

FreeCast: Graceful Free-Viewpoint Video Delivery

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

TR2018-134 September 20, 2018

Abstract

Wireless multi-view plus depth (MVD) video streaming enables free viewpoint video playback on wireless devices, where a viewer can freely synthesize any preferred virtual viewpoint from the received MVD frames. Existing schemes of wireless MVD streaming use digital-based compression to achieve better coding 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 assign quantization levels and transmission power across MVD frames are cumbersome. To realize highquality wireless MVD video streaming, we propose a novel graceful video delivery scheme, called FreeCast. FreeCast directly transmits linear-transformed signals based on five-dimensional discrete cosine transform (5D-DCT), without digital quantization and entropy coding operations. In addition, we exploit a fitting function based on multidimensional Gaussian Markov random field (GMRF) model for overhead reduction to mitigate rate and power loss due to large overhead. The proposed FreeCast achieves graceful video quality with the improvement of wireless channel quality under a low overhead requirement. In addition, the parameter optimization to achieve highest video quality can be simplified by only controlling transmission power assignment. Performance results with several test MVD video sequences show that FreeCast yields better video quality in band-limited environments by significantly decreasing the amount of overhead. For instance, structural similarity (SSIM) performance of FreeCast is approximately 0.127 higher than the existing graceful video delivery schemes across wireless channel quality, i.e., signal-tonoise ratio (SNR), of 0 to 25 dB at a transmission symbol rate of 37.5 Msymbols/s.

IEEE Transactions on Multimedia

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

FreeCast: Graceful Free-Viewpoint Video Delivery

Takuya Fujihashi, *Member, IEEE*, Toshiaki Koike-Akino, *Senior Member, IEEE*,
Takashi Watanabe, *Member, IEEE*, and Philip V. Orlik, *Senior Member, IEEE*

Abstract—Wireless multi-view plus depth (MVD) video streaming enables free viewpoint video playback on wireless devices, where a viewer can freely synthesize any preferred virtual viewpoint from the received MVD frames. Existing schemes of wireless MVD streaming use digital-based compression to achieve better coding 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 assign quantization levels and transmission power across MVD frames are cumbersome. To realize high-quality wireless MVD video streaming, we propose a novel graceful video delivery scheme, called FreeCast. FreeCast directly transmits linear-transformed signals based on five-dimensional discrete cosine transform (5D-DCT), without digital quantization and entropy coding operations. In addition, we exploit a fitting function based on multidimensional Gaussian Markov random field (GMRF) model for overhead reduction to mitigate rate and power loss due to large overhead. The proposed FreeCast achieves graceful video quality with the improvement of wireless channel quality under a low overhead requirement. In addition, the parameter optimization to achieve highest video quality can be simplified by only controlling transmission power assignment. Performance results with several test MVD video sequences show that FreeCast yields better video quality in band-limited environments by significantly decreasing the amount of overhead. For instance, structural similarity (SSIM) performance of FreeCast is approximately 0.127 higher than the existing graceful video delivery schemes across wireless channel quality, i.e., signal-to-noise ratio (SNR), of 0 to 25 dB at a transmission symbol rate of 37.5 Msymbols/s.

I. INTRODUCTION

Free viewpoint video [1]–[3] is an emerging and attractive technique to observe a three-dimensional (3D) scene from freely switchable angles. Fig. 1 shows an example of the streaming systems capable of free viewpoint video delivery, where a large number of closely spaced camera arrays are deployed to capture texture and depth frames of a 3D scene such as a football game. 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], [6] from the received MVD frames.

Takuya Fujihashi is with Graduate School of Science and Engineering, Ehime University, Matsuyama, Ehime, 790-8577 JAPAN e-mail: (fujihashi@cs.ehime-u.ac.jp).

Toshiaki Koike-Akino and Philip V. Orlik are with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139 USA e-mail: (koike@merl.com, porlik@merl.com).

Takashi Watanabe is with Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, 565-0871 JAPAN e-mail: (watanabe@ist.osaka-u.ac.jp).

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

Manuscript received February 20, 2018; revised July 13, 2018; accepted August 15, 2018.

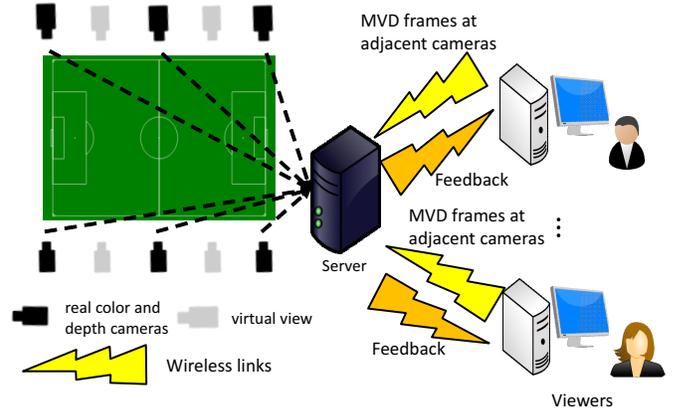


Fig. 1. Wireless MVD video streaming systems for free viewpoint rendering.

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 [7] or 3D-advanced video coding (AVC) [8] 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, possible few bit errors occurred in the bit stream during communications can cause a destructive collapse of texture and depth decoding. The decoding failure in turn disables rendering operation of DIBR synthesis. As a result, the video quality of virtual viewpoints degrades significantly. This phenomenon is called cliff effect [9]. Second, the video quality does not improve even when the wireless channel quality is improved unless an adaptive rate control of video and channel coding is performed in real-time according to the rapid fading channels. Finally, quantization is a lossy process and its distortion cannot be recovered at the receiver. Some studies [10] have been proposed to mitigate the cliff effect in the digital-based MVD scheme by introducing layered source and channel codings. However, the cliff effect is converted into staircase effect. In the staircase effect, the video quality discontinuously improves with the improvement of wireless channel quality.

In addition, the digital-based wireless MVD scheme needs

to solve a complicated parameter optimization to achieve the best quality for a 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* [11], [12], which is often cumbersome to derive the solution because it is a combinatorial problem which needs to take a good care of nonlinear distortion due to quantization errors and DIBR synthesis.

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. To overcome these issues, we proposed a new wireless MVD transmission scheme [13], motivated by the studies on graceful video delivery [14]. The key idea of this scheme is skipping quantization and entropy coding at the encoder. Specifically, this scheme jointly transforms texture and depth frames using 5D-discrete cosine transform (DCT), whose output is then scaled and directly mapped to transmission signals without relying on digital modulation schemes. 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 graceful video quality according to the wireless channel quality, without any cliff effect. In addition, this scheme simplifies the optimization problems by reformulating into a simple power assignment problem because bit allocation for quantization is not required in this scheme. The encoder assigns appropriate transmission power to each texture and depth frame before the transformation based on the viewer's preferred viewpoint. It was demonstrated that the graceful video delivery scheme achieved graceful video quality with the improvement of wireless channel quality and better performance compared to the digital-based schemes at a certain channel quality.

However, graceful video delivery schemes still have a drawback in its large amount of overhead. In graceful video delivery schemes, a sender scales linear-transformed video signals, such that the receiver noise can be minimized, before transmission. The scaling factor is based on the power values of the linear-transformed video signals. Hence, the sender needs to transmit the power information of all the video signals without errors to decode the signals at the receiver. Since the transmission of this metadata causes large overhead, the video quality degrades due to power and rate loss in band-limited environments. To reduce the amount of metadata overhead, the existing scheme [13], [14] divides the video signals into multiple chunks, and then transmits a smaller amount of metadata corresponding to each chunk. However, the chunk division may degrade the video quality due to improper scaling operation.

