

Compressive Sensing for Loss-Resilient Hybrid Wireless Video Transmission

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

TR2015-132 December 2015

Abstract

The quality of wireless video delivery is susceptible to wireless channel instability, including channel fading, interference, noise, and packet loss. To improve the video quality in such wireless channels, analog transmission schemes have been proposed recently. However, the existing analog schemes have two drawbacks in high energy of video signals and loss resilience. To overcome the issues, we propose a new hybrid digital-analog transmission scheme, which jointly uses digital coding and analog coding based on compressive sensing. Digital coding generates the residual signals between the original and the encoded signals to decrease the energy of data for analog coding. Compressive sensing redistributes the energy across whole video packets to increase the resistance to packet loss. We show that our proposed hybrid scheme improves video quality by up to 11.6 dB compared to the existing analog scheme in an erasure wireless channel. In addition, our proposed scheme also improves video quality by 1.5 dB compared to the existing hybrid scheme.

2015 IEEE Global Communications Conference (GLOBECOM)

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 acknowledgement 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.

Compressive Sensing for Loss-Resilient Hybrid Wireless Video Transmission

Takuya Fujihashi^{†*}, Toshiaki Koike-Akino[†], Takashi Watanabe^{*}, and Philip V. Orlik[†]

[†]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA

^{*}Graduate School of Information and Science, Osaka University, Osaka, Japan

Abstract—The quality of wireless video delivery is susceptible to wireless channel instability, including channel fading, interference, noise, and packet loss. To improve the video quality in such wireless channels, analog transmission schemes have been proposed recently. However, the existing analog schemes have two drawbacks in high energy of video signals and loss resilience. To overcome the issues, we propose a new hybrid digital-analog transmission scheme, which jointly uses digital coding and analog coding based on compressive sensing. Digital coding generates the residual signals between the original and the encoded signals to decrease the energy of data for analog coding. Compressive sensing redistributes the energy across whole video packets to increase the resistance to packet loss. We show that our proposed hybrid scheme improves video quality by up to 11.6 dB compared to the existing analog scheme in an erasure wireless channel. In addition, our proposed scheme also improves video quality by 1.5 dB compared to the existing hybrid scheme.

I. INTRODUCTION

Wireless video services, such as mobile television and video surveillance, have been increasingly popular. According to Cisco visual networking index studies, two-thirds of the world's mobile data traffic will be video by 2016 [1]. In conventional video streaming, the digital video compression and transmission parts operate separately. For example, the video compression part uses H.264/AVC to generate a compressed bit stream using quantization and entropy coding. The rate of the bit stream depends on the available bandwidth and the wireless channel quality. The transmission part uses a channel coding and a modulation scheme to reliably transmit the H.264/AVC encoded bit stream. However, the conventional scheme has the following problems due to the wireless channel variability, including channel noise and packet loss. First, the encoded bit stream is highly vulnerable for bit errors. When the channel signal-to-noise ratio (SNR) falls under a certain threshold, the video quality drops significantly due to high bit-error rate. This phenomenon is referred to as *cliff effect*. Second, the video quality does not improve even when the wireless channel quality is improved. Finally, quantization is a lossy process and its distortion cannot be recovered at the receiver.

To overcome the above-mentioned problems, analog transmission schemes [2], [4] have been proposed. For example, SoftCast [2] directly transmits linear-transformed video signals over a lossy analog channel, and allocates power for the signals to maximize video quality. As depicted in Fig. 1, SoftCast uses

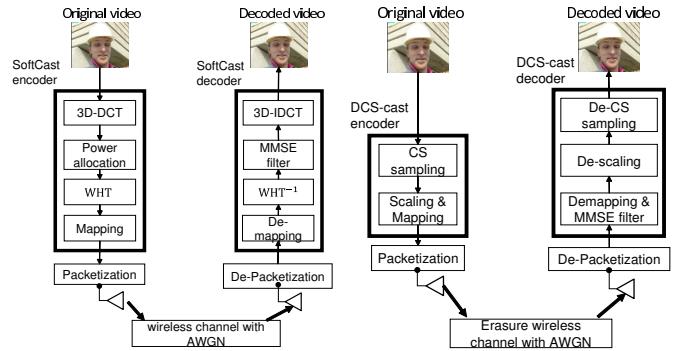


Fig. 1: SoftCast [2].

Fig. 2: DCS-cast [3].

Walsh–Hadamard transform (WHT) to redistribute energy of video signals across whole video packets for the resilience against packet loss. In contrast to the conventional scheme, the video quality of SoftCast is proportional to the wireless channel quality. However, when some packets are lost during communications, the quality of SoftCast degrades significantly. To keep high video quality even in such an erasure wireless channel, compressive sensing (CS) techniques [5], [6] have been recently introduced to analog transmission schemes [3], [7], [8]. Fig. 2 shows a schematic of distributed compressed sensing based multicast scheme (DCS-cast) proposed in [3], where CS is applied for SoftCast to increase the tolerance against packet loss. However, in theory, any analog scheme with linear mapping is relatively inefficient because the performance of an analog scheme can degrade as the ratio of maximum to minimum variances of source data increases [9].

