MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

High-Quality Soft Video Delivery with GMRF-Based Overhead Reduction

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

TR2017-119 August 2017

Abstract

Soft video delivery, i.e., analog video transmission, has been proposed to provide high video quality in unstable wireless channels. However, existing analog schemes need to transmit a significant amount of metadata to a receiver for power allocation and decoding operations causing large overhead and quality degradation due to rate and power losses. To reduce the overhead while keeping the video quality high, we propose a new analog transmission scheme. Our scheme exploits a Gaussian Markov random field (GMRF) for modeling video sequences to significantly reduce the required amount of metadata, which are obtained by fitting into the Lorentzian function. Our scheme achieves not only reduced overhead but also improved video quality, by using the fitting function and parameters for metadata. Evaluations using several test video sequences demonstrate that the proposed scheme reduces overhead by 99.7 % with 1.2 dB improvement of video quality (in terms of peak signal-to-noise ratio) compared to the existing analog video transmission scheme. We also investigate the impact of bandwidth limitation, showing a significant gain up to 2.7 dB for narrow-band systems.

IEEE Transactions on Multimedia

High-Quality Soft Video Delivery with GMRF-Based Overhead Reduction

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

***Abstract*—Soft video delivery, i.e., analog video transmission, has been proposed to provide high video quality in unstable wireless channels. However, existing analog schemes need to transmit a significant amount of metadata to a receiver for power allocation and decoding operations causing large overhead and quality degradation due to rate and power losses. To reduce the overhead while keeping the video quality high, we propose a new analog transmission scheme. Our scheme exploits a Gaussian Markov random field (GMRF) for modeling video sequences to significantly reduce the required amount of metadata, which are obtained by fitting into the Lorentzian function. Our scheme achieves not only reduced overhead but also improved video quality, by using the fitting function and parameters for metadata. Evaluations using several test video sequences demonstrate that the proposed scheme reduces overhead by 99.7 % with 1.2 dB improvement of video quality (in terms of peak signal-to-noise ratio) compared to the existing analog video transmission scheme. We also investigate the impact of bandwidth limitation, showing a significant gain up to 2.7 dB for narrow-band systems.**

***Index Terms*—Soft Video Delivery, Gaussian Markov Random Field, Overhead Reduction**

I. Introduction

Video delivery is one of the major applications in the wireless environment – according to Cisco visual networking index studies, three-fourths of the world's mobile data traffic will be video contents by 2020 [1]. In conventional video streaming, the digital video compression and digital wireless transmission are carried out in sequence [2]–[4]. For example, the video compression part uses H.264/Advanced Video Coding (AVC) [5] or H.265/High-Efficiency Video Coding (HEVC) [6] standards to generate a compressed bit stream using quantization and entropy coding. The wireless transmission part uses channel coding and a digital modulation scheme to reliably transmit the encoded bit stream.

However, the conventional scheme has the following problems due to the unreliable wireless channel. First, the encoded bit stream is highly vulnerable to bit errors. When the channel's signal-to-noise ratio (SNR) falls under a certain threshold, the video quality drops significantly. This phenomenon is referred to as the cliff effect. Second, the video quality does not gracefully improve even when the wireless channel quality is improved. Finally, quantization is a lossy process, whose distortion cannot be recovered at the receiver. Some studies [7], [8] have been proposed to mitigate the cliff effect in the digital video transmission by introducing layered source coding and layered channel coding. However, in these studies, the cliff effect is converted into the so-called staircase effect [9]. In the staircase effect, the video quality discontinuously improves as the wireless channel quality improves.

To overcome the above-mentioned problems, analog transmission schemes [10]–[19] have been proposed. For example, SoftCast [10] directly transmits linearly-transformed video signals over a lossy channel and allocates power to the signals to maximize video quality, instead of using digital video compression and digital modulation. In contrast to the conventional digital scheme, the video quality of SoftCast can be gracefully improved according to the wireless channel quality.

However, the performance of SoftCast depends strongly on the chunk size. In SoftCast, a sender allocates transmission power to the video signals such that the receiver noise can be minimized. The power allocation is based on the power of each linearly-transformed video signal. Hence, the sender needs to transmit the power information of all the video signals without errors to decode the signals at the receiver. The transmission of this metadata causes large overhead, resulting in video quality degradation due to power and rate loss. To reduce metadata overhead, SoftCast therefore divides the signals into multiple chunks and transmits a smaller number of metadata corresponding to each chunk. In turn, the chunk division may degrade performance due to improper power allocation, in particular when a large chunk size is used for lower overhead.

To improve performance, some analog schemes adopted coset coding [11]–[13], motion-compensated temporal filtering [14], compressive sensing [15], [16], and subcarrier assignment [17]. However, all these methods do not consider the effect of chunk size. Although the trade-off between chunk size and video quality were discussed in [20], proposals to reduce the overhead were beyond the scope of the paper.

In this paper, we propose a new analog scheme without chunk division to overcome the issues of conventional analog schemes. To obtain the power values of linear-transformed video signals without transmitting large-overhead metadata, our scheme uses a Gaussian Markov random field (GMRF) [21], [22] to model video signals and exploits a Lorentzian-based fitting function at the sender and the receiver.

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 November 9, 2016.

Specifically, the sender finds a few parameters for the fitting function from video sequences and sends the parameters as metadata to the receiver. The receiver obtains the power values from the fitting function. Evaluations using test video sequences show that the proposed scheme improves video quality by 1.2 dB with 99.7 % reduction in the overhead.

Our contribution is two-fold: 1) we verify that the power of the linear-transformed video signals are well fit by a Lorentzian-based function when the video signals can be modeled using GMRFs and 2) we propose fitting-based power allocation and signal reconstruction to achieve improved video quality and reduced overhead simultaneously.

In [23], we have reported a preliminary analysis of the GMRF-based analog scheme for overhead reduction and high video quality. In this paper, we first extend the method reported in [23] to further reduced overhead for metadata transmissions. We demonstrate the overhead of our scheme can be further reduced by 52 % compared to that of the proposed scheme in [23]. In addition, we enhance the performance evaluations to demonstrate the advantage of the proposed scheme in the presence of bandwidth limitation. We use Huffman coding for metadata compression to evaluate how much metadata is required to send in a realistic analog system. We then consider a bandwidth constraint for the proposed and conventional analog schemes to evaluate the impact of overhead reduction on video quality. We also consider structural similarity (SSIM) [24] besides conventional peak signal-to-noise ratio (PSNR) to evaluate video quality. From the evaluations, it is verified that our scheme outperforms conventional analog schemes with arbitrary chunk size in both broad- and narrow-band environments.

II. Soft Video Delivery

The purposes of our proposed scheme are 1) to achieve video quality that gracefully improves according to the wireless channel quality and 2) to reduce the amount of metadata. Fig. 1 shows the schematic of our proposed scheme. The encoder first performs a three dimensional (3D) discrete cosine transform (DCT) operation on the original video frames. According to the power of the DCT coefficients, we find the best parameters of a fitting function based on GMRF model. The DCT coefficients are then scaled and analog-modulated according to these fitting parameters. Finally, the encoder sends the analog-modulated symbols and the fitting parameters 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 based on the received fitting parameters. The DCT coefficients are obtained from the received analog-modulated symbols through the use of MMSE filtering.

A. Encoder

The encoder first preforms 3D-DCT operation on the original sequence to obtain the DCT coefficients. 3D-DCT is used for whole frames in one group of pictures (GoP), which is a sequence of successive video frames. The DCT coefficients are mapped to I (in-phase) and Q (quadrature) components after the following power allocation.