In this paper, we extend our preliminary study of the graceful video delivery scheme for free-viewpoint streaming [13], called **FreeCast**, to achieve better video quality in band-limited environments by reducing the amount of overhead. To yield better performance in terms of overhead reduction, FreeCast introduces a fitting function obtained from multi-dimensional Gaussian Markov random field (GMRF) [15] at the sender and the receiver to obtain the power information

with few parameters. Specifically, the sender finds a few parameters for the fitting function from the power information of MVD sequences and sends the parameters as metadata to the receiver. The receiver obtains the power values from the same fitting function and received fitting parameters. From evaluation results, the video quality of the existing chunk-based scheme is significantly degraded in band-limited environments whereas FreeCast still keeps better video quality in such environments.

II. RELATED WORKS AND OUR CONTRIBUTIONS

A. Graceful Video Delivery

Graceful video delivery schemes have been recently proposed for single-view video in [14]–[24]. For example, SoftCast [14] skips digital 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 [16] and AirScale [17] extended graceful 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 [18], 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 graceful video delivery to wireless MVD video streaming. To realize high quality and graceful video delivery in free viewpoint video, FreeCast has the following major contributions.

- A sender uses 5D-DCT for MVD video frames to exploit inter-view and texture-depth correlations for performance improvement.
- We model MVD video signals using multidimensional GMRF to obtain a fitting function for overhead reduction.
- We show how to optimize the power assignment across texture and depth frames to achieve the highest video quality in each virtual viewpoint.

B. Overhead Reduction in Graceful Video Delivery

In graceful video delivery, a sender needs to let the receiver know the power information of all the linear-transformed video signals to demodulate the signals. However, it requires a relatively large amount of overhead. To reduce the amount of overhead, the existing schemes use chunk division whereas it causes improper power allocation. To achieve better video quality under a low overhead requirement, a method proposed in [15] exploited a Lorentzian-based fitting function to obtain the power information at the receiver only with few parameters, while achieving an excellent streaming quality in band-limited environments.

In our study, we extend the fitting function for wireless MVD video streaming. To realize overhead reduction in MVD video streaming, we model MVD video signals using 5D first-order GMRF to obtain a function to fit the power spectrum density of 5D-DCT coefficients. By using the fitting function, the sender just needs to send nine parameters as metadata to decode the video signals at the receiver. Since the estimation

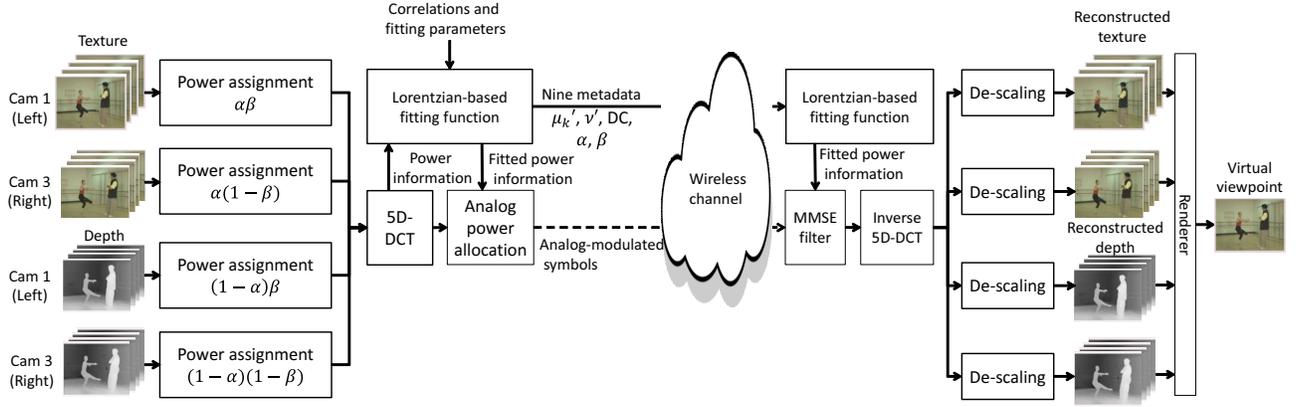


Fig. 2. Schematic of FreeCast: the proposed graceful video delivery scheme for free viewpoint video. Quadratic power assignment among color-texture and depth frames facilitates view synthesis optimization. 5D-DCT offers high compaction of MVD video signals. Parameterized power spectrum fitting constrains the required amount of metadata overhead. The end user at the receiver can freely change the viewpoint in real-time rendering for playback.

error of the fitting function is enough small for typical video sequences, FreeCast can yield better video quality under a low overhead requirement.

C. Digital-based Free Viewpoint Video Delivery

Conventional schemes on free viewpoint video delivery mainly use digital-based compression and transmission techniques for MVD video frames. In contrast to single-view video delivery, free viewpoint video delivery schemes need to optimize view synthesis problem to realize highest video quality at any virtual viewpoint. However, it is often cumbersome to find the solution due to the combinatorial problem with nonlinear quantization. To solve the optimization problem, a fast mode decision algorithm was proposed in [25] for depth videos to find the best solution with small computational overhead. In [11], view synthesis distortion models were proposed to realize better video quality under low computational complexity. The computational complexity of view synthesis optimization was decreased by utilizing less depth sensitivity fidelity regions for depth coding in [26].

In our study, we also aim at quality optimization at any requested virtual viewpoint. FreeCast can simplify the view synthesis optimization problem into a power assignment problem by skipping quantization and entropy coding. From our analysis, we found that view synthesis optimization in FreeCast can be solved efficiently with a quadratic fitting.

III. FREECAST: GRACEFUL 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, 3) to simplify the issues of bit and power assignments to achieve the best video quality based on viewer's preferred viewpoint, and 4) to reduce the amount of metadata to mitigate rate and power loss due to large overhead.

Fig. 1 shows one of system models under consideration. The sender has multi-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 captured data at several adjacent cameras near the requested virtual viewpoint 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 FreeCast. The encoder first assigns transmission power to each texture and depth frame, followed by 5D-DCT operation. Using the power spectrum information of the DCT coefficients, we find the best parameters of a fitting function based on multidimensional GMRF. The DCT coefficients are then scaled and analog-modulated according to the fitted power information for wireless transmissions. Next, the encoder sends the analog modulated symbols as well as the fitting parameters to the receiver over a wireless channel, which is often impaired with additive white Gaussian noise (AWGN) and time-varying fading. At the receiver side, the decoder uses minimum mean-square error (MMSE) filter to obtain the transmitted DCT coefficients. Here, the filter gain is obtained from the received fitting parameters and the corresponding fitting function. 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. After power assignment for each DCT coefficient, the DCT coefficients are mapped to I (in-phase) and Q (quadrature-phase) components for analog wireless transmissions.