This paper aims at two objectives: 1) to achieve higher video quality with the improvement of the wireless channel quality and 2) to achieve higher video quality even in some video packets are lost during communications. To this end, we propose a CS-based hybrid digital-analog transmission scheme, which integrates digital coding, analog coding, and CS. Each technique plays the following roles in video quality and loss resilience.

- **Digital coding:** Our scheme encodes original video using the standard H.264/AVC encoder and uses binary phase-shift keying (BPSK) for modulation. In addition, the encoder calculates residuals between the original and digital-encoded video frames. The residuals reduce the ratio of maximum to

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

minimum variances, and boost the benefit of analog coding.

- **Compressive sensing:** The residuals are re-sampled by CS. CS improves the loss resilience of the residuals by redistributing the energy across whole video packets.
- **Analog coding:** The re-sampled residuals are directly mapped to Q (quadrature-phase) component to avoid interference to the digital-modulated symbols. Analog transmission ensures that the received video quality is proportional to the wireless channel quality.

Through performance evaluations, we show 1) that our proposed scheme achieves higher video quality compared to the analog schemes as the the wireless channel quality improves, and 2) that the proposed scheme prevents a significant penalty in the video quality compared to the analog schemes even when the half number of video packets is lost.

II. COMPRESSIVE SENSING

We first describe some basics of compressive sensing techniques in this Section.

A. Problem Formulation

Let $\mathbf{x} \in \mathbb{R}^{N \times 1}$ denote a real-valued original signal with length N . We consider an invertible matrix $\Psi \in \mathbb{R}^{N \times N}$:

$$\tilde{\mathbf{x}} = \Psi \mathbf{x}, \quad (1)$$

where $\tilde{\mathbf{x}}$ is a K -sparse vector, i.e., $\|\tilde{\mathbf{x}}\|_0 = K \ll N$ with $\|\cdot\|_p$ being p -norm. This means that the signal has a sparse representation in some transform functions like discrete cosine transform (DCT) and discrete wavelet transform (DWT). We suppose that a receiver wishes to reconstruct \mathbf{x} from noisy measurement signals $\mathbf{m} = \Phi \mathbf{x} + \mathbf{n} \in \mathbb{R}^{M \times 1}$ with length M ($M < N$). Here, $\mathbf{n} \in \mathbb{R}^{M \times 1}$ is a measurement noise vector, and $\Phi \in \mathbb{R}^{M \times N}$ is an orthogonal measurement matrix such that $\Phi \Phi^T = \mathbf{I}$, where $[\cdot]^T$ and \mathbf{I} denote a transpose and an identity matrix, respectively. In general, recovering \mathbf{x} from \mathbf{m} is impossible because the system is underdetermined. However, if $\tilde{\mathbf{x}}$ is sufficiently sparse with restricted isometry property, the CS theory [6] tells us that \mathbf{x} can be reconstructed from \mathbf{m} using ℓ_1 minimization problem as follows.

$$\min_{\tilde{\mathbf{x}}} \|\tilde{\mathbf{x}}\|_1 \quad \text{s.t.} \quad \|\mathbf{m} - \Phi \Psi^{-1} \tilde{\mathbf{x}}\|_2^2 < \xi, \quad (2)$$

where ξ is a small tolerance.

B. Reconstruction Algorithms

To reconstruct an original signal from a noisy measurement signal, there are mainly two types of reconstruction algorithms: matching pursuit [10]–[12] and iterative thresholding [13]–[15]. In matching pursuit [10]–[12], the original signal is reconstructed by using a linear combination of M columns from the measurement matrix Φ . To obtain the linear combination, matching pursuit sequentially selects the column of Φ , which is most strongly correlated with the measurement vector \mathbf{m} at each iteration. The contribution of the column is subtracted from the measurement vector and the same process is repeated on the residual measurement vector. Iterative thresholding

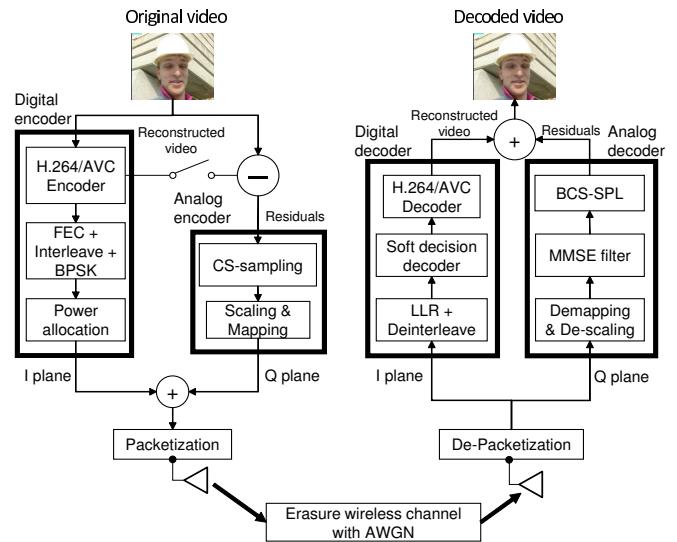


Fig. 3: Proposed hybrid digital-analog transmission scheme.