Let x_i denote the ith analog-modulated symbol. Each analog-modulated symbol is scaled by g_i for noise reduction:

$$x_i = g_i \cdot s_i. \tag{1}$$

Here, s_i is the ith DCT coefficient and g_i is the scale factor which determines the coefficient's power allocation. The transmitter performs optimal power control by selecting g_i to achieve the highest video quality. Specifically, the best g_i is obtained by minimizing the mean-square error (MSE) under the power constraint with total power budget P as follows:

$$\min \quad \mathsf{MSE} = \mathbb{E}\left[(x_i - \hat{x}_i)^2\right] = \sum_i^N \frac{\sigma^2 \lambda_i}{g_i^2 \lambda_i + \sigma^2}, \tag{2}$$

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

where $\mathbb{E}[\cdot]$ denotes expectation, $\hat{x}_i$ is an estimate of the transmitted symbol, λ_i is the power of ith DCT coefficient, N is the number of DCT coefficients, and σ^2 is a receiver noise variance. The near-optimal solution is expressed as

$$g_i = \lambda_i^{-1/4} \sqrt{\frac{P}{\sum_j \sqrt{\lambda_j}}}. \tag{4}$$

B. Decoder

After transmission over the wireless channel, each symbol at the receiver can be modeled as follows:

$$y_i = x_i + n_i, \tag{5}$$

where y_i is the ith received symbol and n_i is an effective noise having a variance of σ^2. The receiver extracts DCT coefficients from I and Q components, and reconstructs the coefficients using MMSE filter [25] as follows:

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

We note that the loss-free channel model can be readily extended to the lossy wireless channel model, where the receiver can fail the detection of DCT coefficients due to intereference. For such a case, the lost coefficients are regarded as zeros, as in [26]. Detail analysis for lossy channels will be left as future works, and we focus on loss-free AWGN channels in this paper. The decoder then obtains corresponding video sequence by taking the inverse 3D-DCT for the filter output $\hat{s}_i$.

C. Overhead Reduction

In order for the receiver to carry out MMSE filtering in (6), the sender needs to transmit λ_i of all coefficients without errors as metadata, which may constitute a large overhead. For example, when the sender transmits eight video frames with the resolution of 352×288, the sender needs to transmit metadata for all DCT coefficients, i.e., $352 \times 288 \times 8 = 811{,}008$ variables in total, to the receiver – a total of 5.8 bits/pixel after Huffman coding. This overhead induces performance

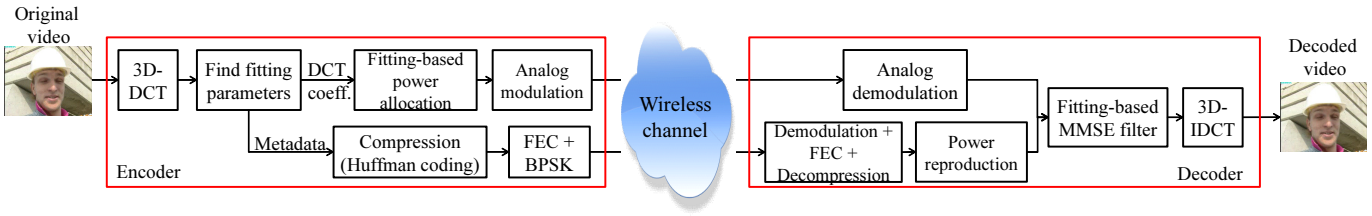


Fig. 1. Proposed analog video transmission scheme employing GMRF-based power allocation for improved quality and reduced overhead.

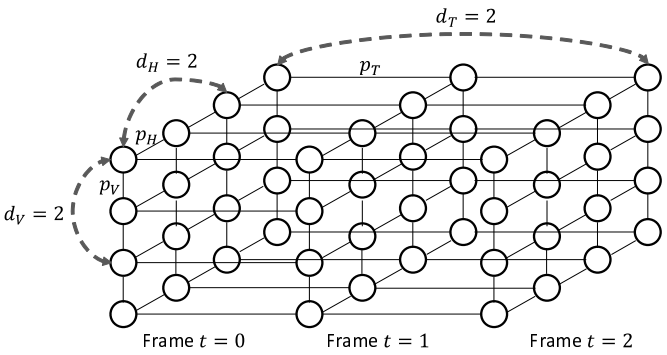


Fig. 2. First-order GMRF model for video signals.

degradation due to rate and power losses in transmission of analog-modulated symbols. To reduce the overhead, conventional methods divide the DCT coefficients into chunks and carry out power allocation and MMSE filter for each chunk. However, overhead is still high and the chunk division causes performance degradation due to a loss of optimality for power allocation with respect to (6). When the chunk size is 44×36 pixels, 512 variables of metadata are still required every eight frames. Although the amount of metadata is reduced to approximately $4.7 \cdot 10^{-3}$ bits/pixel in this large chunk size, the video quality can be significantly degraded.

In order to reduce overhead while keeping video quality high, we use a fitting function to approximate the power values λ_i for a variety of video sequences. To this end, we use a GMRF to model video signals. Based on the model, we verify λ_i, except direct current (DC) component, can be fit by a Lorentzian function with four parameters. The details of the derivation are described in Sec. II-D. Our scheme uses $\hat{\lambda}_i$, an estimate of the power of DCT coefficients, obtained from the fitting function, for the power allocation and MMSE filter. To share the same $\hat{\lambda}_i$ at both the sender and the receiver, our scheme just needs to transmit five metadata variables (i.e., four fitting parameters and DC component of DCT coefficients), which can be also compressed by Huffman coding. Here, the amount of metadata is approximately $1.5 \cdot 10^{-5}$ bits/pixel. We assume that the encoder uses $1/2$-rate convolutional coding and binary phase-shift keying (BPSK) for the compressed metadata transmissions.

D. GMRF-Based Fitting Function

We use a simple first-order GMRF to model video signals as shown in Fig. 2. In the video signals, each pixel is connected to three neighboring pixels, in each of the horizontal, vertical, and time directions. Each direction has different correlations, which are defined as p_{H}, p_{V}, and p_{T}, respectively. Note that the correlation between any two pixels can be described as $p_{\mathrm{H}}^{d_{\mathrm{H}}} \cdot p_{\mathrm{V}}^{d_{\mathrm{V}}} \cdot p_{\mathrm{T}}^{d_{\mathrm{T}}}$, where d_{H}, d_{V}, and d_{T} are horizontal, vertical, and time distances between the pixels, respectively.

The DCT can be regarded as a discrete-time real-valued version of the Fourier transform. For our case, the Fourier transform of the video signal's auto-correlation function represents the power spectral density by the Wiener–Khintchine theorem, assuming that the video source is wide-sense stationary. For 3D video signals following the GMRF, the power spectrum density of 3D-DCT coefficients can be asymptotically obtained by the Lorentzian function as follows:

$$F(i,j,k) = \beta' \cdot \frac{1}{1+f_1^2(i)} \cdot \frac{1}{1+f_2^2(j)} \cdot \frac{1}{1+f_3^2(k)}, \tag{7}$$

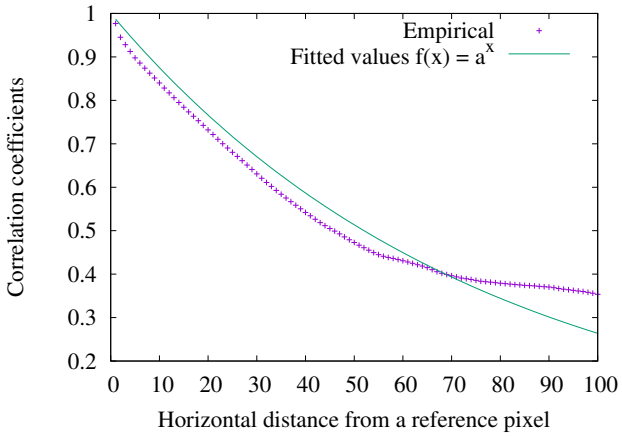
$$\beta' = \frac{\beta}{|\log(p_{\mathrm{H}}) \cdot \log(p_{\mathrm{V}}) \cdot \log(p_{\mathrm{T}})|}, \; f_1(i) = \frac{\alpha_1}{|\log(p_{\mathrm{H}})|}\frac{\pi i}{N_{\mathrm{H}}},$$
$$f_2(j) = \frac{\alpha_2}{|\log(p_{\mathrm{V}})|}\frac{\pi j}{N_{\mathrm{V}}}, \; f_3(k) = \frac{\alpha_3}{|\log(p_{\mathrm{T}})|}\frac{\pi k}{N_{\mathrm{T}}}, \tag{8}$$

where N_{H}, N_{V}, and N_{T} are the number of coefficients in horizontal, vertical, and time domains, respectively. Here, α_k and β are parameters for fitting. Note that above equations express the power spectrum density of the DCT coefficients except the DC component. Our scheme ignores the DC component from fitting operation because the DC component cannot be modeled by the Lorentzian function.