Let x_i denote the i th analog-modulated symbol, which is the i th DCT coefficient s_i scaled by a factor of g_i for noise

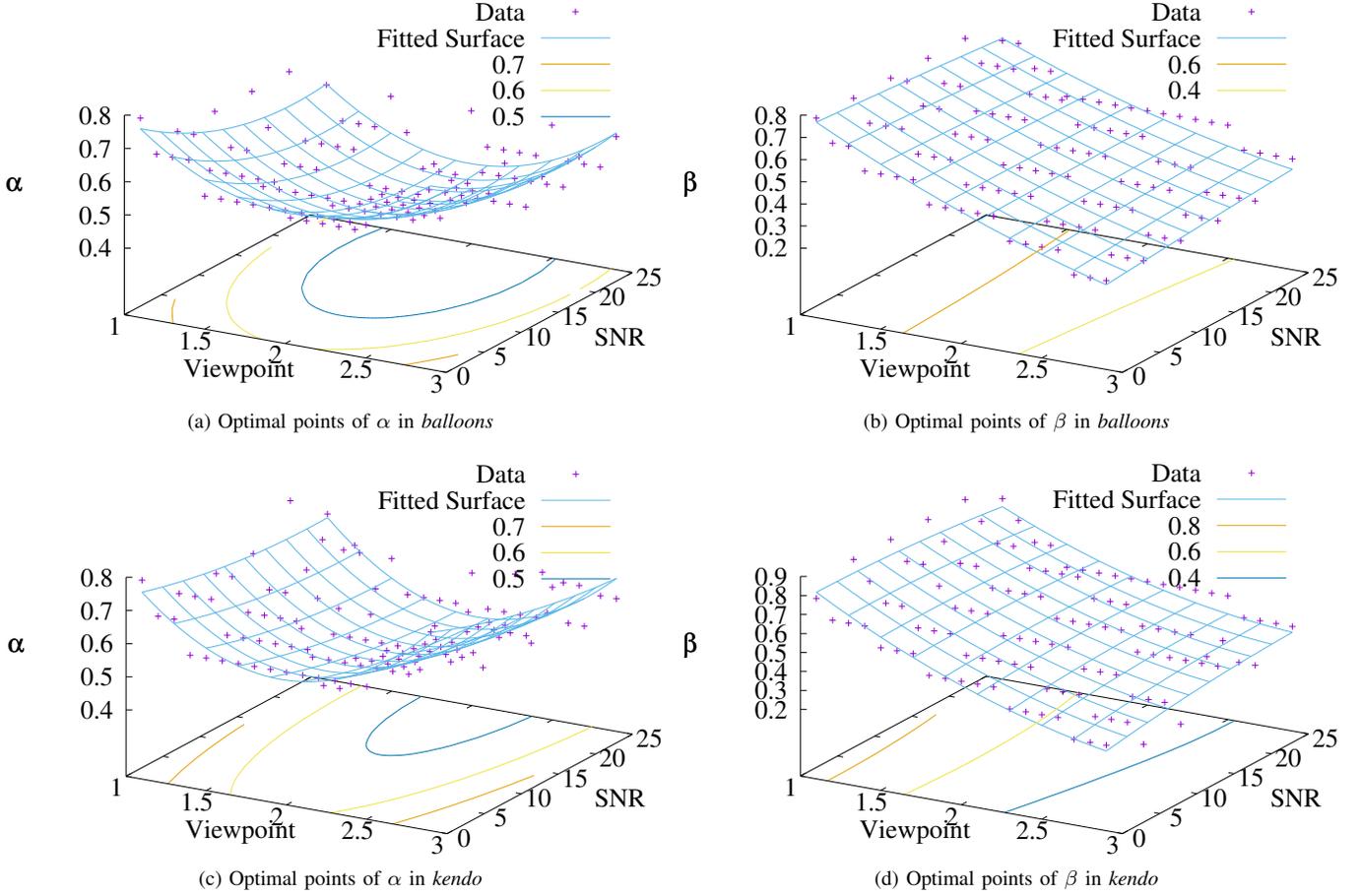


Fig. 3. Optimal values of α and β for inter-texture/depth power assignment in MVD video frames of *balloons* and *kendo*.

reduction as follows:

$$x_i = g_i \cdot s_i. \quad (1)$$

The optimal scale factor g_i is obtained by minimizing the mean-square error (MSE) under the power constraint with total power budget P as follows:

$$\min_{\{g_i\}} \text{MSE} = \mathbb{E} \left[(s_i - \hat{s}_i)^2 \right] = \sum_i^N \frac{\sigma^2 \lambda_i}{g_i^2 \lambda_i + \sigma^2}, \quad (2)$$

$$\text{s.t.} \quad \frac{1}{N} \sum_i^N g_i^2 \lambda_i = P, \quad (3)$$

where $\mathbb{E}[\cdot]$ denotes expectation, \hat{s}_i is a receiver estimate of the transmitted DCT coefficient, λ_i is the power of the i th DCT coefficient, N is the number of DCT coefficients, and σ^2 is a receiver noise variance. As shown in [14], the near-optimal solution is expressed as

$$g_i = \lambda_i^{-1/4} \sqrt{\frac{NP}{\sum_j \sqrt{\lambda_j}}}. \quad (4)$$

B. Decoder

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

$$y_i = x_i + n_i, \quad (5)$$

where y_i is the i th received symbol and n_i is an effective AWGN with a variance of σ^2 (which is already normalized by wireless channel strength in the presence of fading attenuation). The DCT coefficients are extracted from I and Q components via an MMSE filter [14]:

$$\hat{s}_i = \frac{g_i \lambda_i}{g_i^2 \lambda_i + \sigma^2} \cdot y_i. \quad (6)$$

The decoder then obtains corresponding video sequence by taking the inverse 5D-DCT for the filter output \hat{s}_i . 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 [27].

FreeCast simplifies the parameter optimization by removing quantization and entropy coding. Specifically, the distortion of texture and depth frames is reduced to a simple function of the assigned transmission power. FreeCast assigns transmission power for the frames under a certain power budget before the 5D-DCT operation. To control the power assignment, we use two parameters for simplicity: 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. The optimal parameters for inter-texture/depth power assignment are highly dependent on the target viewpoint and MVD contents. Nevertheless, as described below, a useful insight can be observed for optimizing those parameters; specifically the optimal values agreed well a quadratic function of viewpoint and channel condition.

Figs. 3 (a) through (d) show the optimal assignment values of α and β in different MVD test video sequences of *balloons* and *kendo* [28], respectively. To obtain the optimal values, we evaluated the video quality by sweeping α and β in a finite grid under the different virtual viewpoints and channel SNRs, and then plotted the values of α and β with the highest video quality. From these figures, the optimized α and β parameters can be obtained by a quadratic function of $f(p, q) = ap^2 + bq^2 + cpq + dp + eq + h$, where p and q are virtual viewpoint and SNR in dB, respectively. More specifically, both parameters in each test video sequence can be derived as follows:

$$\alpha_{\text{balloons}} = 0.20p^2 - 0.82p - 0.02q + 1.43, \quad (7)$$

$$\beta_{\text{balloons}} = 0.05p^2 - 0.46p - 0.01q + 1.23, \quad (8)$$

$$\alpha_{\text{kendo}} = 0.25p^2 - 0.99p - 0.01q + 1.55, \quad (9)$$

$$\beta_{\text{kendo}} = 0.09p^2 - 0.67p + 1.47. \quad (10)$$

From the above functions, we can see the similar trends in power assignment even in the different MVD test video sequences. Although the optimal power assignment can vary over different video sequences, the observed quadratic trend across SNR and viewpoints is expected to facilitate efficient fine-tuning method for adaptive power assignment. These two parameters, i.e., α and β , are sent from the transmitter to the receiver for de-scaling operations.