[13]–[15] first backprojects the noisy measurement vector to estimate an original signal \mathbf{x} , i.e., $\tilde{\mathbf{x}} = \Phi^T \mathbf{m}$. The elements of $\tilde{\mathbf{x}}$ are modified by a thresholding function at each iteration to impose sparseness as follows:

$$\tilde{\mathbf{x}}^{(l+1)} = \Psi^{-1} \mathbb{T}_\tau(\Psi \tilde{\mathbf{x}}^{(l)}), \quad (3)$$

where $\mathbb{T}_\tau(\cdot)$ is a thresholding function with a sparseness parameter of τ , and Ψ is used to transform the output of the (l) th iteration $\tilde{\mathbf{x}}^{(l)}$ onto a sparse domain. Our proposed video transmission scheme employs the block-wise compressed sensing (BCS-SPL) [16], which is one of iterative thresholding algorithms, for the reconstruction. In contrast to the other algorithms, BCS-SPL requires small storage and short processing time during the reconstruction.

III. PROPOSED DIGITAL-ANALOG VIDEO TRANSMISSION USING COMPRESSIVE SENSING

A. Overview

The purposes of our study are 1) to achieve high video quality as the wireless channel quality improves and 2) to keep high video quality even when the wireless packet is loss-prone.

Fig. 3 shows a schematic of our proposed scheme. The encoder and decoder are divided into digital and analog parts. The digital encoder first encodes the original video frames in group of pictures (GoP), which is a sequence of successive video frames, using the standard H.264/AVC encoder to generate a bit stream. The bit stream is then channel coded, interleaved, BPSK-modulated, and power allocated by the digital encoder. On the other hand, the analog encoder calculates the residuals from the original and reconstructed video frames. The residuals are CS-sampled, scaled, and analog-modulated. Finally, the outputs of digital and analog encoders are superposed, packetized, and transmitted to the receiver over an erasure wireless channel with additive white Gaussian noise (AWGN).

At the receiver side, the decoder extracts digital- and analog-modulated symbols from I (in-phase) and Q components of the received signals. The digital decoder deinterleaves and channel decodes the digital-modulated symbol, and then reconstructs the video frames using H.264/AVC decoder. The analog decoder uses minimum mean-square error (MMSE) filter and BCS-SPL [16] for the received analog-modulated symbol to obtain the residuals. After the both analog and digital decoding are completed, the receiver adds the residuals to the reconstructed video frames for display.

B. Encoder

1) *Digital Part*: The digital-part encoder consists of H.264/AVC encoder, channel coding, interleaving, and BPSK modulation. The encoder generates an entropy-coded bit stream for video frames in one GoP. The bit stream is channel coded by a convolutional code and interleaved to protect against bursts of bit errors. The interleaved stream is modulated by BPSK and mapped to I component.

2) *Analog Part*: Analog encoder reconstructs video frames from the bit stream, and calculates residuals between the original and reconstructed video frames. The residuals in one GoP are divided into several chunks of size $B \times B$. Each chunk i is converted into a vector \mathbf{v}_i with a length of B^2 . The vectors are CS-sampled to obtain a vector \mathbf{c}_i as follows:

$$\mathbf{c}_i = \Phi \mathbf{v}_i, \quad (4)$$

where the matrix Φ has a size of $B^2 \times B^2$. The matrix Φ is the left-singular vectors of a random matrix, whose elements are independent and identically distributed (*i.i.d.*) random variables generated by a random seed to follow the standard normal distribution. We use the same matrix Φ for all chunks. The CS-sampled values are mapped to Q component.

3) *Power Allocation*: In our scheme, the transmitter decides transmission powers for digital and analog encoders based on the wireless channel quality. The transmitter first decides power allocation for digital encoder to ensure enough power to decode the entropy-coded bit stream correctly. When the channel quality is low, the receiver will have difficulty in decoding the bit stream correctly. For that case, our scheme will switch to pure analog transmission to prevent the cliff effect. To decide the transmission power for digital encoder, the transmitter calculates the power threshold to decode the bit stream correctly:

$$P_{\text{th}} = N_0 \cdot \gamma_0, \quad (5)$$

where P_{th} is the power threshold and N_0 is the average noise power of the wireless channel. Here, γ_0 is the required SNR to guarantee that the decoding bit-error rate (BER) is not larger than the target BER. After the threshold calculation, the transmitter decides the transmission power for digital encoder P_d and the transmission power for analog encoder P_a , as follows:

$$P_d = \begin{cases} P_{\text{th}}, & P_{\text{th}} \leq P_t, \\ 0, & \text{otherwise,} \end{cases} \quad P_a = P_t - P_d, \quad (6)$$

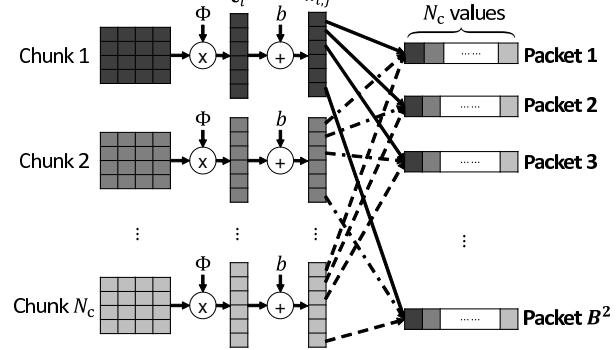


Fig. 4: Packetization of our scheme. Pixel values in each chunk are sampled by the same matrix Φ . Each element of CS-sampled vectors is superposed with a BPSK-modulated symbol. The superposed symbols are assigned to the corresponding packets.

where P_t is the total power budget.