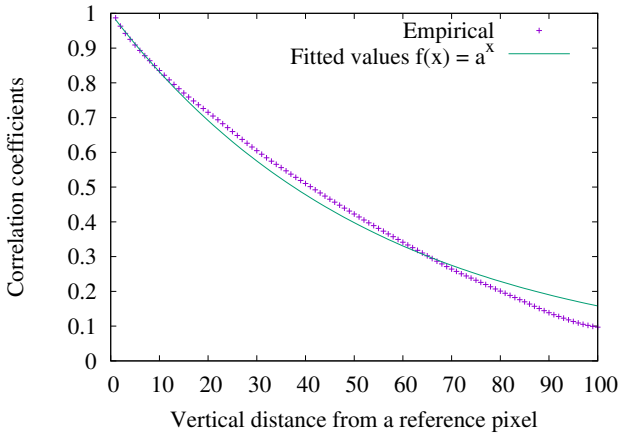
In [23], we considered all eight parameters (p_{H}, p_{V}, p_{T}, α_1, α_2, α_3, β, and DC component) as metadata. In this paper, to further reduce the overhead without any penalty, the sender transmits five fitting parameters of $\alpha_1' = \alpha_1/|\log(p_{\mathrm{H}})|$, $\alpha_2' = \alpha_2/|\log(p_{\mathrm{V}})|$, $\alpha_3' = \alpha_3/|\log(p_{\mathrm{T}})|$, β', and DC component, as the metadata. In addition, we employ Huffman coding to compress the fitting parameters before transmission. In consequence, the proposed scheme can reduce overhead from $2.9 \cdot 10^{-5}$ to $1.5 \cdot 10^{-5}$ bits/pixel. This overhead is enough small compared to an acceptable overhead in real-life video streaming services [27], i.e., around 0.1 bits/pixel.

E. Correlation Coefficient

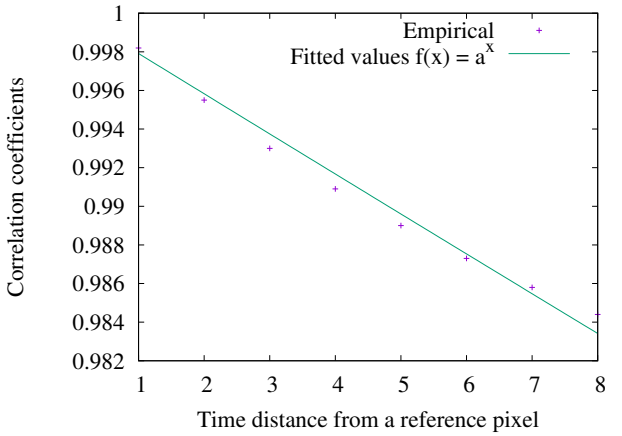
To calculate the fitting function, the encoder estimates the horizontal, vertical, and time correlations of the video



(a) Horizontal correlation

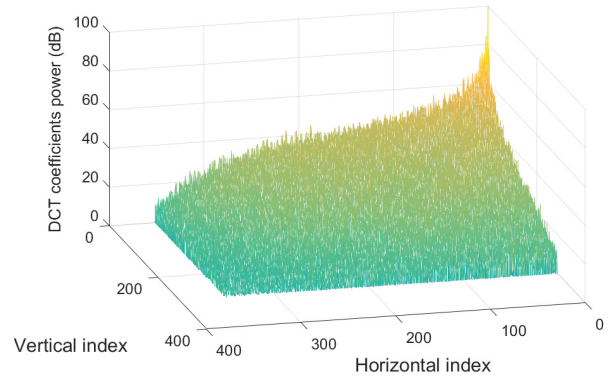


(b) Vertical correlation

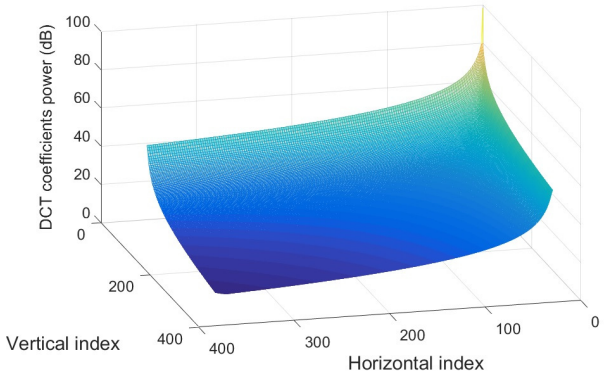


(c) Time correlation

Fig. 3. Fitting correlation coefficients of *akiyo*.



(a) Empirical



(b) Fitting

Fig. 4. Power values of DCT coefficients for *akiyo*: (a) empirical, and (b) fitting. Here, NMSE between empirical and fitting values is lower than -22 dB.

sequence, by fitting an empirical auto-correlation function to an exponential function of the form $f(x) = p^x$ with p being a correlation factor to be estimated by means of least-squares methods. Fig. 3 shows an example of fitting curves for horizontal, vertical, and time correlation coefficients for the video sequence *akiyo*. Here, we obtain the estimated parameters p_{H}, p_{V}, and p_{T} of 0.98, 0.98, and 0.99, respectively. From this figure, it is expected that the simple GMRF model depicted in Fig. 2 can capture some useful statistics of real video sequences. We note that the prediction error of autocorrelation between real and fitted values can be reduced by using higher-order Markov model. Detailed analysis with more complicated model will be left as future work.

With the estimated correlations, the encoder finds the other fitting parameters based on the empirical power of the non-DC components by least-squares fitting. Note that the computational complexity of the fitting function is the same order as that of calculating mean and variance in each chunk for conventional SoftCast schemes. The encoder then reproduces the power of DCT coefficients using the estimated parameters and fitting function. Fig. 4 shows the empirical and fitting power of DCT coefficients within one video frame for the video sequence of *akiyo*. It was 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 -27.1 dB on average across test video sequences of *akiyo*, *foreman*, *mobile*, *coastguard*, and *news* [28]. The proposed scheme can significantly reduce the overhead by transmitting just five values regardless of the video size.

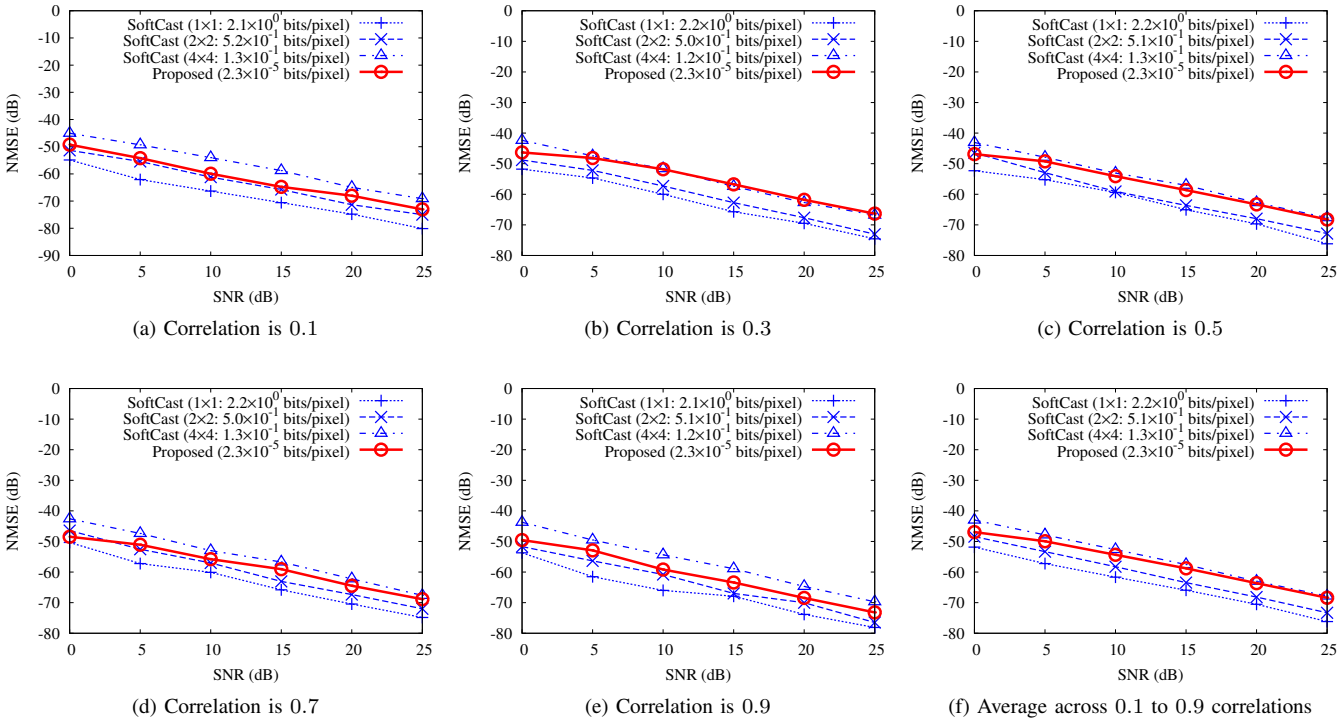


Fig. 5. NMSE vs. SNR for synthetic video sequence generated by GMRF: (a) 0.1 correlation, (b) 0.3 correlation, (c) 0.5 correlation, (d) 0.7 correlation, (e) 0.9 correlation, and (e) average across 0.1 to 0.9 correlations.