D. Overhead Reduction

To carry out MMSE filtering in (6) at the receiver, the sender needs to notify λ_i of all coefficients without errors as metadata. For example, when the sender transmits eight MVD video frames with the resolution of 1024×768 and YUV 4:4:4 format, the sender needs to transmit metadata for all DCT coefficients, i.e., $1024 \times 768 \times 8 \times 2 \times 4 = 50,331,648$ variables in total, to the receiver. In this case, the amount of metadata is approximately 2.9 bits/pixel after Huffman coding. This overhead induces quality degradation due to rate and power losses in transmission of analog-modulated symbols. To reduce the overhead, existing schemes [13] divide the DCT coefficients into chunks and carry out scaling and MMSE filter for each chunk. However, overhead is still high in general and the chunk division can cause performance degradation due

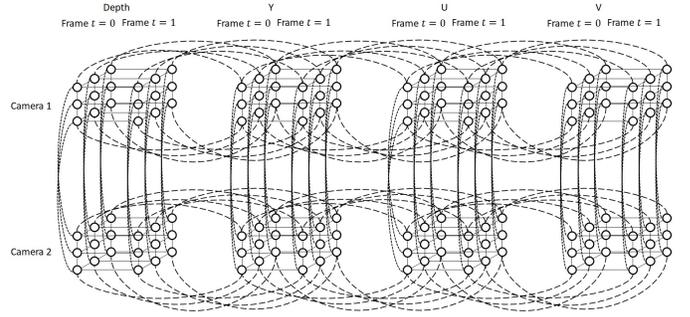


Fig. 4. 5D first-order GMRF for MVD video signals.

to a loss of optimality for scaling with respect to (6). For example, when the sender divides the DCT coefficients into chunks with a relatively large size of 64×48 pixels, 16,384 variables of metadata are still required every a few frames (specifically, GoP size). Although the amount of metadata is reduced to approximately $9.6 \cdot 10^{-4}$ bits/pixel in this large chunk size, the video quality is significantly degraded in narrowband environments as we will see later.

In order to reduce overhead while keeping video quality high, FreeCast uses a parametric function to approximate the power values λ_i for a variety of MVD video sequences. More specifically, we first regard MVD signals as 5D first-order GMRF, i.e., horizontal, vertical, time, inter-camera, and depth-Y-U-V correlations, as shown in Fig. 4. By taking 5D-DCT operation for such MVD signals, the power spectrum density of 5D-DCT coefficients can be asymptotically obtained by the Lorentzian functions as follows:

$$F(i, j, k, l, m) = \nu \cdot \frac{1}{1 + f_1^2(i)} \cdot \frac{1}{1 + f_2^2(j)} \cdot \frac{1}{1 + f_3^2(k)} \cdot \frac{1}{1 + f_4^2(l)} \cdot \frac{1}{1 + f_5^2(m)}, \quad (11)$$

$$f_1(i) = \mu_1 i, \quad f_2(j) = \mu_2 j, \quad f_3(k) = \mu_3 k, \\ f_4(l) = \mu_4 l, \quad f_5(m) = \mu_5 m. \quad (12)$$

Here, μ_k and ν are fitting parameters. Note that above equations express the power spectrum density of the DCT coefficients except the DC component. FreeCast ignores the DC component from fitting operation because the DC component cannot be modeled by the fitting function. Although 5D-DCT may not be the best possible transform to decorrelate inter-view and depth-texture interactions of MVD video sequences, it is of great advantage that the DCT is easily extendable to any dimensions having asymptotic expression of the power spectrum density under the GMRF model.

The sender finds the best parameters of ν , μ_1 , μ_2 , μ_3 , μ_4 , μ_5 based on the empirical power of the non-DC components by least-squares fitting. We found that the estimation error is small enough for real video sequences; more specifically, the normalized mean-square error (NMSE) between empirical and fitting values is approximately -25.9 dB on average across test video sequences of *balloons*, *kendo*, *champagne*, and *pantomime* [28]. The accurate fitting function having such

TABLE I
PARAMETERS FOR FITTING FUNCTION

Video Sequence	ν	μ_1	μ_2	μ_3	μ_4	μ_5
Balloons	$3.26 \cdot 10^5$	0.04	0.13	0.77	0.49	0.38
Kendo	$7.61 \cdot 10^3$	0.03	0.05	0.40	0.26	0.07

TABLE II
AMOUNT OF METADATA IN FREECAST AND CHUNK-BASED FREECAST AT VIDEO SEQUENCES OF BALLOONS AND KENDO

Video Sequence	Amount of metadata (bits/pixel) at different chunk sizes					
	8×6	16×12	32×24	64×48	1024×768	FreeCast
balloons	$6.1 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$3.7 \cdot 10^{-3}$	$9.2 \cdot 10^{-4}$	$5.0 \cdot 10^{-6}$	$5.4 \cdot 10^{-7}$
kendo	$6.4 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$3.9 \cdot 10^{-3}$	$9.9 \cdot 10^{-4}$	$8.4 \cdot 10^{-6}$	$5.4 \cdot 10^{-7}$
average	$6.2 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$	$9.6 \cdot 10^{-4}$	$6.7 \cdot 10^{-6}$	$5.4 \cdot 10^{-7}$

a small estimation error contributes to maintaining high video quality in FreeCast even with relatively low overhead.

The sender then transmits nine parameters of ν , μ_1 , μ_2 , μ_3 , μ_4 , μ_5 , DC component, and power assignment parameters of α and β as the metadata. In addition, we employ Huffman coding to compress the parameters before transmission. Here, the amount of metadata is approximately $5.4 \cdot 10^{-7}$ bits/pixel, which is significantly smaller than that of the standard chunk-based methods. We assume that the encoder uses 1/2-rate convolutional coding and binary phase-shift keying (BPSK) for the compressed metadata transmissions.

IV. PERFORMANCE EVALUATIONS

A. Simulation Settings

Performance Metric: We evaluate the video quality in terms of the peak SNR (PSNR) and structural similarity (SSIM) [29]. PSNR is defined as follows:

$$\text{PSNR} = 10 \log_{10} \frac{(2^L - 1)^2}{\epsilon_{\text{MSE}}}, \quad (13)$$

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. The original video is generated by DIBR given distortion-less adjacent MVD frames. We obtain the average YUV-PSNR across test video sequences. 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 also obtain the average YUV-SSIM across test video sequences.

Test Video: We use two standard reference MVD videos, namely, *balloons* and *kendo* with 30 fps from Fujii Laboratory at Nagoya University [28]. We use two cameras, i.e., camera 1 and 3, with a resolution of 1024×768 pixels for texture and depth frames. The distance between two cameras is 10 cm long.

Video Encoder: We set the GoP size for all reference schemes to eight video frames. In the existing chunk-based schemes of graceful video delivery, we consider five different chunk sizes of 8×6 , 16×12 , 32×24 , 64×48 , and 1024×768 pixels to discuss an effect of chunk division. For digital schemes, we use 3D HEVC test model (HTM) software video encoder/decoder v13.0 [30] to generate a bit stream from the test MVD video.

We encode video frames of viewpoint 3 in one GoP into one I-frame and subsequent seven P-frames. Video frames of viewpoint 1 are encoded into eight P-frames by using both motion compensation and disparity compensation.

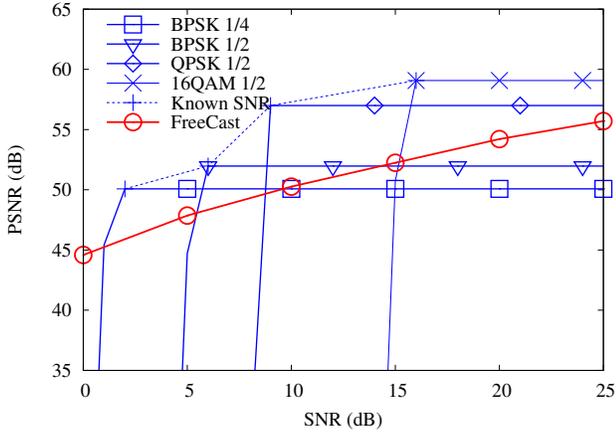
Rendering Software: To synthesize a virtual viewpoint from the received texture and depth frames, we used HTM software renderer v13.0 [30]. The renderer requires two texture and depth frames as the input to produce video frames at a virtual viewpoint.