Let $x_{i,j}$ denote a transmission symbol of CS-sampled value j in chunk i . The symbol $x_{i,j}$ is formed by superposing a BPSK-modulated symbol $x_{i,j}^{(d)}$ and analog-modulated symbol $x_{i,j}^{(a)}$ as follows:

$$x_{i,j} = x_{i,j}^{(d)} + j x_{i,j}^{(a)}, \quad (7)$$

where $j = \sqrt{-1}$ denotes the imaginary unit. The BPSK-modulated symbol and the analog-modulated symbol are scaled by P_d and P_a , respectively, as follows:

$$x_{i,j}^{(d)} = \sqrt{P_d} \cdot b_{i,j}, \quad x_{i,j}^{(a)} = \sqrt{P_a} \cdot r_{i,j}, \quad (8)$$

where $b_{i,j} \in \mathbb{X} = \{\pm 1\}$ is the BPSK-modulated symbol and $r_{i,j} \in \mathbf{c}_i$ is the CS-sampled value j in chunk i .

4) *Packetization*: The transmitter assigns the superposed symbols to packets as illustrated in Fig. 4, which is similar to [3]. The transmitter collects the same element of each chunk into one packet as follows:

$$p_k = \bigcup_{i=1}^{N_c} x_{i,k}, \quad (9)$$

where N_c is the total number of chunks in one GoP. After the packetization, the total number of packets is B^2 and transmission symbols in each packet is N_c . The received symbol in packet p_k over an erasure wireless channel with AWGN can be modeled as:

$$y_{i,k} = \begin{cases} x_{i,k} + n_{i,k}, & \text{with a probability of } p, \\ \emptyset, & \text{with a probability of } 1-p, \end{cases} \quad (10)$$

where $y_{i,k}$ is the received symbol, $n_{i,k}$ is an effective noise of element k in chunk i , and p is a packet arrival rate. Here, \emptyset represents that the receiver does not receive the transmitted packet, i.e., the values of I and Q components are unknown. This corresponds to a packet erasure when the receiver is impaired, e.g., by a strong interference, deep fading, and shadowing.

C. Decoder

1) *Digital Part*: The receiver first extracts BPSK-modulated symbol from the I component of each received symbol, i.e., $\Re(y_{i,j})$. To decode the modulated symbol, the digital decoder calculates log-likelihood ratio (LLR) values from the received symbols:

$$L_{i,j} = \begin{cases} 0 & \Re(y_{i,j}) = \emptyset, \\ \ln \frac{\Pr(y_{i,j}|1)}{\Pr(y_{i,j}|0)}, & \text{otherwise.} \end{cases} \quad (11)$$

where $L_{i,j}$ is the LLR value of the received symbol. Here, $\Pr(y_{i,j}|\omega)$ denotes the probability that the received signal is $y_{i,j}$ conditioned on the transmitted bits of ω , i.e., $\Pr(y_{i,j}|\omega) = \frac{1}{\pi\sigma_i^2} \exp(-\frac{1}{\sigma_i^2}(\Re(y_{i,j}) - M(\omega))^2)$ where $M(\omega) \in \sqrt{P_d} \cdot \mathbb{X}$ is the BPSK-modulated symbol for ω . After computing the LLR values for all received symbols, the receiver deinterleaves the LLR values and feeds them into the Viterbi decoder. The Viterbi decoder provides the entropy-coded bit stream at its output and the digital decoder uses the standard H.264/AVC decoder to reconstruct video frames from the bit stream.

2) *Analog Part*: The receiver extracts CS-sampled values from the Q component of each received symbol, i.e., $\Im(y_{i,j})$. The receiver first uses the MMSE filter [3] for the extracted value except $\Im(y_{i,j}) = \emptyset$:

$$s_{i,j} = \frac{\lambda_i}{\lambda_i + N_0} \cdot \Im(y_{i,j}), \quad (12)$$

where λ_i is the variance of CS-sampled values in chunk i . The receiver then generates the $B^2 \times B^2$ matrix Φ using the same random seed at the transmitter. The receiver vectorizes the received CS-sampled values of chunk i into a column vector \mathbf{s}_i . Note that some rows in each column vector may be missed due to the packet loss. In this case, the decoder trims the corresponding rows of the matrix Φ . After the trimming, we solve ℓ_1 minimization problem using BCS-SPL [16]. More specifically, the decoder initializes with $\mathbf{v}_i^{(0)} = \Phi^T \mathbf{s}_i$ and $\hat{\mathbf{v}}^{(0)} = \text{Wiener}[\mathbf{v}^{(0)}]$, where $\text{Wiener}[\cdot]$ is a pixel-wise adaptive Wiener filter for smoothed reconstruction. $\hat{\mathbf{v}}^{(0)}$ is updated using block-wise successive projection and thresholding operation as follows:

$$\hat{\mathbf{v}}_i^{(l)} = \hat{\mathbf{v}}_i^{(l)} + \Phi^T (\mathbf{s}_i - \Phi \hat{\mathbf{v}}_i^{(l)}), \quad (13)$$

$$\check{\mathbf{v}}^{(l)} = \begin{cases} \hat{\mathbf{v}}^{(l)}, & |\Psi \hat{\mathbf{v}}^{(l)}| \geq \tau^{(l)}, \\ \mathbf{0}, & \text{otherwise,} \end{cases} \quad (14)$$

$$\mathbf{v}_i^{(l+1)} = \check{\mathbf{v}}^{(l)} + \Phi^T (\mathbf{s}_i - \Phi \check{\mathbf{v}}^{(l)}). \quad (15)$$