III. Performance Evaluation

A. Simulation Settings

Metric: We evaluate the performance of reference schemes in terms of the NMSE, PSNR, and SSIM [24]. NMSE and PSNR are defined as follows:

$$\mathsf{NMSE} = 10\log_{10}\frac{\varepsilon_{\mathrm{MSE}}}{\sum_i^N s_i^2}, \tag{9}$$

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

where L is the number of bits used to encode pixel luminance (typically eight bits), and $\varepsilon_{\mathrm{MSE}}$ is the MSE between all pixels of the decoded and the original video. SSIM can predict the perceived quality of video streaming. Larger values of SSIM close to 1 indicates higher perceptual similarity between original and decoded images. We obtain the average NMSE, PSNR, and SSIM across the entire video sequence.

Test Video: We use standard reference video, namely, *foreman*, *akiyo*, *mobile*, *coastguard*, *news*, *crew*, *football*, *bus*, *container*, *flower*, *stefan*, *silent*, *tempete*, *waterfall*, *bridge-close*, *bridge-far*, *paris*, and *highway* in the CIF format (352×288 pixels, 30 frames per second) from the Xiph collection [28]. In addition, we also use high resolution videos, namely, *Johnny* and *KristenAndSara*, in the HD format (1280×720 pixels, 60 frames per second) to discuss an effect of the proposed fitting function in high resolution videos. Here, we set the size of each GoP to eight frames.

Amount of Metadata: As we mentioned in Sec. II-C, the proposed scheme sends five metadata variables for one GoP sequence. Conventional schemes of soft video delivery transmit mean and variance as metadata variables for each chunk. Both methods use Huffman coding to obtain the size of compressed metadata in one GoP, and then divides the size of compressed metadata by the number of pixels in one GoP.

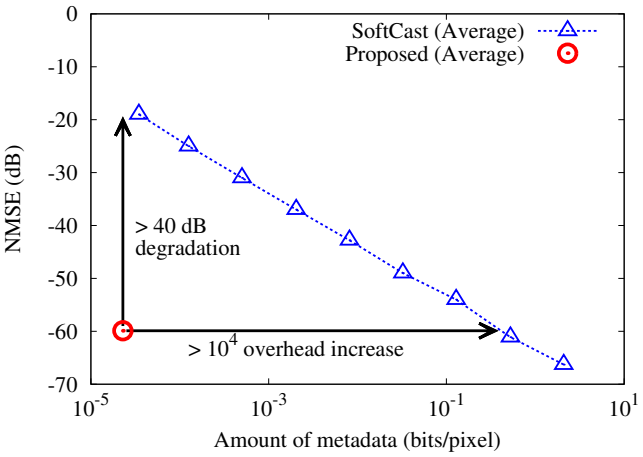


Fig. 6. Average NMSE (across correlations between 0.1 and 0.9) vs. amount of metadata (bits/pixel) at a channel SNR of 10 dB.

B. Synthetic Video Signals from GMRF

Before analyzing real video sequences, we first evaluate our proposed scheme for virtual/synthetic video sequences generated from GMRF model. We assume that the resolution of the signals is $256 \times 256 \times 8$ and the correlations of

three domains (horizontal, vertical, and time) are identical. We set the mean and variance of the signals are 128 and 1, respectively. For the comparison, we measure NMSE of the proposed and three SoftCast schemes with different chunk sizes: 1×1, 2×2, and 4×4 pixels. The corresponding amount of metadata in SoftCast becomes $2.2\cdot10^{0}$, $5.1\cdot10^{-1}$, and $1.3\cdot10^{-1}$ bits/pixel on average, respectively. On the other hand, the amount of metadata in the proposed scheme is $2.3\cdot10^{-5}$ bits/pixel. Fig. 5 shows the NMSE with the different correlations: (a) 0.1, (b) 0.3, (c) 0.5, (d) 0.7, (e) 0.9, and (f) average across 0.1 to 0.9. From these figures, we observe the following two points:

- NMSE of the proposed scheme is lower than SoftCast with a chunk size of 4×4 pixels irrespective of correlations. It is verified that the proposed scheme can reduce the amount of metadata by 99.98% while keeping video quality higher compared to SoftCast.
- At a correlation of 0.9, NMSE of the proposed scheme approaches that of SoftCast with a chunk size of 2×2 pixels. Thus, it will bring high performance in video delivery since video signals have a high correlation in horizontal, vertical, and time domains.
- SoftCast with a smallest chunk size of 1×1 pixels, which is an idealistic case, achieves the lowest NMSE. However, the amount of metadata is $9.2\cdot10^{4}$ times larger than the proposed scheme. A large overhead will cause power and rate losses on transmissions of analog-modulated symbols.

Fig. 6 shows the average NMSE across correlations of 0.1 to 0.9 with the different chunk sizes at an SNR of 10 dB. Here, we evaluate NMSE of SoftCast with nine chunk sizes: 1×1, 2×2, 4×4, 8×8, 16×16, 32×32, 64×64, 128×128, and 256×256 pixels. The corresponding amount of metadata is $2.2\cdot10^{0}$, $5.1\cdot10^{-1}$, $1.3\cdot10^{-1}$, $3.2\cdot10^{-2}$, $8.0\cdot10^{-3}$, $2.0\cdot10^{-3}$, $5.0\cdot10^{-4}$, $1.3\cdot10^{-4}$, and $3.4\cdot10^{-5}$ bits/pixel. This figure demonstrates that the proposed scheme significantly outperforms the SoftCast in both video quality and overhead. For example, the proposed scheme improves NMSE approximately by 40.9 dB compared to SoftCast with a chunk size of 256×256 pixels for a comparable overhead. In addition, we can see that the proposed scheme can still yield better performance over SoftCast with a large overhead of 10^{-1} bits/pixel since the power values of DCT coefficients can be estimated by fitting the Lorentzian function with high accuracy.

C. Real Video Sequences

The previous section demonstrated that the proposed scheme approaches the performance of SoftCast with large overhead and small chunk sizes when video signals are generated from GMRF. However, real video sequences may not follow the model and this model mismatch induces estimation errors. To evaluate the effect on real video sequences, this section uses the 18 test sequences listed in Table I. These video sequences are also used to discuss the performance of the proposed scheme in various different conditions having high and low motion videos. In particular, the videos of *bus* and *football* contain high-motion objects. Table I lists the values of fitting parameters of each video sequence in the first GoP. It is interesting to note that the correlation factors p_{H}, p_{V}, and p_{T} are nearly identical. If we exploit this observation by using an averaged common value for α_1, α_2, and α_3, we can further reduce the metadata from five values to three values with a slight performance penalty (approximately 1 dB loss). We note that SoftCast schemes in our evaluations do not use the Hadamard transform, which was originally used in [10] to protect chunks against packet loss. In our evaluations, we consider the case when the wireless channel is loss-free but noisy, and thus the SoftCast with and without the Hadamard transform perform almost identically.