Wireless Settings: The received symbols are impaired by an AWGN channel. We first set the channel symbol rate in reference schemes to 150.0 and 37.5 Msymbols/s. Especially, 37.5 Msymbols/s is within a reasonable bandwidth range of Wi-Fi communications, i.e., 20 MHz to 40 MHz. In Sec. IV-D2, we consider the different channel symbol rates from 4.7 to 150.0 Msymbols to evaluate the performance in narrowband to broadband environments. 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 BPSK, quadrature PSK (QPSK), or 16-ary quadrature-amplitude modulation (16QAM).

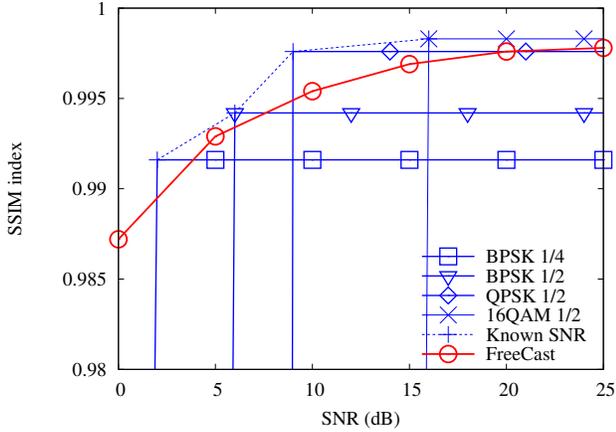
Fitting Parameters: Table I shows Lorentzian fitting parameters of two MVD video sequences to model the power spectrum information of 5D-DCT coefficients.

B. Discussion on Overhead Reduction

We first compare the amount of overhead in FreeCast and the existing scheme [13], namely, chunk-based FreeCast with different chunk sizes in terms of *bits/pixel*. Table II shows the overhead in each reference scheme with different video sequences. It is verified that FreeCast needs a lower overhead requirement by a few orders of magnitude, compared to chunk-based FreeCast irrespective of the video sequences. The chunk-based FreeCast schemes require a large amount of metadata in a small chunk size to yield better video quality. This reduction in FreeCast scheme saves transmission power and lead to additional quality improvement by allocating the saved power to the transmission of analog-modulated symbols. For example, the amount of the metadata in FreeCast is approximately $8.6 \cdot 10^{-6}$ times lower than chunk-based FreeCast with a chunk size of 8×6 pixels and $8.0 \cdot 10^{-2}$ times lower



(a) Average PSNR performance at the center virtual viewpoint



(b) Average SSIM performance at the center virtual viewpoint

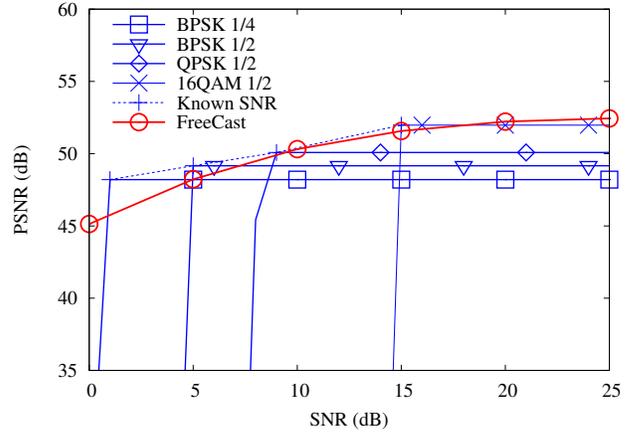
Fig. 5. Average video quality vs. SNR at channel symbol rate of 150.0 Msymbols/s.

than chunk-based FreeCast with a chunk size of 1024×768 pixels on average across two MVD video sequences.¹

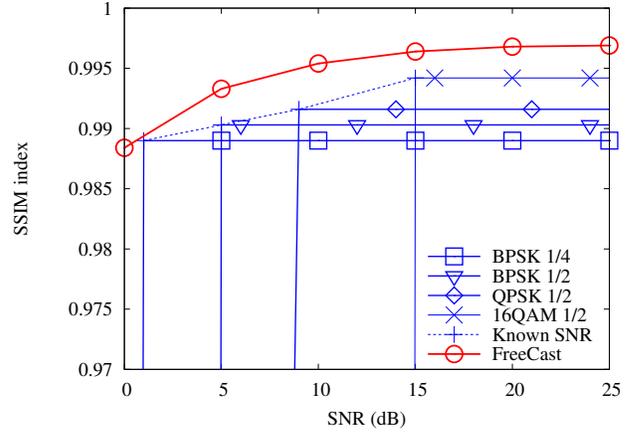
C. FreeCast vs. Digital-based Schemes

We first evaluate the performance of the proposed analog transmission scheme in comparison to five 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. In addition, “Known SNR” scheme assumes that the sender can know instantaneous channel quality and change coding rates according to the channel quality. Fig. 5(a) shows the average PSNR performance at virtual viewpoint 2 (center between left viewpoint 1 and right viewpoint 3) and channel symbol rate of 150 Msymbols/s across two test video sequences as a function of the channel SNR. In addition, Fig. 5(b) shows the average SSIM performance at the same virtual viewpoint across two test video sequences as a function of the channel SNR.

¹We divide the average amount of metadata in FreeCast by that of chunk-based FreeCast with a chunk size of 8×6 and 1024×768 pixels, respectively, to obtain the numbers of $8.6 \cdot 10^{-6}$ and $8.0 \cdot 10^{-2}$.



(a) Average PSNR performance at the center virtual viewpoint

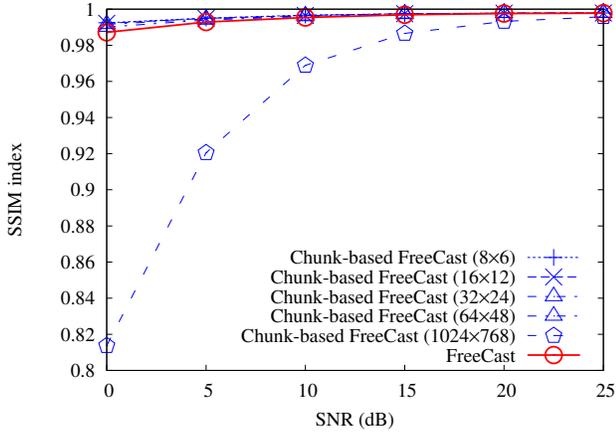


(b) Average SSIM performance at the center virtual viewpoint

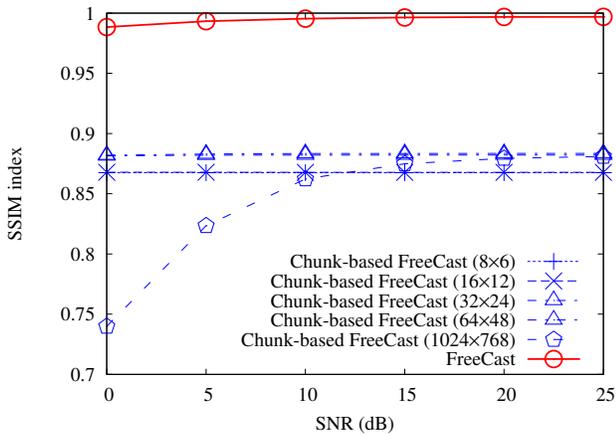
Fig. 6. Average video quality vs. SNR at channel symbol rate of 37.5 Msymbols/s.

From these figures, FreeCast prevents the cliff and constant quality by skipping quantization and error-sensitive entropy coding, and achieves graceful video quality at the virtual viewpoint without channel SNR information. On the other hand, digital-based schemes suffer from cliff-effect at different channel qualities. For example, SSIM performance in QPSK with rate-1/2 convolutional code stays the same at 0.998 when the channel SNR changes from 9 dB to 25 dB while the video quality sharply degrades when the channel SNR is below 9 dB. In view of Known SNR scheme, it achieves the best performance of all the reference schemes, where the sender can change the coding rates according to the instantaneous channel SNR. This is usually ideal for low-latency applications because the channel SNR is unknown and varies with time.