Here, $\mathbf{v}_i^{(l)}$ is the vector representing chunk i of whole frames $\mathbf{v}^{(l)}$ at the (l) th iteration, and $\tau^{(l)}$ is a threshold at the (l) th iteration. This reconstruction terminates when $|D^{(l+1)} - D^{(l)}| < 10^{-4}$ where $D^{(l)} = \frac{1}{\sqrt{N_c}} \|\mathbf{v}_i^{(l)} - \hat{\mathbf{v}}_i^{(l-1)}\|_2$. When the reconstruction terminates at iteration l_{end} , the reconstructed residuals are obtained from $\mathbf{v}^{(l_{\text{end}}+1)}$. The receiver finally adds the reconstructed residuals to the reconstructed video frames for playback.

IV. PERFORMANCE EVALUATIONS

A. Evaluation Settings

Metric: We evaluate the performance of various video transmission schemes in terms of the peak signal-to-noise ratio (PSNR) defined as follows:

$$\text{PSNR} = 10 \log_{10} \frac{(2^L - 1)^2}{\epsilon_{\text{MSE}}}, \quad (16)$$

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. We obtain the average PSNR across whole video frames in each video sequence.

Test Video: We use standard reference video, namely, *foreman*, *akiyo*, and *coastguard* in the QCIF format (176×144 pixels, 30 fps) from the Xiph collection [17]. We first focus on *foreman* in Sections IV-B and IV-D. We then compare the video quality for the other video sequences in Section IV-E.

Video Encoding: To generate a bit stream from the test video and calculate the residuals, we used MATLAB video encoder/decoder [18]. We set the GoP size for all transmission schemes to 8 video frames. For the analog parts of the transmission, the chunk size is set to 8×8 pixels in all schemes. This means that each GoP is divided into 3168 chunks. For schemes with a digital component, we encode the video frames in one GoP into one I-frame and subsequent seven P-frames. We set quantization parameters to hold the number of digital-modulated symbols within the number of the residuals in one GoP. When bit errors occur in the bit stream of a video frame, we regard the video frame and the subsequent video frames as lost. In this case, we copy the nearest video frame, which is decoded successfully, to the lost video frame as an error concealment operation.

Wireless Channel Environment: The transmitter sends B^2 packets in 1 GoP, i.e., 64 packets. We assume that the receiver experiences erasure for $(1 - p)B^2$ packets, where p is the packet arrival rate. In addition, the received symbols are impaired by an AWGN channel.

Channel Coding and Modulation: We use a rate-1/4 convolutional code with a constraint length of 8. For analog modulation, all schemes map one linear-transformed representation to the Q component in each symbol for a fair comparison across schemes.

Basis of CS Reconstruction: Our proposed scheme uses 2-dimensional DWT (2D-DWT) as transform matrix Ψ for the CS reconstruction at the receiver side unless otherwise stated. In Section IV-D, we compare the video quality of our scheme using 2-dimensional DCT (2D-DCT), 3-dimensional DCT (3D-DCT), and 2-dimensional dual-tree DWT (2D-DDWT) [19].

B. Effect of Erasure Channel for Digital Video Transmission

We first evaluate the effect of soft-decision decoding on video quality in the erasure wireless channel. Fig. 5 shows the video quality of BPSK-modulated digital transmission scheme with different packet arrival rates p as a function of the SNR. We set quantization parameters to 35, 37, 38, 41, 44, and

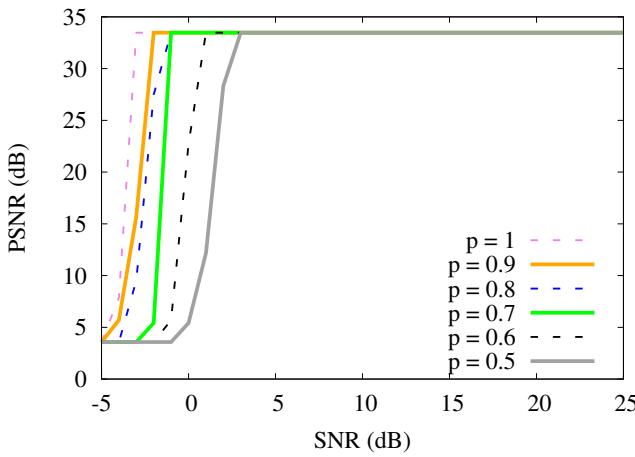


Fig. 5: Quality of digital video transmission with different packet arrival rates p .