TABLE I
PARAMETERS FOR FITTING FUNCTION

Video Sequence	p_{H}	p_{V}	p_{T}	α_1	α_2	α_3	β
Akiyo	0.98	0.98	0.99	0.06	1.75	0.06	86.60
Foreman	0.98	0.97	0.96	0.08	1.69	0.07	194.56
Mobile	0.97	0.94	0.93	0.10	0.15	0.04	4.98
Coastguard	0.99	0.98	0.96	0.02	1.28	0.04	18.21
News	0.97	0.97	0.99	0.17	1.71	0.07	1756.81
Crew	0.95	0.99	0.96	0.13	0.08	0.15	2.34
Football	0.95	0.93	0.86	0.23	0.12	0.16	27.30
Bus	0.98	0.89	0.85	0.36	0.08	0.06	7.39
Container	0.99	0.95	0.99	0.10	1.80	0.01	142.16
Flower	0.98	0.98	0.93	0.01	0.05	0.03	0.12
Stefan	0.98	0.93	0.85	0.17	0.05	0.07	3.45
Silent	0.99	0.98	0.99	0.03	1.78	0.02	36.00
Tempete	0.97	0.92	0.96	0.17	1.52	0.05	498.09
Waterfall	0.99	0.98	0.99	0.03	1.30	0.02	31.59
Bridge-close	0.99	0.97	0.99	0.01	1.60	0.01	20.31
Bridge-far	0.99	0.98	0.99	0.01	137.39	0.01	32274.80
Paris	0.97	0.94	0.98	0.16	1.11	0.03	395.62
Highway	0.99	0.99	0.99	0.01	2.35	0.03	35.26

1) Overhead Reduction: We first evaluate overhead in the proposed and existing schemes in terms of *bits/pixel*. Table II shows the amount of metadata in each chunk size with different video sequences. The results show that irrespective of the test video sequence the proposed scheme requires the least amount of metadata. This reduction saves transmission power and leads to additional quality improvement by allocating the saved power to the transmission of analog-modulated symbols. For example, the proposed scheme reduces the metadata by approximately 99.7 % compared to SoftCast with a default chunk size of 44×36 pixels and 72.4 % compared to SoftCast with a largest chunk size of 352×288 pixels on average across 18 test video sequences.

2) Video Quality: The discussion above revealed that the proposed scheme achieves the smallest overhead compared to SoftCast with arbitrary chunk sizes. This section compares video quality of the proposed and conventional schemes with the different chunk sizes to demonstrate the benefit of the proposed scheme. Fig. 7 shows the average PSNR performance across 18 test video sequences as a function of channel SNR. Here, we select four sizes of chunks for SoftCast: 1×1, 2×2, 44×36, and 352×288 pixels. SoftCast with chunk size of 1×1 represents the ideal case in terms of quality, 2×2 achieves the second highest performance, 44×36 is a default chunk size, and 352×288 has almost the same overhead as the proposed scheme. The key results from this figure are summarized as follows:

TABLE II
AMOUNT OF METADATA IN THE PROPOSED AND CONVENTIONAL SOFTCAST SCHEMES

Video Sequence	Amount of metadata (bits/pixel) at different chunk sizes											
	1×1	2×2	4×4	8×8	16×16	11×9	22×18	44×36	88×72	176×144	352×288	Proposed
Akiyo	3.0	1.3	0.29	$7.0 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$4.5 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$2.5 \cdot 10^{-3}$	$6.2 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Foreman	6.0	2.6	0.65	$1.6 \cdot 10^{-1}$	$3.7 \cdot 10^{-2}$	$1.0 \cdot 10^{-1}$	$2.4 \cdot 10^{-2}$	$5.5 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$2.6 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Mobile	8.8	3.8	0.93	$2.2 \cdot 10^{-1}$	$5.0 \cdot 10^{-2}$	$1.4 \cdot 10^{-1}$	$3.1 \cdot 10^{-2}$	$6.5 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$2.8 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Coastguard	6.2	2.8	0.67	$1.5 \cdot 10^{-1}$	$3.5 \cdot 10^{-2}$	$9.6 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$5.2 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$2.7 \cdot 10^{-4}$	$4.9 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
News	4.9	2.2	0.48	$1.1 \cdot 10^{-1}$	$2.6 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$4.0 \cdot 10^{-3}$	$9.4 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Crew	6.3	2.3	0.56	$1.3 \cdot 10^{-1}$	$3.1 \cdot 10^{-2}$	$8.4 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$4.7 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$2.6 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Football	8.8	3.3	0.78	$1.8 \cdot 10^{-1}$	$3.9 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$2.5 \cdot 10^{-2}$	$5.5 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$2.7 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Bus	9.1	3.4	0.82	$1.9 \cdot 10^{-1}$	$4.2 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$2.6 \cdot 10^{-2}$	$5.9 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$2.8 \cdot 10^{-4}$	$4.9 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Container	5.7	2.1	0.50	$1.1 \cdot 10^{-1}$	$2.6 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$	$9.0 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$	$4.6 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Flower	11.0	3.8	0.92	$2.2 \cdot 10^{-1}$	$4.9 \cdot 10^{-2}$	$1.4 \cdot 10^{-1}$	$3.1 \cdot 10^{-2}$	$6.8 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$2.8 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Stefan	8.7	3.3	0.82	$1.9 \cdot 10^{-1}$	$4.4 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$2.8 \cdot 10^{-2}$	$6.1 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$2.7 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Silent	5.3	1.9	0.44	$1.0 \cdot 10^{-1}$	$2.4 \cdot 10^{-2}$	$6.6 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$3.7 \cdot 10^{-3}$	$8.8 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$4.7 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Tempete	9.1	3.3	0.80	$1.8 \cdot 10^{-1}$	$4.2 \cdot 10^{-2}$	$1.2 \cdot 10^{-1}$	$2.6 \cdot 10^{-2}$	$6.0 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$2.8 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Waterfall	6.3	2.4	0.56	$1.3 \cdot 10^{-1}$	$3.0 \cdot 10^{-2}$	$8.2 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$4.5 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$2.5 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Bridge-close	7.5	2.7	0.63	$1.4 \cdot 10^{-1}$	$3.3 \cdot 10^{-2}$	$9.1 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$	$5.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$2.6 \cdot 10^{-4}$	$5.3 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Bridge-far	6.0	2.1	0.50	$1.1 \cdot 10^{-1}$	$2.7 \cdot 10^{-2}$	$7.1 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$4.1 \cdot 10^{-3}$	$9.8 \cdot 10^{-4}$	$2.3 \cdot 10^{-4}$	$5.1 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Paris	9.0	3.3	0.80	$1.8 \cdot 10^{-1}$	$4.1 \cdot 10^{-2}$	$1.1 \cdot 10^{-1}$	$2.6 \cdot 10^{-2}$	$6.0 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$2.7 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Highway	7.1	2.5	0.61	$1.4 \cdot 10^{-1}$	$3.2 \cdot 10^{-2}$	$8.7 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$	$5.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$2.7 \cdot 10^{-4}$	$5.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$
Average	6.7	2.7	0.64	$1.5 \cdot 10^{-1}$	$3.4 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$4.9 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$2.5 \cdot 10^{-4}$	$5.4 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$

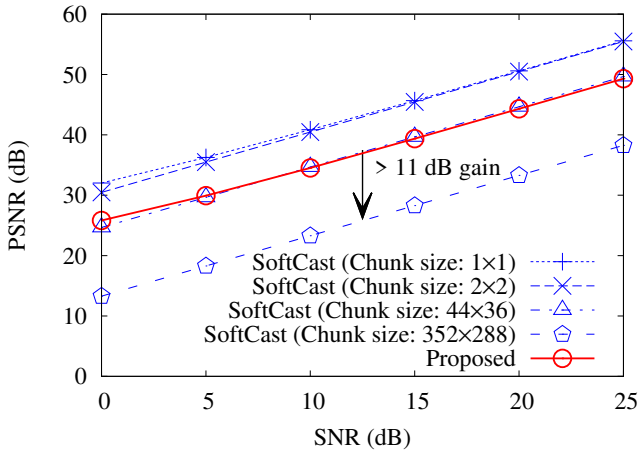


Fig. 7. Average PSNR across 18 test video sequences vs. channel SNR with different chunk size. The corresponding amount of metadata in SoftCast and proposed schemes is $6.7 \cdot 10^{0}$, $2.7 \cdot 10^{0}$, $4.9 \cdot 10^{-3}$, $5.4 \cdot 10^{-5}$, and $1.5 \cdot 10^{-5}$ bits/pixel, respectively.