We then evaluate the video quality of the digital-based schemes and FreeCast in a narrowband environment to discuss an impact of overhead reduction on video quality. Figs. 6(a) and (b) show the average PSNR and SSIM performance at the center virtual viewpoint and channel symbol rate of 37.5 Msymbols/s across two test video sequences as a function of the channel SNR, respectively. Here, FreeCast outperforms the Known SNR scheme in terms of SSIM. Since the overhead



(a) SSIM performance at a channel symbol rate of 150.0 Msymbols/s



(b) SSIM performance at a channel symbol rate of 37.5 Msymbols/s

Fig. 7. Average SSIM performance as a function of channel SNRs in different channel symbol rates.

in FreeCast is small and estimation errors in the proposed fitting function are enough small, FreeCast keeps high video quality at a low channel symbol rate. For example, in Fig. 6(b), FreeCast improves SSIM performance by 0.002 compared to the Known SNR scheme at a channel 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 FreeCast.

D. FreeCast vs. Chunk-based Graceful Video Delivery

Previous evaluations showed that an overhead reduction in FreeCast performs well in a narrowband environment by comparison with the digital-based schemes. We next compare FreeCast performance with the existing chunk-based schemes of graceful video delivery, i.e., chunk-based FreeCast.

1) *Effect of Wireless Channel Quality*: Figs. 7(a) and (b) show the average SSIM performance at the center virtual viewpoint as a function of the channel SNRs at channel symbol rates of 150.0 Msymbols/s and 37.5 Msymbols/s, respectively.

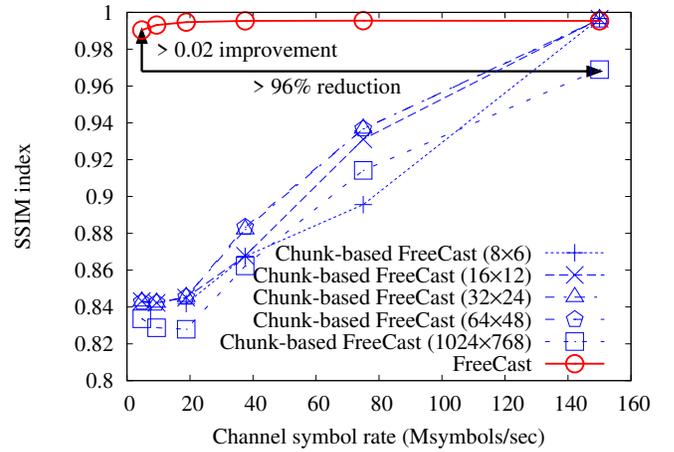


Fig. 8. SSIM vs. channel symbol rate at the center virtual viewpoint and wireless channel SNR of 10 dB.

At a high channel symbol rate, we can see that FreeCast achieves lower overhead and better video quality compared to chunk-based FreeCast with a chunk size of 1024×768 pixels irrespective of wireless channel SNRs. In addition, FreeCast achieves the similar performance with chunk-based FreeCast with a small chunk size because estimation errors in the proposed fitting function are sufficiently small.

At a low channel symbol rate, it is interesting to note that the performance of chunk-based FreeCast is significantly degraded due to large overhead. More specifically, since the most of 5D-DCT coefficients are discarded to send the corresponding metadata, it causes high distortion in texture and depth values and the collapse of virtual viewpoint rendering. For example, FreeCast achieves SSIM improvement by 0.127 over chunk-based FreeCast with the chunk size of 8×6 pixels across channel SNRs of 0 to 25 dB.

2) *Effect of Limited Bandwidth*: Above evaluations demonstrated that FreeCast is well-performed in a narrowband environment compared to digital-based schemes and chunk-based FreeCast because of a low overhead requirement. To discuss an effect of bandwidth limitation on FreeCast performance in more detail, this section measures the video quality under the different channel symbol rates.

Fig. 8 shows the SSIM performance at the center virtual viewpoint and channel SNR of 10 dB as a function of channel symbol rates varying from 5 to 150 Msymbols/s. The key results from this figure are summarized as follows:

- FreeCast achieves the best video quality even in bandwidth-limited environments and keeps almost the same quality as the broadband environment up to the channel symbol rate of 37.5 Msymbols/s.
- The video quality in chunk-based FreeCast schemes significantly degrades at low channel symbol rates.

In view of video quality and traffic reduction, FreeCast improves SSIM performance by 0.02 with 96.9% traffic reduction compared to chunk-based FreeCast with a chunk size of 1024×768 pixels.

Finally, Figs. 9 and 10 compare the visual quality of FreeCast and chunk-based FreeCast for the video sequences

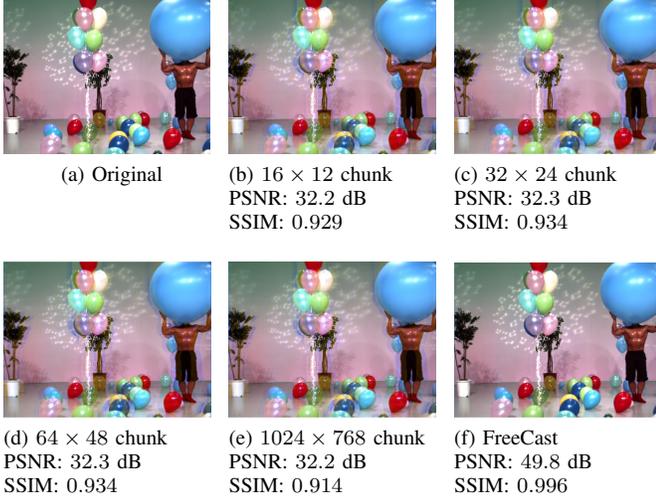


Fig. 9. Snapshot of *balloons* (frame #1) in each scheme at an SNR of 10 dB and channel symbol rate of 75.0 Msymbols/s.

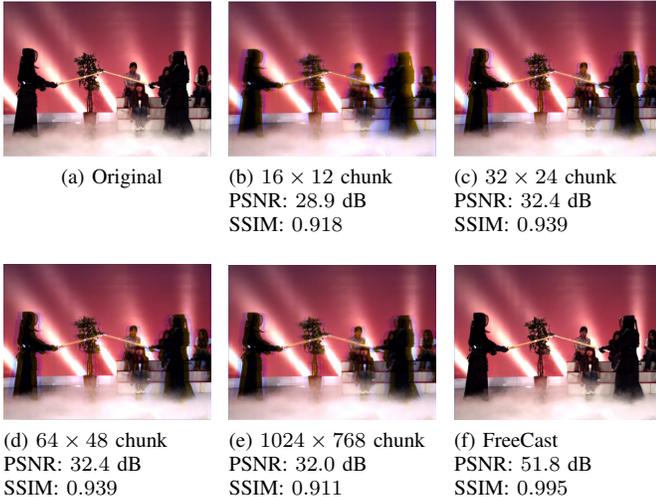


Fig. 10. Snapshot of *kendo* (frame #1) in each scheme at an SNR of 10 dB and channel symbol rate of 75.0 Msymbols/s.

of *balloons* and *kendo*. The video frame is transmitted at the channel SNR of 10 dB and the channel symbol rate of 75.0 Msymbols/s. For *balloons*, the SSIMs achieved by chunk-based FreeCast with chunk size of 16×12 , 32×24 , 64×48 , and 1024×768 pixels are 0.929, 0.934, 0.934, and 0.914, respectively, whereas 0.996 is achieved by FreeCast. From the snapshots, we can clearly see that chunk-based FreeCast schemes with small to large chunk sizes provide low-quality images (having blur) at a requested virtual viewpoint. In contrast, FreeCast can synthesize a clean virtual image with details.