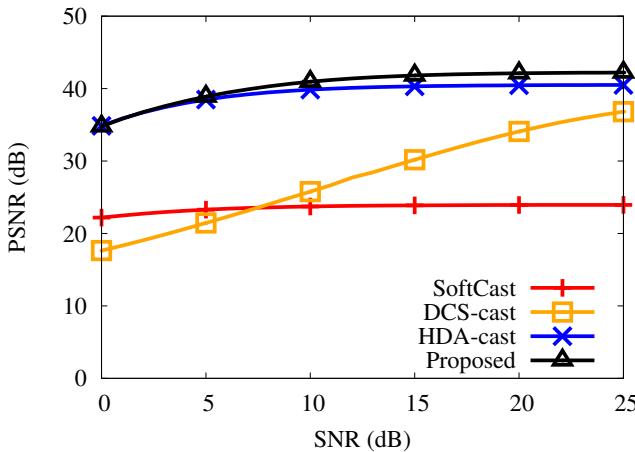


Fig. 6: Video quality of analog and our schemes as a function of SNR at a packet arrival rate p of 0.8.

50 for the packet arrival rates p of 1, 0.9, 0.8, 0.7, 0.6, and 0.5, respectively. Note that we plot the video quality for 1 dB interval. When bit errors occur in video frames, the video quality degrades rapidly. Without any packet loss, the soft-decision decoding can reconstruct BPSK-modulated symbols at an SNR of -3 dB. As the packet arrival rate decreases, the cliff point is shifted to higher SNR. From the above results, we set γ_0 in (5) to $-3, -2, -1, -1, 1$, and 3 dB for the packet arrival rates p of 1, 0.9, 0.8, 0.7, 0.6, and 0.5, respectively, for the following evaluations.

C. Comparison with Analog and Hybrid Schemes

In this section, we discuss the video quality and loss resilience of our scheme. We compare four schemes: SoftCast [2], DCS-cast [3], hybrid digital-analog cast (HDA-Cast) [20], and our proposed scheme. SoftCast directly maps linear-transformed (3D-DCT) coefficients on Q component and uses WHT for the loss resilience. DCS-cast uses CS for original

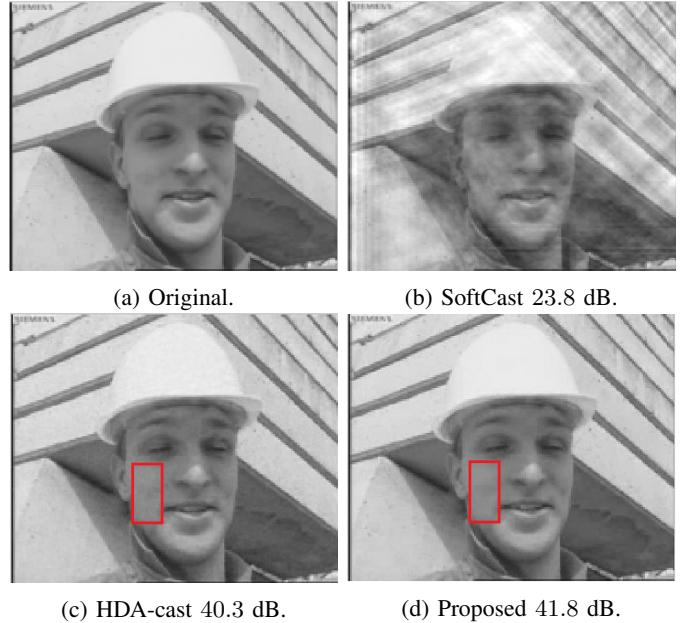


Fig. 7: Snapshot in each scheme.

pixels and maps CS-sampled values on Q component. HDA-Cast is a hybrid transmission scheme, which maps entropy-coded symbols on I component and residuals on Q component. Our scheme uses CS for residuals and maps entropy-coded symbols on I component and CS-sampled residuals on Q component.

Fig. 6 shows the video quality of each scheme as a function of SNR at a packet arrival rate p of 0.8. The proposed method achieves the highest video quality and improves video quality with the improvement of the wireless channel quality. For example, the proposed scheme improves video quality approximately by 17.9 dB compared to SoftCast, 11.6 dB compared to DCS-cast, and 1.5 dB compared to HDA-cast at an SNR of 15 dB. Fig. 7 shows a snapshot of each scheme at the same SNR. Note that the snapshot shows the luminance. For example, the proposed scheme reduces noise at man's cheek compared with HDA-cast.

Fig. 8 shows the video quality of each scheme with different packet arrival rates p at an SNR of 20 dB. From this figure, we observe the following two points:

- 1) The proposed method achieves the highest video quality when the packet arrival rate is low. For example, the proposed scheme improves video quality approximately by 23.1 dB compared to SoftCast, 7.1 dB compared to DCS-cast, and 1.5 dB compared to HDA-cast at a packet arrival rate of 0.7.
- 2) Without any packet loss (i.e., $p = 1$), the performance of HDA-Cast becomes better than our proposed scheme. This is because the power allocation of HDA-Cast plays a dominant role to reduce the channel noise.

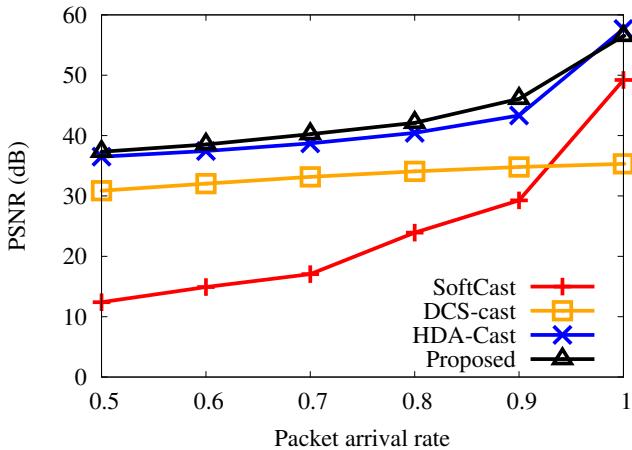


Fig. 8: Video quality of analog and our schemes with different packet arrival rates p at an SNR of 20 dB.