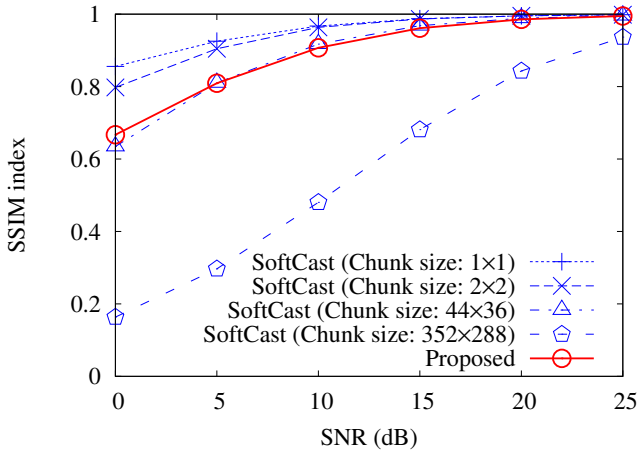


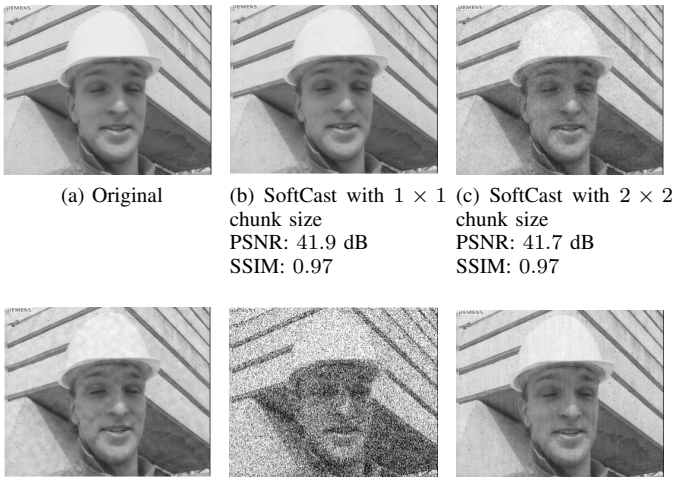
Fig. 8. Average SSIM across 18 test video sequences vs. channel SNR with different chunk size.

- The proposed scheme achieves higher video quality compared to SoftCast with the default chunk size of 44×36.
- Video quality of SoftCast with the largest chunk size is significantly lower than the proposed scheme even with small overhead.

For example, the proposed scheme improves PSNR performance approximately by 0.1 dB compared to SoftCast with the chunk size of 44×36 pixels and 11.4 dB compared to SoftCast with the chunk size of 352×288 pixels on average across channel SNRs of 0 to 25 dB. In addition, the quality differences between the proposed scheme and SoftCast with chunk size of 2×2 and 1×1 pixels are 5.8 dB and 6.3 dB on average, respectively. Note that the PSNR performance degradation does not much affect visual quality in the proposed scheme as shown in Figs. 9 and 10.

In addition to PSNR performance, we also evaluate video quality of the proposed and conventional SoftCast in terms of SSIM. Fig. 8 shows the average SSIM performance across 18 test video sequences as a function of channel SNR. It is shown in this figure that the proposed scheme outperforms SoftCast with a chunk size of 44×36 pixels in a low channel SNR. For example, the proposed scheme achieves SSIM improvement up to 0.03 and 0.50 over SoftCast with the chunk sizes of 44×36 and 352×288 pixels at a channel SNR of 0 dB, respectively. In addition, the average SSIM of the proposed scheme is only 0.01 worse than idealistic SoftCast with a smallest chunk size of 1×1 pixels at SNRs of 15–25 dB. It means that the proposed scheme can yield almost the same perceptual video quality compared to SoftCast having much larger overhead.

Finally, Figs. 9 and 10 compare the visual quality of the proposed and existing schemes for the video sequences of *foreman* and *mobile*. The video frame is transmitted at the channel SNR of 10 dB. For *foreman*, the PSNRs achieved by SoftCast with chunk size of 1×1, 2×2, 44×36, and 352×288 pixels are 41.9 dB, 41.7 dB, 34.5 dB, and 21.3 dB, respectively, whereas 36.2 dB is achieved by the proposed scheme. The SSIMs achieved by SoftCast with chunk size of 1×1, 2×2, 44×36, and 352×288 pixels are 0.97, 0.97, 0.91, and 0.31, respectively, while the proposed scheme shows 0.92 SSIM index. From the snapshots, we can clearly



(a) Original

(b) SoftCast with 1×1 chunk size
PSNR: 41.9 dB
SSIM: 0.97

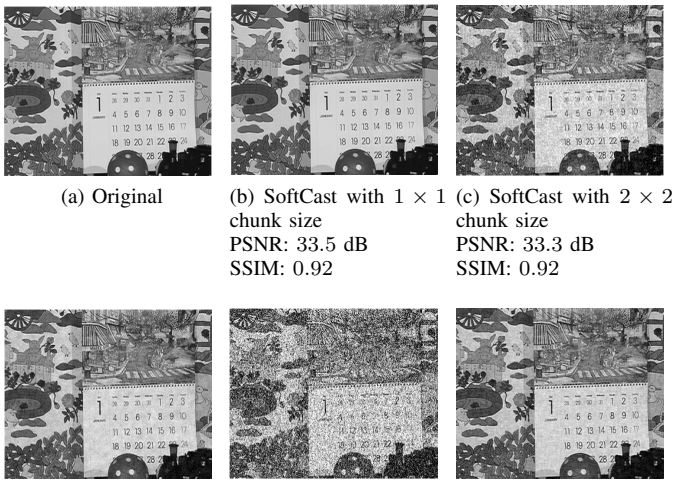
(c) SoftCast with 2×2 chunk size
PSNR: 41.7 dB
SSIM: 0.97

(d) SoftCast with 44×36 chunk size
PSNR: 34.5 dB
SSIM: 0.91

(e) SoftCast with 352×288 chunk size
PSNR: 21.3 dB
SSIM: 0.31

(f) Proposed
PSNR: 36.2 dB
SSIM: 0.92

Fig. 9. Snapshot of *foreman* (frame #1) in each scheme at an SNR of 10 dB.



(a) Original

(b) SoftCast with 1×1 chunk size
PSNR: 33.5 dB
SSIM: 0.92

(c) SoftCast with 2×2 chunk size
PSNR: 33.3 dB
SSIM: 0.92

(d) SoftCast with 44×36 chunk size
PSNR: 29.4 dB
SSIM: 0.86

(e) SoftCast with 352×288 chunk size
PSNR: 20.7 dB
SSIM: 0.59

(f) Proposed
PSNR: 28.7 dB
SSIM: 0.85

Fig. 10. Snapshot of *mobile* (frame #1) in each scheme at an SNR of 10 dB.

see that SoftCast schemes with large chunk sizes provide low-quality images. In contrast, the proposed scheme achieves a clean image with details and almost the same visual compared to SoftCast schemes with a small chunk size.

D. Discussion on Bandwidth Limitation

Above evaluations assumed that all schemes can transmit all DCT coefficients and metadata to the receiver without any constraints. However, conventional SoftCast schemes need to send more metadata than our scheme to achieve high video quality. To evaluate an impact of overhead reduction of the proposed scheme on video quality, this section considers the identical bandwidth constraint over the proposed and existing schemes for fair comparisons.

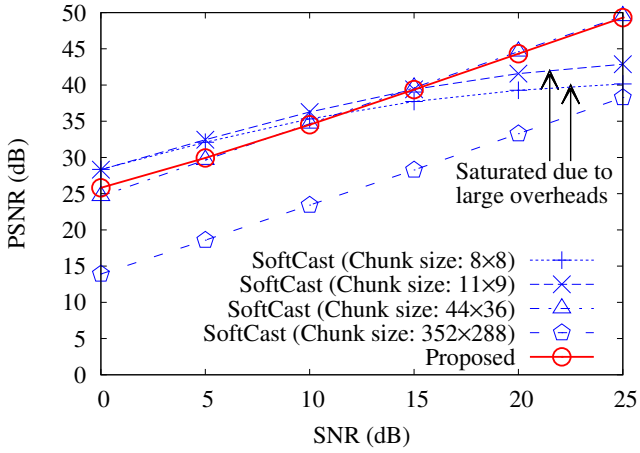


Fig. 11. Average PSNR performance across 18 test video sequences vs. channel SNRs at a channel symbol rate of 1.5 Msymbols/sec.

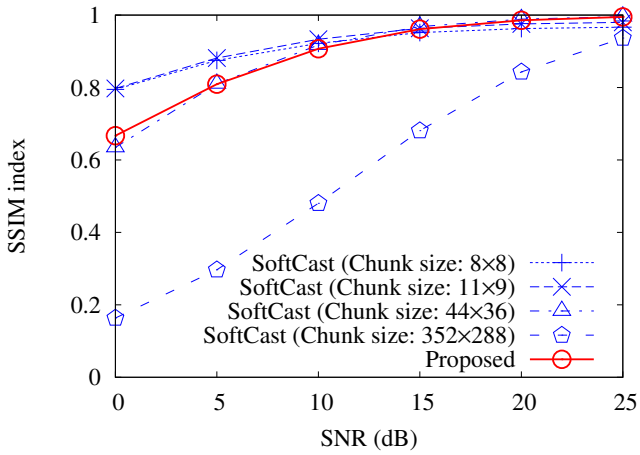


Fig. 12. Average SSIM performance across 18 test video sequences vs. channel SNRs at a channel symbol rate of 1.5 Msymbols/sec.