E. Discussion on Feedback Delay

Previous evaluations assume that a feedback channel has a sufficient capacity and a receiver sends feedback information with a short interval to notify his/her preferred viewpoint with no error and delay. Based on the preferred viewpoint, a sender

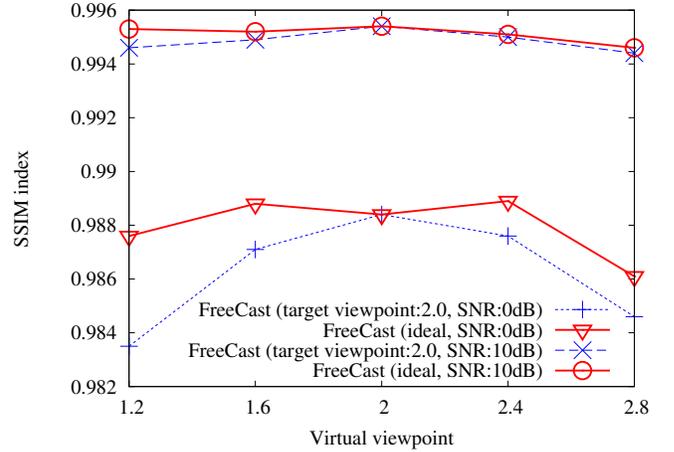


Fig. 11. SSIM vs. positions with different power assignments at channel symbol rate of 37.5 Msymbols/s and wireless channel SNRs of 0 and 10 dB.

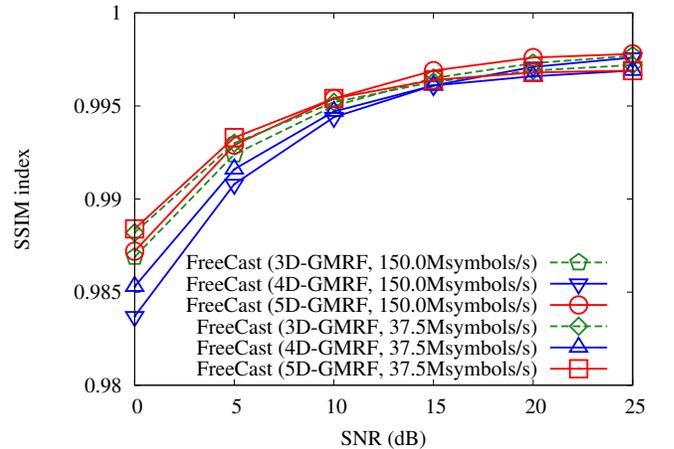


Fig. 12. SSIM vs. SNR with different GMRF models at the center virtual viewpoint and channel symbol rates of 150.0 and 37.5 Msymbols/s.

notifies the corresponding fitting parameters to the receiver. 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 two schemes at wireless channel SNRs of 0 and 10 dB: ideal and fixed power assignment for the center virtual viewpoint. The ideal scheme represents the case when the receiver's feedback is omniscient. The fixed allocation scheme assigns the fixed transmission power to texture and depth frames to achieve the best video quality at the center virtual viewpoint.

Fig. 11 shows the SSIM performance of each scheme as a function of virtual viewpoint positions at channel symbol rate of 37.5 Msymbols/s and wireless channel SNRs of 0 and 10 dB. This figure reveals the following two observations:

- As the distance between target viewpoint and preferred viewpoint increases, the performance gap between ideal and fixed allocation schemes becomes larger up to 0.004.
- At a high channel SNR, the performance gap between

ideal and fixed allocation schemes becomes marginal even when the distance is long. It means we can expect FreeCast keeps better performance even if the feedback information is delayed.

F. Discussion on Multi-Dimensional GMRF

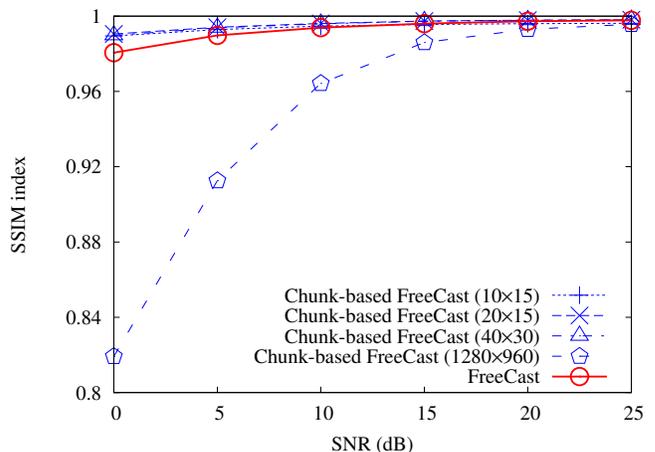
FreeCast models MVD video signals using 5D-GMRF as shown in Fig. 4 to reduce the amount of overhead. In this section, we compare the video quality of 5D-GMRF with that of 3D/4D-GMRF schemes to evaluate the effect of different multidimensional GMRF models. 3D-GMRF scheme models each viewpoint (e.g, left and right) of texture/depth (i.e., Y, U, V, and depth) frames independently using first-order 3D-GMRF for the use of 3D-DCT. 4D-GMRF scheme independently models the texture and depth frames using first-order 4D-GMRF to make use of 4D-DCT operations for each texture and depth frame. 3D-GMRF and 4D-GMRF schemes send 42 and 26 parameters as metadata, respectively, whereas 5D-GMRF requires 9 metadata. For example, the amount of overhead in the 3D-GMRF scheme will be $4.5 \cdot 10^{-6}$ bits/pixels on average at video sequences of *balloons* and *kendo*.

Fig. 12 plots the average SSIM performance at the center virtual viewpoint as a function of wireless channel SNRs. It is observed that FreeCast based on 5D-GMRF outperforms the 4D-GMRF scheme for the whole SNR regimes. Especially, FreeCast yields better performance at low SNR regimes. This is because 5D-DCT has strong energy compaction property by exploiting correlations between texture and depth frames. For example, FreeCast with 5D-GMRF improves the SSIM performance by 0.001 compared to the 4D-GMRF scheme across channel SNR regimes of 0 to 25 dB at a channel symbol rate of 37.5 Msymbols/s. Although 3D-GMRF performs comparable to 5D-GMRF, the amount of compressed metadata can be significantly reduced by nearly eight-fold with 5D-GMRF. Note that the higher-dimension DCT transform for energy compaction may potentially result in higher peak-to-average power ratio (PAPR), which can be one of drawbacks when considering practical power amplifier hardware for energy-efficient wireless systems.

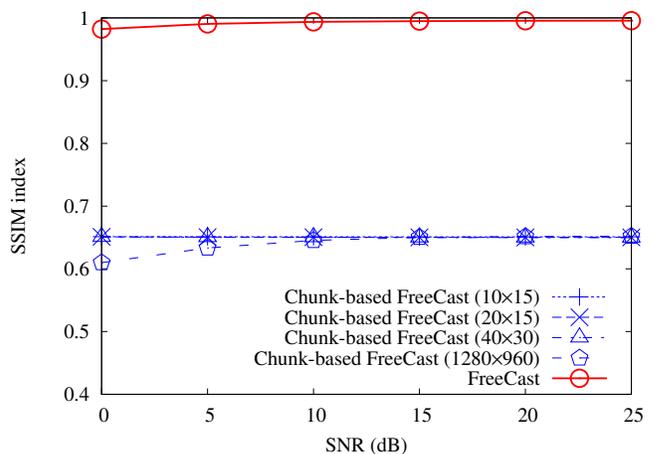
G. Discussion on High-Resolution MVD Videos

Previous sections used low-resolution MVD videos, i.e., 1024×768 pixels, to demonstrate an impact of the proposed scheme. Here, we discuss the effect of our FreeCast on high-resolution MVD videos. Figs. 13(a) and (b) show the average SSIM performance of FreeCast and the chunk-based FreeCast schemes using two video sequences, namely, *champagne_tower* and *pantomime*, with a resolution of 1280×960 pixels at channel symbol rates of 150 and 37.5 Msymbols/sec. In this case, we use four different chunk sizes, i.e., 10×15 , 20×15 , 40×30 , and 1280×960 pixels, in the chunk-based FreeCast schemes.

