preferred viewpoint with no error and delay. Nevertheless, we can still use past feedback information when the feedback channel is band-limited. This section evaluates the effect of the inaccurate feedback information on the virtual viewpoint quality. We compare four schemes: ideal and fixed power assignment for target virtual viewpoints of 1.5, 2.0, and 2.5. Ideal scheme represents the case when the receiver's feedback is omniscient. The fixed schemes assign the fixed transmission power to texture and depth frames to achieve the best video quality at the target virtual viewpoints, respectively.

Fig. 8 shows the video quality of each scheme as a function of virtual viewpoint positions at an SNR of 5 dB. This figure reveals the following two observations:

- As the distance increases, the performance gap between ideal and fixed allocation schemes becomes larger up to 1.6 dB.
- When the feedback information is inaccurate and viewer's preferred viewpoint is unknown, the sender shall assign transmission powers assuming that the preferred viewpoint is center to prevent quality degradation at edge viewpoints.

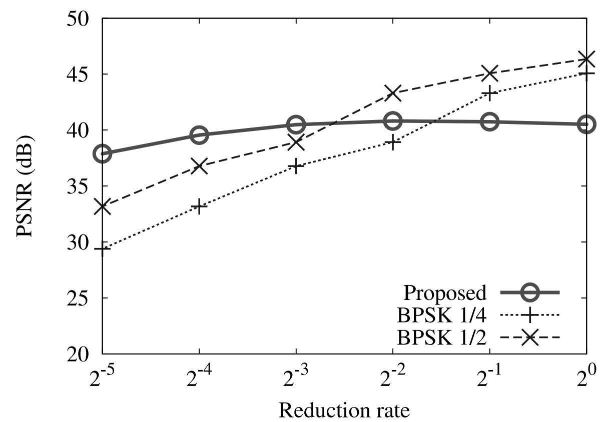
G. Discussion on Data Reduction

Above evaluations assumed that our scheme transmits all 5D-DCT coefficients to a receiver. We can adaptively reduce the number of transmitting coefficients if necessary. This section discusses the effect of traffic reduction on video quality, in terms of *reduction rate*. When the reduction rate is 2^{-1}, the half of chunks with small variance are discarded in the proposed scheme. For digital-based schemes, we control quantization parameters to make the number of digital-modulated symbols equal to the total number of analog-modulated symbols as much as possible for a fair comparison.

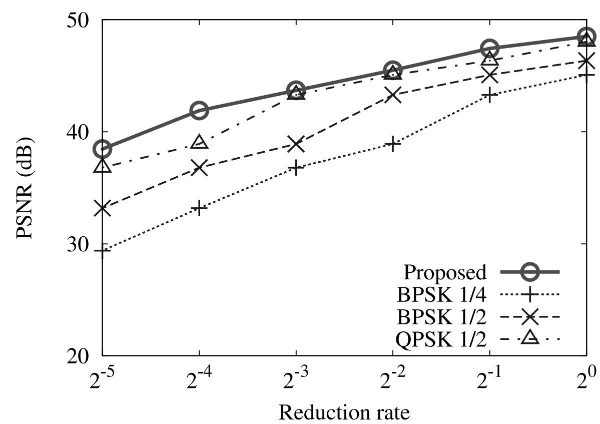
Figs. 9(a) and (b) show the video quality at virtual viewpoint 2 as a function of reduction rates. Note that we do not present the results of 16QAM and/or QPSK with $1/2$-rate convolutional code in those figures because of the same reason as Sec III-C. The key results from these figures are summarized as follows:

- When wireless channel quality is high, the proposed scheme achieves the best performance regardless of the reduction rates.
- Video quality becomes better as the reduction rate increases for low channel quality.
- For an SNR of 5 dB, the video quality at a reduction rate of 2^{-2} is slightly better than the case at a reduction rate of 2^{0} because chunks with too small variance can waste transmission powers.

Above results indicate that the benefit of our scheme can be more significant in narrower-band environment.