For the comparison, we set the same channel symbol rate over the proposed and existing schemes. Each scheme sends both analog-modulated symbols and BPSK-modulated metadata symbols with $1/2$-rate convolutional coding at a certain channel symbol rate. When the total number of modulated symbols exceeds the maximum number of transferable symbols at the channel symbol rate, a sender discards analog-modulated symbols from the ones having smaller power to constrain the total number of transmission symbols. In this case, the receiver regards the discarded coefficients as zeros. We first use the channel symbol rate of approximately 1.5 Msymbols/sec, within which the proposed scheme can send all the analog-modulated and BPSK-modulated symbols, and later compare the video quality for lower channel symbol rates.

Figs. 11 and 12 show the average PSNR and SSIM performance across 18 test video sequences as a function of channel SNR at a channel symbol rate of 1.5 Msymbols/sec, respectively. We compare SoftCast with chunk size of 8×8,

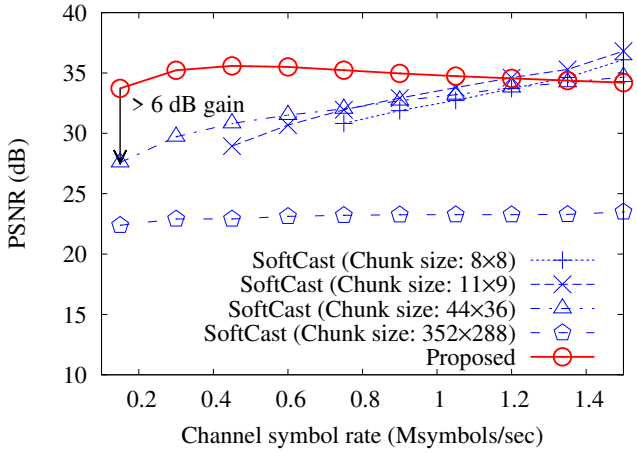


Fig. 13. Average PSNR performance across 18 test video sequences vs. channel symbol rate at a channel SNR of 10 dB.

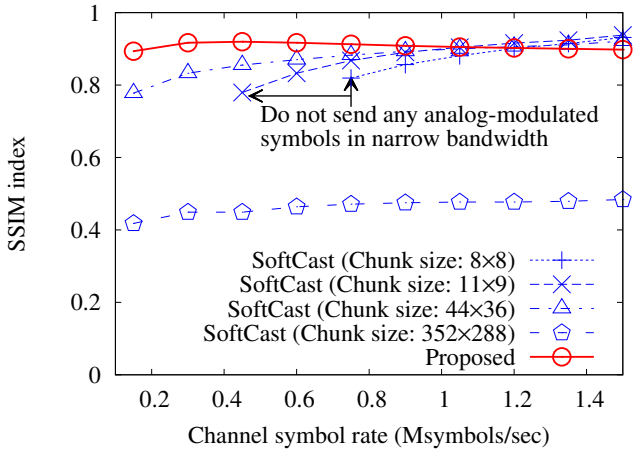


Fig. 14. Average SSIM performance across 18 test video sequences vs. channel symbol rate at a channel SNR of 10 dB.

11×9, 44×36, and 352×288 pixels. Note that we do not present the results of SoftCast schemes with smaller chunk size such as 4×4 pixels in those figures because the performance results were extremely poor due to the large overhead.

These results show that the proposed scheme achieves higher video quality compared to SoftCast schemes with a small chunk size in high channel SNRs at the same bandwidth constraint. For example, the proposed scheme achieves PSNR improvement by 1.7, 0.4, 0.1, and 11.2 dB over SoftCast with the chunk size of 8×8, 11×9, 44×36, and 352×288 pixels, respectively, across channel SNRs of 0 dB to 25 dB.In addition, PSNR of SoftCast with a small chunk size is saturated at high channel SNRs. Since DCT coefficients with small power are discarded to satisfy the bandwidth constraint, the improvement of video quality is limited in high channel SNR regimes. We also confirmed that the impact of metadata transmission failure was marginal for both SoftCast and our method in channel SNRs greater than 0 dB when using $1/2$-rate convolutional encoding for metadata. Although higher code rates for metadata transmission can degrade the performance more significantly, the impact can be negligible in high SNR regimes.

Figs. 13 and 14 show the average PSNR and SSIM performance across 18 test video sequences as a function of channel symbol rate at a channel SNR of 10 dB, respectively. Note that results of SoftCast with chunk size of 8×8 and 11×9 pixels at a low channel symbol rate are not shown in those figures because the same reason as above. We note that the proposed scheme outperforms SoftCast schemes even in narrow-band environment. In particular, video quality at the channel symbol rate between 0.3 and 1.35 Msymbols/sec is slightly better than the case at a channel symbol rate of 1.5 Msymbols/sec because DCT coefficients with too small power can waste transmission power. For example, the proposed scheme improves PSNR performance by 7.0, 5.0, 2.7, and 11.7 dB in comparison to SoftCast with a chunk size of 8×8, 11×9, 44×36, and 352×288 pixels, respectively, across channel symbol rates of 0.15 to 1.5 Msymbols/sec.

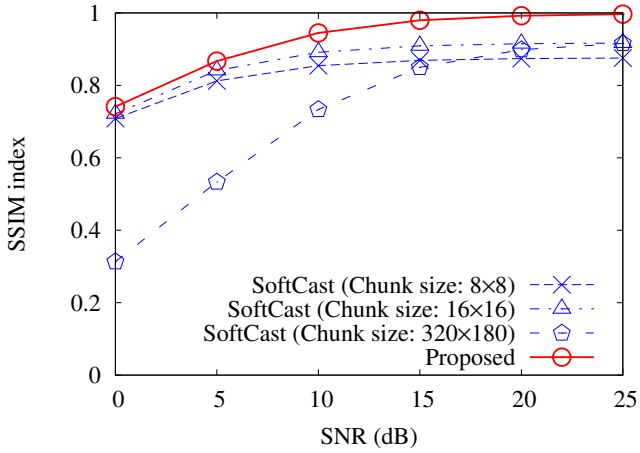


Fig. 15. Average SSIM performance across two test video sequences in the HD format vs. channel SNRs at a channel symbol rate of 19.3 Msymbols/sec.

E. Discussion on High-Resolution Videos

Previous sections used low-resolution videos, i.e., CIF format, to demonstrate an impact of the proposed scheme. Here, we discuss the effect of our proposed fitting function on high resolution videos. Fig. 15 shows the SSIM performance of SoftCast and the proposed schemes using two video sequences, namely, *Johnny* and *KristenAndSara*, in the HD format at a channel symbol rate of 19.3 Msymbols/sec. In this case, we use three chunk sizes, i.e., 8×8, 16×16, and 320×180, in SoftCast.

From this figure, we can see that the proposed scheme can yield better video quality even in high-resolution videos, and thus it is confirmed that the proposed fitting function is effective irrespective of video resolutions. In addition, comparing with Fig. 8, it was found that our proposed scheme may be more effective for higher resolutions. This may be because the amount of metadata in SoftCast schemes increases in higher-resolution videos, while that in the proposed scheme does not depend on the video resolution. The increase of metadata

causes degradation of video quality due to the deletion of more DCT coefficients for band-fair comparisons.

IV. Related Research

Our study is related to various works on soft video delivery and model-based video delivery schemes. In this section, we introduce some related works in order to highlight the contributions of our paper.

A. Soft Video Delivery

To prevent the cliff effect in wireless video delivery, soft video delivery schemes have been proposed recently [10]–[12], [14]–[18]. SoftCast [10] is a pioneering soft video delivery system. It directly sends linearly-transformed video signals by using analog modulation to ensure the received video quality is proportional to the channel quality. On the other hand, it limits the quality improvement due to large overhead. DCast [11], [12], [14] aim at the quality improvement of SoftCast by using coset coding and motion-compensated temporal filtering. These techniques can reduce the power of the video signals, and thus a sender assigns large transmission power to whole video signals. Some studies adopted compressive sensing [29] for video signals to enhance packet loss resilience [15], [16], [26]. Other studies [17], [18] extend the concept of analog scheme to multiple-antenna systems. A sender adaptively assigns transmission power and analog-modulated symbols to antennas based on the channel estimates. However, the above conventional studies are oblivious of an impact of overhead on video quality. A recent study [20] discusses an impact of different chunk sizes on video quality in SoftCast. However, the study does not focus on efficient overhead reduction.