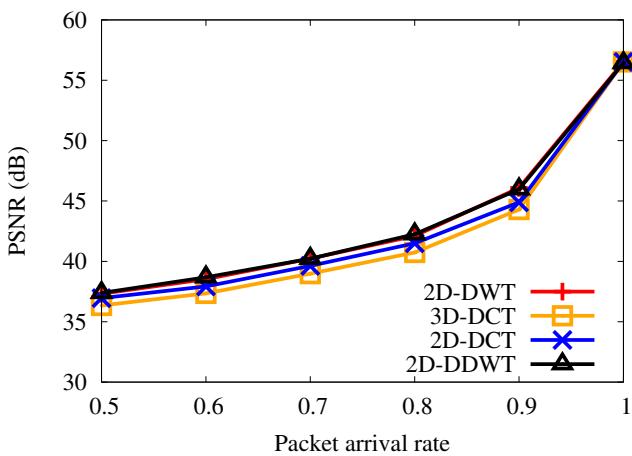


Fig. 9: Video quality of our scheme with different basis functions.

D. Effect of Basis Functions for CS Reconstruction

In previous evaluations, we use 2D-DWT as the basis functions Ψ of CS reconstruction. To evaluate the effect of the basis for our scheme, this section uses 2D-DDWT [19], 2D-DCT, and 3D-DCT as the basis of CS reconstruction. Fig. 9 shows the video quality of our scheme with different basis functions against packet arrival rate p at an SNR of 20 dB. From this figure, we observe that DWT-based reconstruction, i.e., 2D-DWT and 2D-DDWT, achieves higher video quality compared to DCT-based reconstruction. It suggests that DWT-based reconstruction is more advantageous for the hybrid video transmission scheme.

E. Effect of Video Sequences

Previous evaluations use *foreman* for the test video sequence. To evaluate the difference of the video sequence, this section uses *akiyo* and *coastguard* as additional test video

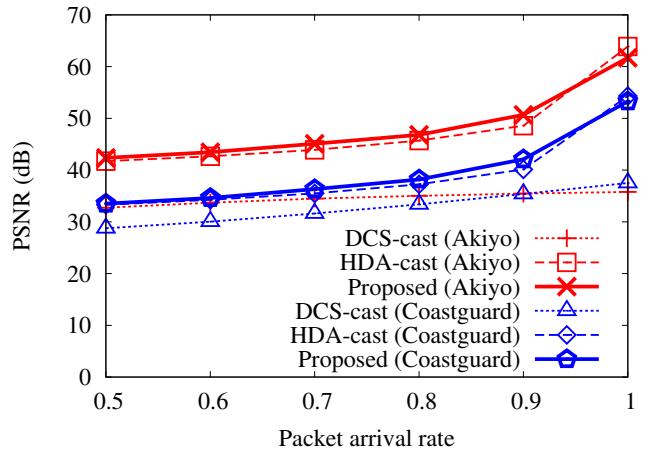


Fig. 10: Video quality with different video sequences.

sequences. For the comparison, we measure the video quality of three schemes: DCS-cast, HDA-Cast, and proposed scheme. Fig. 10 shows the video quality of each scheme with the different test video sequences at an SNR of 20 dB. The figure reveals that the proposed scheme achieves higher video quality compared to the existing DCS-cast and HDA-Cast irrespective of the video sequences. For example, our scheme improves video quality by 10.6 dB compared to DCS-Cast and 1.2 dB compared to HDA-Cast at a packet arrival rate p of 0.7 for the video sequence of *akiyo*.

V. RELATED RESEARCHES

Our study is related to various works on graceful video transmission, CS-based graceful video transmission, and hybrid digital-analog video transmission schemes.

A. Graceful Video Transmission

To prevent the cliff effect in wireless video delivery and to improve the video quality as the wireless channel quality improves, graceful video transmission schemes have been proposed recently [2], [4], [21], [22]. SoftCast [2] uses analog modulation for the video transmission to ensure the received video quality is proportional to the channel quality. FlexCast [21] achieves the goal with bit allocation. They assign many parity bits to most significant bits while few parity bits to least significant bits to reconstruct high-order bits of pixel values correctly. Parallel video unicast (ParCast) [4] and SimCast [22] achieve a similar goal in multiple-antenna systems.

Our proposed scheme uses analog modulation, which is based on SoftCast, for the transmission of residuals to provide graceful quality transition. In addition, our scheme combines digital coding to decrease the energy of analog data, leading to higher video quality in the analog transmission.

B. Compressive Sensing-based Graceful Video Transmission

When packet loss occurs in analog-based graceful transmission schemes, the quality degradation is even high. To keep the high video quality even in the erasure wireless channel, some

studies [3], [7], [8] introduce CS for the graceful transmission schemes. They use CS-sampling for video frames and assume that receivers do not receive some CS-sampled values due to packet loss [3], [8] or antenna diversity [7]. Each receiver reconstructs video frames from the limited number of CS-sampled values.