(a) SNR: 5 dB



(b) SNR: 15 dB

Fig. 9. PSNR vs. reduction rate: (a) SNR of 5 dB and (b) SNR of 15 dB.

H. Impact of Video Sequences

Previous evaluations use *balloons* for the reference MVD video. To evaluate the difference of the video sequence, we show the PSNR of two reference MVD videos, *kendo* and *pantomime* for digital-based schemes and the proposed scheme in Figs. 10 and 11, respectively. Here, we evaluate the PSNR of virtual viewpoint 2 and 38 for *kendo* and *pantomime*, respectively. We can see that the video quality of proposed scheme is proportional to the wireless channel quality independent of the video sequences. In addition, the proposed scheme realizes the highest performance above the channel SNR of 15 dB. For example, the proposed scheme offers an improvement of 1.2 and 2.9 dB in PSNR, respectively, compared to QPSK with rate-$1/2$ convolutional code at an SNR of 15 dB for the video sequences of *kendo* and *pantomime*.

IV. Conclusion

This paper proposed a new transmission scheme for wireless free viewpoint video streaming. Our scheme based on 5D-DCT can achieve graceful video quality at any virtual viewpoint. It was demonstrated that the proposed method can outperform digital-based MVD streaming methods in high SNR regimes.

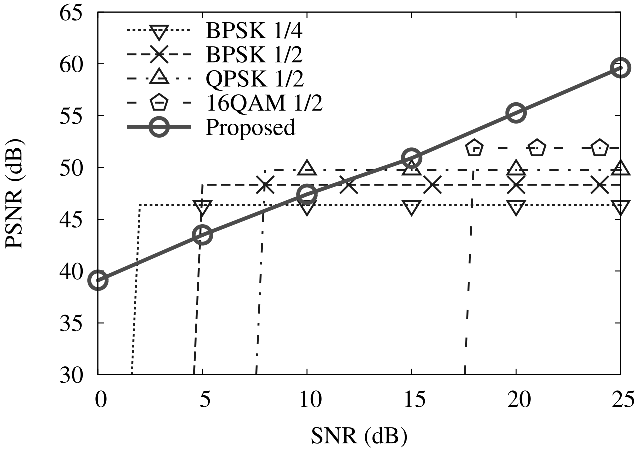


Fig. 10. PSNR vs. SNR of *kendo* at a virtual viewpoint of 2.

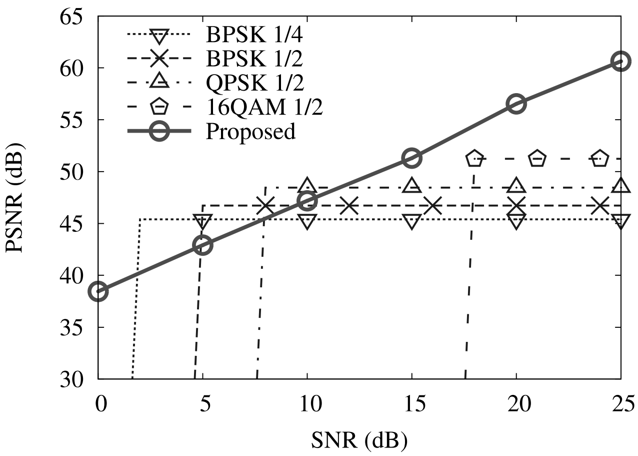


Fig. 11. PSNR vs. SNR of *pantomime* at a virtual viewpoint of 38.

In addition, we discussed the best power assignment to achieve the highest video quality for different virtual viewpoints and channel qualities. We also introduced an efficient analog-based compression scheme for band-limited systems.

ACKNOWLEDGMENT

This work was partly supported by JSPS KAKENHI Grant Number 14J09461.

REFERENCES

[1] M. Domanski, M. Bartkowiak, A. Dziembowski, T. Grajek, A. Grzelka, A. Luczak, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski, and K. Wegner, "New results in free-viewpoint television systems," in *IEEE International Conference on Multimedia and Expo*, 2016, pp. 1–6.