In contract to the conventional analog schemes, we aim at overhead reduction and video quality improvement. There are no studies to decrease overheads while keeping high video quality in analog schemes to the best of our knowledge. To this end, the proposed scheme uses a GMRF for modeling video signals. Based on the model, we find a Lorentzian fitting function to obtain power values of DCT coefficients, i.e., metadata, with a few parameters. Since the fitting function can be estimated with a small error, the proposed scheme simultaneously achieves high video quality and small overhead compared to SoftCast. Note that the fitting-based operations can be applied to most of existing analog schemes to enhance the quality improvement by reducing overhead.

B. Model-based Video Delivery

Some studies use a model for video signals to improve the performance of video delivery [30]–[35]. The model is used to obtain required values for efficient encoding and decoding operations with small overhead. For encoding operations, a sender can estimate rate distortion (RD) curves from a model, e.g., Laplacian distributions [30]–[32] and Cauchy–Lorentz distributions [33]. By using the estimated RD curves, high video quality can be achieved in a certain bandwidth with short encoding time. For decoding operations, a receiver estimates original pixel values by using GMRF [34], [35]. The estimated pixel values can be used for error concealment operations to improve loss-resilience of video delivery.

To date, there is no study, which introduced a GMRF model for analog transmission schemes in literature. We incorporated the GMRF model into soft video delivery in order to obtain the power values of DCT coefficients at a receiver with reduced overhead. Since the power can be fit by a function with a small error, the proposed scheme can achieve high video quality with significant overhead reduction.

V. Conclusions

We have proposed a new analog transmission scheme based on a simple GMRF model to maintain high video quality while achieving a significant reduction in metadata overhead. The proposed scheme finds parameters for a fitting function to obtain the power of DCT coefficients with small overhead. Through performance evaluations, we have observed that the proposed scheme achieves higher video quality compared to conventional SoftCast schemes. In addition, the proposed scheme significantly reduces the required amount of overhead. The overhead reduction in turn enables more efficient resource allocation for analog-modulated symbols within a bandwidth, and results in additional quality improvement compared to conventional schemes in both broad- and narrow-band environments.

Acknowledgment

The authors would like to thank Dr. David S. Millar at MERL for useful discussions. This work was partly supported by JSPS KAKENHI Grant Numbers 14J09461 and 17K12672.

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2015-2020," feb 2016.

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

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

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

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

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

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

[8] M. M. Ghandi and M. Ghanbari, "Layered H.264 video transmission with hierarchical QAM," *Journal of Visual Communication Image representation*, vol. 17, no. 2, pp. 451–466, 2006.

[9] 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, 2014.

[10] 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.

[11] X. Fan, F. Wu, and D. Zhao, "D-Cast: DSC based soft mobile video broadcast," in *ACM International Conference on Mobile and Ubiquitous Multimedia*, 2011, pp. 226–235.

[12] 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.

[13] J. Wu, J. Wu, H. Cui, C. Luo, X. Sun, and F. Wu, "DAC-Mobi: Data-assisted communications of mobile images with cloud computing support," *IEEE Transactions on Multimedia*, vol. 18, no. 5, pp. 893–904, 2016.

[14] 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.

[15] 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, 2014.

[16] 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, 2014.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] H. Rue and H. Leonhard, *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.

[22] C. Zhang and D. Florencio, "Analyzing the optimality of predictive transform coding using graph-based models," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 106–109, 2013.

[23] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Quality improvement and overhead reduction for soft video delivery," in *IEEE International Conference on Communications*, 2016, pp. 1–6.

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

[25] S. Jakubczak, H. Rahui, and D. Katabi, "One-size-fits-all wireless video," in *ACM HotNets*, 2009, pp. 1–6.

[26] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "Compressive sensing for loss-resilient hybrid wireless video transmission," in *IEEE Globecom*, 2015, pp. 1–5.

[27] Y. Huang, T. Z. Fu, D. M. Chiu, J. C. Lui, and C. Huang, "Challenges, design and analysis of a large-scale p2p-vod system," in *ACM SIGCOMM*, 2008, pp. 375–388.

[28] Xiph, "Xiph.org media." [Online]. Available: http://media.xiph.org/video/derf/

[29] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[30] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based lagrangian rate distortion optimization for hybrid video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 2, pp. 193–205, 2009.

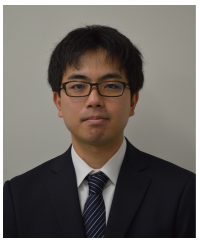
[31] C. Li, H. Xiong, and D. Wu, "Delay-rate-distortion optimized rate control for end-to-end video communication over wireless channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 10, pp. 1665–1681, 2015.

[32] W. Gao, S. Kwong, H. Yuan, and X. Wang, "DCT coefficient distribution modeling and quality dependency analysis based frame-level bit allocation for hevc," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 139–153, 2016.

[33] S. Hu, H. Wang, S. Kwong, and C. C. J. Kuo, "Novel rate-quantization model-based rate control with adaptive initialization for spatial scalable video coding," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 3, pp. 1673–1684, 2012.

[34] J. Rombaut, A. Pizurica, and W. Philips, "Passive error concealment for wavelet-coded i-frames with an inhomogeneous gauss-markov random field model," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 783–796, 2009.

[35] D. Persson and T. Eriksson, "Mixture model- and least squares-based packet video error concealment," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 1048–1054, 2009.



Takuya Fujihashi (M'16) received the B.E. degree in 2012 and the M.S. degree in 2013 from Shizuoka University, Japan. In 2016, he received Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan. He is currently an assistant professor at the Graduate School of Science and Engineering, Ehime University since Jan. 2017. He was research fellow (PD) of Japan Society for the Promotion of Science in 2016. From 2014 to 2016, he was research fellow (DC1) of Japan Society for the Promotion of Science. From 2014 to 2015, he was an intern at Mitsubishi Electric Research Labs. (MERL) working with the Electronics and Communications group. He selected one of the Best Paper candidates in IEEE ICME (International Conference on Multimedia and Expo) 2012. His research interests are in the area of video compression and communications, with a focus on multi-view video coding and streaming over high and low quality networks.

Toshiaki Koike-Akino (M'05–SM'11) received the B.S. degree in electrical and electronics engineering, M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2002, 2003, and 2005, respectively. During 2006–2010 he was a Postdoctoral Researcher at Harvard University, and joined Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, in 2010. His research interests include digital signal processing for data communications and sensing. He received the YRP Encouragement Award 2005, the 21st TELECOM System Technology Award, the 2008 Ericsson Young Scientist Award, the IEEE GLOBECOM'08 Best Paper Award in Wireless Communications Symposium, the 24th TELECOM System Technology Encouragement Award, and the IEEE GLOBECOM'09 Best Paper Award in Wireless Communications Symposium.

Takashi Watanabe (S'83–M'87) is a Professor of Graduate School of Information Science and Technology, Osaka University, Japan since 2013. He received his B.E. M.E. and Ph.D. degrees from Osaka University, Japan, in 1982, 1984 and 1987, respectively. He joined Faculty of Engineering, Tokushima University in 1987 and moved to Faculty of Engineering, Shizuoka University in 1990. He was a visiting researcher at University of California, Irvine from 1995 through 1996. He has served on many program committees for networking conferences, IEEE, ACM, IPSJ, IEICE. His research interests include mobile networking, ad hoc sensor networks, IoT/M2M networks, intelligent transport systems, specially MAC and routing. He is a member of IEEE, IPSJ and IEICE.

Philip V. Orlik (S'97–M'99) was born in New York, NY in 1972. He received the B.E. degree in 1994 and the M.S. degree in 1997 both from the State University of New York at Stony Brook. In 1999 he earned his Ph.D. in electrical engineering also from SUNY Stony Brook. In 2000 he joined Mitsubishi Electric Research Laboratories located in Cambridge, MA, where he is currently the Group Manager of the Electronics&Communications Group. His primary research focus is on advanced wireless and mobile communications, sensor networks, ad hoc networking and UWB. Other research interests include vehicular/car-to-car communications, mobility modeling, performance analysis, and queuing theory.