Our proposed scheme combines digital coding with the CS-based graceful transmission schemes to keep higher video quality even in the erasure channel. Note that the existing schemes assume that the video quality of digital coding degrades rapidly when few packets of digital-coded data are lost during the communication. To improve the loss resilience of the digital-coded data, our scheme uses soft-decision decoding for the data and utilizes the benefit of digital coding even in the loss-prone wireless channel.

C. Hybrid Digital-Analog Video Transmission

In recent years, there are several hybrid digital-analog video transmission schemes proposed [20], [23]–[26] to achieve higher video quality as the wireless channel quality improves. These schemes integrate digital coding with analog coding by transmitting superposition of digital modulation signal and analog modulation signal. The existing hybrid transmission schemes can be divided into two types: bit-level hybrid schemes and frame-level hybrid schemes. References [23], [24] are the bit-level schemes which use digital coding for high-order bits and analog coding for low-order bits of pixel values. References [20], [25] are the frame-level schemes which use digital coding for whole video frames and analog coding for the residuals.

Our proposed scheme is classified into the frame-level hybrid scheme. In contrast to the existing hybrid schemes, our proposed scheme aims to achieve higher video quality even in the loss-prone wireless channel. To increase the resistance against the packet loss during communications, we combine CS with the hybrid schemes and reveal the combination keeps high video quality even in the erasure wireless channel.

VI. CONCLUSIONS

This paper proposes a new CS-based hybrid digital-analog transmission scheme. The proposed scheme combines digital coding, analog coding, and CS to improve video quality gracefully and increase the resistance against packet loss. Evaluations show that our proposed scheme achieves higher video quality compared to existing analog and hybrid video transmission schemes with the improvement of wireless channel quality. In addition, the proposed scheme decreases the degree of quality degradation even in the loss-prone wireless channel.

REFERENCES

- [1] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update 2011–2016,” 2012.
- [2] S. Jakubczak, H. Rahui, and D. Katabi, “One-size-fits-all wireless video,” in *ACM HotNets*, 2009, pp. 1–6.
- [3] 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.
- [4] L. X. Lin, H. Wenjun, P. Qifan, W. Feng, and Z. Yongguang, “Parcast: Soft video delivery in MIMO-OFDM WLANs,” in *ACM MOBICOM*, 2012, pp. 233–244.
- [5] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [6] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [7] 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.
- [8] C. Li, H. Jiang, P. Wilford, Y. Zhang, and M. Scheutzow, “A new compressive video sensing framework for mobile broadcast,” *IEEE Transaction on Broadcast*, vol. 59, no. 1, pp. 197–205, 2013.
- [9] V. Prabhakaran, R. Puri, and K. Ramchandran, “Hybrid digital-analog codes for source-channel broadcast of gaussian sources over gaussian channels,” *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4573–4588, 2011.
- [10] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [11] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *27th Annual Asilomar Conference Signals, Systems, and Computers*, 1993, pp. 40–44.
- [12] D. Needell and J. A. Tropp, “CoSaMP: iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [13] M. Fornasier and H. Rauhut, “Iterative thresholding algorithms,” *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, pp. 187–208, 2008.
- [14] I. Daubechies, M. Fornasier, and I. Loris, “Accelerated projected gradient method for linear inverse problems with sparsity constraints,” *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5–6, pp. 764–792, 2008.
- [15] I. Daubechies, M. Defrise, and C. D. Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [16] S. Mun and J. E. Fowler, “Block compressed sensing of images using directional transforms,” in *IEEE International Conference on Image Processing*, 2009, pp. 3021–3024.
- [17] Xiph., “Xiph.org media.” [Online]. Available: <http://media.xiph.org/video/derf/>
- [18] A. A. Muhi, “H.264 Baseline Codec v2,” <http://www.mathworks.com/matlabcentral/fileexchange/40359-h-264-baseline-codec-v2>, 2013.
- [19] N. Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” *Journal of Applied Computational Harmonic Analysis*, vol. 10, pp. 234–253, 2001.
- [20] 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.
- [21] S. T. Aditya and S. Katti, “FlexCast: Graceful wireless video streaming,” in *ACM MOBICOM*, 2011, pp. 277–288.
- [22] G. Wang, K. Wu, Q. Zhang, and L. M. Ni, “SimCast: Efficient video delivery in MU-MIMO WLANs,” in *IEEE Conference on Computer Communications*, 2014, pp. 2454–2462.
- [23] 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.
- [24] Z. Song, R. Xiong, S. Ma, X. Fan, and W. Gao, “Layered image/video softcast with hybrid digital-analog transmission for robust wireless visual communication,” in *IEEE International Conference on Multimedia and Expo*, 2014, pp. 1–6.
- [25] L. Yu, H. Li, and W. Li, “Wireless cooperative video coding using a hybrid digital-analog scheme,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 436–450, 2015.
- [26] H. Cui, Z. Song, Z. Yang, C. Luo, R. Xiong, and F. Wu, “Cactus: A hybrid digital-analog wireless video communication system,” in *ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems*, 2013, pp. 273–278.