At a high channel symbol rate in Fig. 13 (a), we can see that FreeCast achieves better video quality compared to the chunk-based FreeCast with a chunk size of 1280×960 pixels irrespective of wireless channel SNRs. On the other hand,



(a) SSIM performance at a channel symbol rate of 150.0 Msymbols/s



(b) SSIM performance at a channel symbol rate of 37.5 Msymbols/s

Fig. 13. Average SSIM performance as a function of channel SNRs in video sequences of *champagne_tower* and *pantomime*.

SSIM performance in FreeCast is lower than that of chunk-based FreeCast with smaller chunk sizes in low channel SNR regimes. This is because estimation errors in the proposed fitting function become bigger in high-resolution MVD videos.

At a low channel symbol rate in Fig. 13 (b), it is demonstrated that FreeCast still achieves graceful quality improvement. On the other hand, SSIM performance of the chunk-based FreeCast schemes is significantly low due to a large overhead. For example, FreeCast achieves SSIM improvement by 0.341, 0.342, 0.342, and 0.352 over the chunk-based FreeCast schemes with the chunk size of 10×15 , 20×15 , 40×30 , and 1280×960 pixels, respectively, across channel SNRs of 0 dB to 25 dB.

V. CONCLUSION

This paper proposed FreeCast, which is a novel graceful video delivery scheme for wireless free viewpoint video streaming. FreeCast can ensure that video quality at any virtual viewpoint is proportional to the wireless channel quality. In addition, the proposed fitting function can precisely estimate the power of 5D-DCT coefficients. Evaluations demonstrated

that the proposed fitting function in FreeCast can significantly reduce the amount of overhead and it brings better video quality compared to the existing digital-based MVD streaming and chunk-based graceful MVD streaming scheme even in narrowband environments. While we showed a great potential of FreeCast using several test MVD video sequences, more rigorous analyses should follow to validate the performance over many different MVD video types.

In this paper, we assumed no strong interference which may cause severe packet loss. In order to improve loss resilience, the proposed FreeCast may be extended to adopt compressive sensing techniques [31]–[34] as a future work. It is also known that the analog-based transmission scheme has a limitation in its efficiency of compression compared to digital-based video encoding. In order to exploit both the benefit of digital and analog video streaming, an extension towards hybrid methods [35] is highly anticipated. When we integrate digital video encoding into FreeCast, how to realize the best video quality by optimizing both digital and analog video transmissions is an open issue.

ACKNOWLEDGMENT

T. Fujihashi's work was partly supported by JSPS KAKENHI Grant Number 17K12672.

REFERENCES

- [1] Z. Chen, X. Zhang, Y. Xu, J. Xiong, Y. Zhu, and X. Wang, "MuVi: Multiview video aware transmission over mimo wireless systems," *IEEE Transactions on Multimedia*, vol. 19, no. 12, pp. 2788–2803, Jun. 2017.
- [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, p. 93930G, Mar. 2015.
- [3] O. Stankiewicz, M. Domanski, A. Dziembowski, A. Grzelka, D. Mieloch, and J. Samelak, "A free-viewpoint television system for horizontal virtual navigation," *IEEE Transactions on Multimedia*, vol. 20, no. 9, pp. 2182–2195, Aug. 2018.
- [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, May 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*, vol. 5291, May 2004, pp. 93–105.
- [6] S. Li, C. Zhu, and M.-T. Sun, "Hole filling with multiple reference views in DIBR view synthesis," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 1948–1959, Jan. 2018.
- [7] 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, Apr. 2015.
- [8] Y. Chen and S. Yen, "3D-AVC draft text 6, document JCT3V-D1002.doc," JCT-3V, Apr. 2013.
- [9] 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, Jul. 2015.
- [10] E. Ekmekcioglu, C. G. Gurler, A. Kondo, and A. M. Tekalp, "Adaptive multiview video delivery using hybrid networking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 6, pp. 1313–1325, Feb. 2017.
- [11] 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, Jun. 2015.
- [12] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3366–3378, Sep. 2013.
- [13] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Soft video delivery for free viewpoint video," in *IEEE International Conference on Communications*, Paris, France, May 2017, pp. 1–7.
- [14] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *ACM Annual International Conference on Mobile Computing and Networking*, Las Vegas, NV, Sep. 2011, pp. 289–300.
- [15] 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, Feb. 2018.
- [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, Nov. 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, May 2016.
- [18] 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*, Cancun, Mexico, Nov. 2015, pp. 327–336.
- [19] G. Wang, K. Wu, Q. Zhang, and L. M. Ni, "SimCast: Efficient video delivery in MU-MIMO WLANs," in *IEEE Conference on Computer Communications*, Toronto, ON, Apr.–May 2014, pp. 2454–2462.
- [20] 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, Apr. 2016.
- [21] 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*, vol. 25, no. 11, pp. 1801–1814, Nov. 2015.
- [22] D. He, C. Lan, C. Luo, E. Chen, F. Wu, and W. Zeng, "Progressive pseudo-analog transmission for mobile video streaming," *IEEE Transactions on Multimedia*, vol. 19, no. 8, pp. 1894–1907, 2017.
- [23] B. Tan, J. Wu, Y. Li, H. Cui, W. Yu, and C. W. Chen, "Analog coded SoftCast: A network slice design for multimedia broadcast/multicast," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2293–2306, Oct. 2017.
- [24] X. Song, X. Peng, J. Xu, G. Shi, and F. Wu, "Distributed compressive sensing for cloud-based wireless image transmission," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1351–1364, Jun. 2017.
- [25] J. Lei, J. Duan, F. Wu, N. Ling, and C. Hou, "Fast mode decision based on grayscale similarity and inter-view correlation for depth map coding in 3D-HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 3, pp. 706–718, Mar. 2016.
- [26] F. Shao, W. Lin, G. Jiang, and M. Yu, "Low-complexity depth coding by depth sensitivity aware rate-distortion optimization," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 94–102, Mar. 2016.
- [27] A. Vetro, T. Ebrahimi, and V. Baroncini, "3D video subjective quality assessment test plan," Doc. JCT3V-F1011, Nov. 2013.
- [28] Fujii Laboratory at Nagoya University. [Online]. Available: <http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/>
- [29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [30] [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-13.0/
- [31] A. Wang, B. Zeng, and H. Chen, "Wireless multicasting of video signals based on distributed compressed sensing," *Signal Processing: Image Communication*, vol. 29, no. 5, pp. 599–606, May 2014.
- [32] X. L. Liu, W. Hu, C. Luo, and F. Wu, "Compressive image broadcasting in MIMO systems with receiver antenna heterogeneity," *Signal Processing: Image Communication*, vol. 29, no. 3, pp. 361–374, Mar. 2014.
- [33] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Compressive sensing for loss-resilient hybrid wireless video transmission," in *IEEE Globecom*, San Diego, CA, Dec. 2015, pp. 1–5.
- [34] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [35] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 2, pp. 331–345, Feb. 2014.