[2] R. Suenaga, K. Suzuki, T. Tezuka, M. P. Tehrani, K. Takahashi, and T. Fujii, "A practical implementation of free viewpoint video system for soccer games," *Three-Dimensional Image Processing, Measurement, and Applications*, vol. 9393, 2015.

[3] G. Lafruit, M. Domanski, K. Wegner, T. Grajek, T. Senoh, J. Jung, P. T. Kovak, P. Goorts, L. Jorissen, A. Munteanu, B. Ceulemans, P. Carballeira, S. Garcia, and M. Tanimoto, "New visual coding exploration in MPEG: Super-multiview and free navigation in free viewpoint TV," in *IS&T Electronic Imaging: Stereoscopic Displays and Applications XVIII*, 2016, pp. 1–9.

[4] Y. Chen, M. M. Hannuksela, T. Suzuki, and S. Hattori, "Overview of the MVC+D 3D video coding standard," *Journal of Visual Communication and Image Representation*, vol. 25, no. 4, pp. 679–688, 2014.

[5] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Stereoscopic Displays and Virtual Reality Systems*, 2004, pp. 93–104.

[6] A. D. Abreu, P. Frossard, and F. Pereira, "Optimizing multiview video plus depth prediction structures for interactive multiview video streaming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 487–500, 2015.

[7] "3D-AVC draft text 6, document JCT3V-D1002.doc," JCT-3V, 2013.

[8] H. Chui, R. Xiong, C. Luo, Z. Song, and F. Wu, "Denoising and resource allocation in uncoded video transmission," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 102–112, 2015.

[9] W. S. Kim, A. Ortega, P. Lai, and D. Tian, "Depth map coding optimization using rendered view distortion for 3D video coding," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3534–3545, 2015.

[10] T. Y. Chung, J. W. Dong, and K. C. Su, "Efficient depth video coding based on view synthesis distortion estimation," in *Visual Communications and Image Processing*, 2012, pp. 1–4.

[11] 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.

[12] 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.

[13] 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.

[14] D. He, C. Luo, F. Wu, and W. Zeng, "Swift: A hybrid digital-analog scheme for low-delay transmission of mobile stereo video," in *ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, 2015, pp. 327–336.

[15] G. Wang, K. Wu, Q. Zhang, and L. M. Ni, "SimCast: Efficient video delivery in MU-MIMO WLANs," in *IEEE Conference on Computer Communications*, 2014, pp. 2454–2462.

[16] R. Xiong, F. Wu, J. Xu, X. Fan, C. Luo, and W. Gao, "Analysis of decorrelation transform gain for uncoded wireless image and video communication," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1820–1833, 2016.

[17] D. Yang, Y. Bi, Z. Si, Z. He, and K. Niu, "Performance evaluation and parameter optimization of SoftCast wireless video broadcast," in *International Conference on Mobile Multimedia Communciations*, 2015, pp. 79–84.

[18] X. Fan, R. Xiong, D. Zhao, and F. Wu, "Layered soft video broadcast for heterogeneous receivers," *IEEE Transactions on Circuits and Systems for Video Technology*, 2015.

[19] X. Lin, Y. Liu, and L. Zhang, "Scalable video SoftCast using magnitude shift," in *IEEE Wireless Communications and Networking Conference*, 2015, pp. 1996–2001.

[20] C. Lan, D. He, C. Luo, F. Wu, and W. Zeng, "Progressive pseudo-analog transmission for mobile video live streaming," in *IEEE Visual Communications and Image Processing*, 2015, pp. 1–4.

[21] ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, "3D video subjective quality assessment test plan," Doc. JCT3V-F1011, 2013.

[22] Fujii Laboratory at Nagoya University. [Online]. Available: http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/

[23] A. A. Muhit, "H.264 Baseline Codec v2," http://www.mathworks.com/matlabcentral/fileexchange/40359-h-264-baseline-codec-v2, 2013.

[24] [Online]. Available: https://hevc.hhi.fraunhofer.de /svn/svn_3DVCSoftware/tags/HTM-13.0/